Aristotle's *Physics*
A Collection of Essays

Edited by
LINDSAY JUDSON

CLARENDON PRESS · OXFORD

ARISTOTLE'S *PHYSICS*

This book has been printed digitally and produced in a standard specification in order to ensure its continuing availability

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

Published in the United States by Oxford University Press Inc., New York

Reprinted 2003

ISBN 0-19-823602-6

ACKNOWLEDGEMENTS

I should like to thank my wife, Jenny Tyler, for all her help in the preparation of this volume, and my daughter Anna Penelope for sometimes being quiet enough for me to hear my wife's comments.

LJ

CONTENTS

NOTES ON CONTRIBUTORS

ROBERT BOLTON is Professor of Philosophy at Rutgers University. He is the author of a number of articles on Aristotle.

DAVID BOSTOCK is a Fellow of Merton College, Oxford. His publica tions include *Logic and Arithmetic*, *Plato's Phaedo*, and *Plato's Theaetetus*.

DAVID CHARLES is a Fellow of Oriel College, Oxford. He is the author of *Aristotle's Philosophy of Action*.

WILLIAM CHARLTON is Senior Lecturer in Philosophy at the University of Edinburgh. His publications include *Aristotle's Physics, Books I and II* in the Clarendon Aristotle Series and a translation of Philoponus' commentary on *De Anima* III. 4–8, as well as articles on ancient philosophy.

CYNTHIA A. FREELAND is Associate Professor of Philosophy at the University of Houston. She has published articles on Aristotle's science, metaphysics, and ethics, and has also written on topics in aesthetics and feminist theory.

MARY LOUISE GILL is Associate Professor of Classics and Philosophy at the University of Pittsburgh. She is co-editor of *Ancient Philosophy*, and the author of *Aristotle on Substance: The Paradox of Unity* as well as of a number of articles on Plato and Aristotle.

EDWARD HUSSEY is University Lecturer in Ancient Philosophy at Oxford and a Fellow of All Souls College. He has published *The Presocratics* and *Aristotle's Physics, Books III and IV* in the Clarendon Aristotle Series.

MICHAEL INWOOD is Fellow and Tutor in Philosophy at Trinity College, Oxford. He is the author of *Hegel*, and of a number of articles on Hegel.

LINDSAY JUDSON is Official Student and Tutor in Philosophy at Christ Church, Oxford. He is Associate General Editor of the Clarendon Aristotle Series, and has published a number of articles on the history of philosophy.

JULIUS M. MORAVCSIK is Professor of Philosophy at Stanford University. He has written numerous articles on ancient philosophy, has edited several anthologies of essays, and is the author of *Thought and Language*.

ABBREVIATIONS

Cat.	*Categories*
Int.	*De Interpretatione*
An. Pr.	*Prior Analytics*
An. Post.	*Posterior Analytics*
Top.	*Topics*
Soph. El.	*Sophistici Elenchi*
Phys.	*Physics*
Cael.	*De Caelo*
GC	*De Generatione et Corruptione*
Meteor.	*Meterologica*
An.	*De Anima*
Sens.	*De Sensu*
Mem.	*De Memoria et Reminiscentia*
Som.	*De Somno*
Insom.	*De Insomniis*
Div.	*De Divinatione Per Somnum*
Long.	*De Longitudine et Brevitate Vitae*
Resp.	*De Respiratione*
HA	*Historia Animalium*
PA	*Parts of Animals*
MA	*De Motu Animalium*
IA	*De Incessu Animalium*
GA	*Generation of Animals*
Met.	*Metaphysics*
EN	*Nicomachean Ethics*
EE	*Eudemian Ethics*
Rhet.	*Rhetoric*

I

Aristotle's Method in Natural Science: *Physics* I

ROBERT BOLTON

I Aristotle's Concern with Method

Aristotle begins his *Physics*, and thus his studies in natural science generally, with a chapter on method. Further remarks on this subject are added at crucial junctures later in the work, particularly in the subsequent chapters of Book I and in Book IV. This concern with method in the *Physics* is just one prominent example of the explicit and self-conscious attention to method which is displayed in nearly all of Aristotle's major works. Among other things, this interest reflects the lively concern, and the lively disagreement, among philosophers and scientists in Aristotle's own day not only about positive doctrine in various areas but also about the proper method or methods to use to reach such doctrine.[1]

An equally lively concern, and equally lively disagreement, about how to *understand* Aristotle's views on method has figured prominently in recent Aristotelian scholarship. The *Physics* itself has served as a *locus classicus* for this disagreement. Traditionally scholars have found the notion congenial that Aristotle's intended method in his works on natural science is empirical, even as they have criticized him for failures on this count. The current generation has reversed this verdict entirely. The *Physics* in particular is now standardly taken as a paradigm of Aristotle's use of dialectical method, understood as a largely conceptual or a priori technique of inquiry appropriate for philosophy, as opposed to the more empirical inquiries which we,

[1] Of many relevant passages, one not usually noticed, *GC* I. 2, 316^a5–14, nicely illustrates Aristotle's stand on one of these important disagreements about method which will be of particular interest here.

these days, now typically regard as scientific.[2] This sharp difference of interpretation raises a question which takes us, clearly enough, to the heart of Aristotle's conception of the nature of philosophy and of science. Does Aristotle follow Plato in thinking of philosophy and science as, in important respects, non-empirical in its method, or does he oppose Plato on this as on so many other points? This is, in effect, the main question with which we will be concerned here.

II Scientific Method in *Physics* I. 1: Some Puzzles

Though Aristotle begins his *Physics* with a chapter on method, recent studies have largely ignored this chapter and have concentrated mainly on Book IV. This may be due, in part, to the comparative obscurity of the initial discussion but, clearly enough, further study should take full account of Aristotle's own point of departure and its application in Book I.[3] The initial chapter is compact enough to include here in full.

> Knowledge and scientific knowledge, in every inquiry where there are principles or causes or elements, comes from the grasp of these things. For we think we know a thing when we grasp its primary causes and its first principles and we are down to its elements. It is evident, then, that in the science of nature too the primary task must be to settle the things which concern the principles.
>
> The natural procedure is to go *from* things which are more intelligible and clearer to us *to* things which are clearer and more intelligible by nature. For what is intelligible to us is not the same as what is intelligible simply. Therefore, it is necessary to proceed in this manner: from what is less clear by nature, though more clear to us, to what is clearer and more intelligible by nature. The things which are first evident and clear to us are things which are rather jumbled up. Later from them their elements become intelligible and the

[2] A main stimulus for this approach is G. E. L. Owen's influential article '*Tithenai ta phainomena*' (1961, final version in his *Logic, Science and Dialectic*, London and Ithaca, 1986). A variant of the approach is developed in W. Wieland, *Die Aristotelische Physik* (Göttingen, 1962; 2nd edn., 1970); see e.g. pp. 216 ff. Owen's reading of Aristotle's method in the *Physics* is now standard: see, e.g., W. Charlton, *Aristotle's Physics I–II* (Oxford, 1970), pp. x–xii, 66, *et passim*; E. Hussey, *Aristotle's Physics III–IV* (Oxford, 1983), p. ix *et passim*; and T. H. Irwin, *Aristotle's First Principles* (Oxford, 1988), sections 34–7. For other references and general discussion of the issues see R. Bolton, 'Definition and Scientific Method in Aristotle's *Posterior Analytics* and *Generation of Animals*', in A. Gotthelf and J. Lennox (eds.), *Philosophical Issues in Aristotle's Biology* (Cambridge, 1987).

[3] For discussion of the issues raised by Book IV see Bolton, 'Definition and Scientific Method', 120–30, 164–6.

principles serve to separate these [jumbled up] things. That is why it is necessary to proceed from the comprehensive to specifics; for the whole is more intelligible in relation to perception and the comprehensive is a sort of whole since it embraces many things as parts.

This same thing happens also, in a certain way, to names in relation to their account. For a name also signifies a certain whole in an indefinite way; for instance the circle [is so signified], while the definition of it distinguishes [the whole] into its specifics. Also, children at first call all men father and all women mother; later they distinguish each of them. (184^a10–$^b14$)

This chapter begins with material which is quite familiar and intelligible to us from other passages in Aristotle, particularly from the early chapters of the *Posterior Analytics*. The first paragraph introduces the doctrine of the *Analytics* that unqualified scientific knowledge in any area requires an understanding of the most fundamental principles, principles which explain why the other facts hold (*An. Post.* I. 2; cf. II. 1–2). The second introduces the distinction made in the *Analytics* between the things which are more intelligible to and first known by us on some subject and the things which are first known and more intelligible by nature, where the latter things are the fundamental principles which are explanatory of the other things (I. 2, 71^b29–72^a5). However, Aristotle's further description of the process of scientific inquiry in this passage is not so clear in the light of other familiar passages. In particular, two features of his description of the starting-points for scientific inquiry are, notoriously, unusual and not easy to understand.

First, these things—the things which are first evident and clear to us—are described as things which are jumbled up or compounded (συγκεχυμένα 184^a22). Second, they are described as comprehensive (καθόλου $^a23$–4) and whole (ὅλον $^a24$) because at this level our understanding embraces many as yet undistinguished things as parts (μέρη $^a26$, with $^b2$ καὶ ἀδιορίστως and $^b5$). Aristotle offers two illustrations to make clearer what he means: (1) a name, he says, also signifies a certain whole or complex in an indeterminate way, that is, without clearly distinguishing what a proper definition would distinguish; and (2) little children also cannot distinguish things, e.g. their mother from other women, which they later can distinguish. However, it is not easy to understand just what the precise content and point of these examples is. They do indicate, clearly enough, that the initial facts which we grasp about some object of scientific inquiry do not enable us adequately to distinguish that object from some others which we may also be aware of. When we find the principles which explain these facts we can make the distinctions. This does not take us very far, however, in our attempt to understand in detail the *process* of inquiry itself or Aristotle's two unusual descriptions of its starting-points.

In addition to these two unusual descriptions of the starting-points, however, there is a final one which is again familiar. The starting-points are further described as what is more intelligible (or more knowable) in relation to perception. This too fits with what we find at the beginning of the *Posterior Analytics*, where Aristotle identifies what is first known by and more intelligible to us with what is 'closer to perception' (I. 2, 72^a1–5). In *Top.* VI. 4 what is more intelligible to us is similarly described as what is most accessible to perception (141^b5–14). However, in the passage in the *Analytics* such items are said to be most opposed to and furthest from what is καθόλου (i.e. universal). This flatly contradicts what is said in *Phys.* I. 1 unless, as most commentators reasonably conclude, the use of the term καθόλου in *Phys.* I. 1 is peculiar and not the typical use displayed, for instance, in the passage at the beginning of the *Analytics*. (I have followed the commentators here and used 'comprehensive' not 'universal' to translate καθόλου in *Phys.* I. 1.)

This standard move serves, however, only to deepen the problem of how to understand the peculiar descriptions in the *Physics*. But one close and clear point of correspondence between the *Physics* and the *Analytics* suggests a route to try in order to resolve this problem. Both works, as already noted, identify what is most knowable to us, i.e. the starting-points for scientific inquiry, with what is most knowable, or is knowable most directly, from perception. The famous final chapter of the *Posterior Analytics* (II. 19) attempts, in more detail than does the opening chapter of the *Physics*, to describe just what the process is by which we move from what we are most directly provided with by perception to a grasp of fundamental scientific principles. This connection suggests that attention to that chapter may be useful for understanding the opening chapter of the *Physics*.

III *Physics* I. 1 and *Posterior Analytics* II. 19

The aim of *An. Post.* II. 19 is the same as that of *Phys.* I. 1, to explain how we come to know first principles in science (99^b17–18). These are the first principles of demonstration (100^b8–14; cf. 99^b26–9). Two constraints are imposed by Aristotle at the beginning of his discussion in the chapter on this explanation. First, since knowledge of the principles is not innate the explanation must show how the principles are *learned* (99^b26–9). Second, since we learn them, the explanation of how they are learned must show how we infer the principles from what is previously known. This is what learning of the sort in question requires, as Aristotle has already said at the beginning of the *Posterior Analytics* ($^b28$–30; cf. I. 1, 71^a1 ff., $^b29$–33). The requirement of *previous* knowledge for learning, as the earlier passages make clear, is the requirement of this knowledge as a basis for inference to what is learned.[4] This shows that the account in II. 19 is not, as some have supposed, merely a genetic account of the psychological preconditions for the generation of knowledge of first principles. It is an account of the inferential process by which such knowledge is reached from previous knowledge.[5] This previous knowledge cannot, however, be more exact than the principles, that is, it cannot serve as a basis for *demonstrating* the principles, since they are immediate and indemonstrable. So it must be less exact (99^b30–4). Thus, we must have some power which enables us to acquire initial knowledge of what is less exact than the principles from which the principles can be inferred and thus learned ($^b30$–4). This power, Aristotle says, is perception ($^b34$–5). Perception, via memory, provides us with items of experience (ἐμπειρία) which serve as the starting-point for learning the principles.[6] Aristotle describes these items of experience, in the *Analytics*, in just the way in which he describes the starting-points for scientific inquiry in *Phys.* I. 1. They are the things 'from which' the fundamental principles are learned (100^a6–10, cf. 184^a22–3), without being more knowable or intelligible (by nature) than the principles (100^a10–11, cf. 184^a18–21). Thus they are acquired by perception, which is a procedure for acquiring information that is more knowable to us (100^a11; cf. 184^a24–5).

[4] J. Barnes (*Aristotle's Posterior Analytics* (Oxford 1975), 251) criticizes Aristotle for invoking his earlier doctrine from I. 1 about 'intellectual' or inferential learning. His reason is that Aristotle does not believe that the learning of the principles is intellectual, that is, 'knowledge of principles is not deduced knowledge'. But Aristotle's invocation of the earlier doctrine shows that he does believe that the learning of principles is inferential. The inferential process may not be deductive but, according to I. 1, not all intellectual learning is deductive. It may also be inductive (71^a5–9), which is what Aristotle has in mind here (100^b2–4).

[5] A main source for the alternate approach is W. D. Ross, *Aristotle's Prior and Posterior Analytics* (Oxford, 1949) 49. See, further, section VIII.

[6] 'From perception memory comes, from memory ... experience [comes], ... from experience ... a first principle of art or science [comes]' (100^a3–8).

These connections establish a very close correspondence between the two passages and lead directly to the question whether there are further connections which illuminate Aristotle's thought in the *Physics*. In particular, does anything in the *Analytics* correspond to Aristotle's descriptions of the starting-points for inquiry in the *Physics* as ὅλον καθόλου and as συγκεχυμένα which contain μέρη that are ἀδιόριστα? Interestingly enough, the first basic account that Aristotle offers of the starting-points, i.e. of 'experience', in *An. Post.* II. 19 is that it is 'the πᾶν καθόλου which is stabilized in the soul' (100ᵃ6–7). This is later described as a 'first καθόλου', which is a ἓν τῶν ἀδιαφόρων (100ᵃ15–16). This could easily mean that it is 'a unity composed of undifferentiated things', i.e. a συγκεχυμένον. This first καθόλου is then said to be succeeded as learning advances by things which are ἀμερῆ, 'without parts' (as well as being καθόλου, 100ᵇ1–2). This language, clearly, is extremely close to the language which Aristotle uses in *Phys.* I. 1 to describe the starting-points, and the further process of scientific inquiry leading to knowledge of principles. This establishes, I think, at least a prima-facie presumption that Aristotle is using the same language to say the same thing in the two passages where, after all, he is discussing the same topic.

This, however, runs counter to the mainstream of interpretation of the *Analytics* passage. The πᾶν καθόλου in the soul, which Aristotle identifies with 'experience' there, is standardly taken to be a 'universal' in the more familiar sense, that is, a universal proposition (or concept), since experience is described there as a grasp of 'the one apart from the many, what is one and the same in all those things' (100ᵃ7–8). But there are very serious difficulties with this interpretation. First of all, Aristotle says clearly that experience is constituted *simply* by 'many memories' of particular cases of the same thing grouped together (100ᵃ5–6), not by something which covers *all* cases of the same kind of thing. In the parallel passage in *Met. A*. 1, in fact, Aristotle denies that experience involves a grasp of a 'universal' of this latter sort. His example there of something one knows from experience is that: 'In many individual cases, namely the cases of Socrates, Callias, etc., this treatment was beneficial for this sickness.' But, Aristotle expressly says, this does *not* involve a 'universal judgement' that in 'all cases distinguished *as one in kind*' (κατ' εἶδος ἓν ἀφορισθεῖσι) this treatment was beneficial for this sickness (981ᵃ5–12). This fits perfectly with what Aristotle says in *Physics* I. 1 where also, at the initial stage of our understanding of certain things, those things are undifferentiated (ἀδιόριστα, 184^b2) in kind from other related ones.

This does not conflict, moreover, with the description of experience in the *Analytics* as a grasp of what is 'the same in all *those things*', that is, all of the 'many things' remembered (100^a7–8). To take the example from the *Metaphysics*, one is aware there, just as the *Analytics* says, of one thing which is 'the same in all' the particular cases, i.e. the cases of a certain illness which we recall, namely that this treatment works as a cure. We just don't know how to mark off definitively these cases as a *kind* (εἶδος) distinct from other similar ones. So it is quite easy to make the initial treatment of experience in the *Analytics* internally consistent, and consistent with the treatment in the *Metaphysics*, if the πᾶν καθόλου which we grasp in experience is the ὅλον καθόλου of *Phys.* I. 1, rather than a strict universal.

IV Further Support for a Unified Account

There are equally good reasons for treating the second discussion of 'experience' in the *Analytics* passage (100^a14 ff.) along the same lines. Aristotle goes on to redescribe the 'first καθόλου' which is present in the soul as a ἓν τῶν ἀδιαφόρων (100^a15–16). Here again the καθόλου is normally understood as a standard universal, since Aristotle says of the perceptual state in which it is grasped that it is, for example, 'of man, not of the man Callias' (100^b1). But this, again, faces serious difficulties. Earlier in *An. Post.* I. 31, Aristotle makes the same point that 'perception is of a certain sort of thing not of the individual'. But he goes on immediately to add that, nevertheless, one cannot perceive 'what is universal and holds in every case' of a given kind. This sort of universal is grasped '*from* our seeing', that is, by reasoning from what we 'often observe' or what we 'see in each individual case'—in other words, by reasoning from experience. It is the object of scientific knowledge (ἐπιστήμη) not perception or experience (87^b28–31, 88^a2–17). So the καθόλου of II. 19 (100^a17) which *is* the object of experience, or of what we often perceive in many individual cases without further reasoning, cannot be a universal of the sort which 'holds in every case' of a given kind.

Furthermore, if we understand the object of experience as a universal of that sort it is hard to make sense of Aristotle's description of it as a ἓν τῶν ἀδιαφόρων (100^a15–16). This cannot mean that grasp of it involves an awareness of just *one* of the many particulars referred to earlier, on which experience is based (100^a3–6), since Aristotle explicitly says that experience has to be based on many of these. So commentators typically interpret it to mean one of the *infimae species*, which are ἀδιάφορα in the sense of not being further sub-dividable. This, then, is understood to commit Aristotle to a certain view of concept acquisition, according to which we start from particulars then through abstraction we grasp, first, *infimae species*, then progressively higher genera. But in this case the chapter turns out to be about how we acquire general or abstract concepts, not about what it is supposed to be about, namely how we acquire knowledge of *all* first principles (99^b17–18). Acquiring general concepts may have something to do with coming to know adequate scientific definitions, though they are hardly the same thing.[7] But the acquiring of general concepts is not at all the same thing as coming to know the first principles of the types other than definitions, namely the axioms and hypotheses (see *An. Post.* I. 10).

In addition, even if we can somehow ignore the problem about the other types of principles (as we will see below we cannot), and even if we can somehow treat the acquiring of general concepts and the acquiring of definitions as much the same thing, this interpretation still commits Aristotle to the curious idea that we acquire genuine first principles, that is, scientific knowledge, of species before we acquire scientific knowledge of the genera under which they fall. This conflicts with his standard doctrine that a species is defined, and known *as* a complex of genus and differentia.[8] One might respond to this (though the main tradition would not) by saying that Aristotle is talking here simply about rough initial definitions or concepts of species and genera, not the adequate definitions or concepts where knowledge of a species presupposes knowledge of its genus. But there is no good reason to believe he thought that we always acquire rough definitions or concepts of more specific things before we acquire rough definitions or concepts of more generic things. His examples earlier in the *Analytics* of our initial or rough definitions and concepts often include, in the *definiens*, the genera of the things defined, for instance the initial

[7] It is clear from Aristotle's earlier discussion in *An. Post.* II. 8–10 that in his view one can have a general concept adequately formulated in a *nominal* definition, e.g. of man, without having an adequate scientific definition.

[8] See *Top.* VI. 4, 141^b29–34. This picture of definition by division is present also in the *Posterior Analytics* (II. 13, 97^a23–8).

rough definition of thunder as a certain noise in the clouds and of man as a certain kind of animal (see II. 8).

These difficulties are avoided if the ἓν τῶν ἀδιαφόρων is understood, in the light of *Phys.* I. 1, as a unity composed of (as yet) undifferentiated things. That is, the content of perception, at the level where we have only experience, is a καθόλου in the sense of *Phys.* I. 1. This content is still, 'for example, man, not the man Callias' (100^a17–18) because, as Aristotle has already said, perception has for its content 'a certain sort of thing not the individual' (87^b28–31). To take the case from *Met. A.* 1, when we are aware through perception of many cases where a certain treatment cured fever, we are aware of a certain sort of thing; that is, we know of a certain treatment (in many cases) curing fever, and we are not simply aware of the individual cases. However, we cannot distinguish the cases where this works 'as a kind'. So too in the example from the *Analytics* we are aware of (many cases of) a certain sort of animal, Aristotle says, and thus of man. We just cannot, at this stage, precisely delimit the kind. Further attention to what we experience enables us to move from this grasp of man, i.e. of a certain sort of animal, to the point where we can clearly distinguish the kind, animal, and within it man and other distinct species. At this point, Aristotle says, we have καθόλου which are ἀμερῆ, that is, not composed of as yet undifferentiated parts and, thus, we have καθόλου which are now universals in the familiar sense. They cover all cases of a genuine distinct kind. (100^b1–3: 'In turn, among these [undifferentiated things] a stand is made to the point where καθόλου without [further undifferentiated] parts stand, e.g. [we go from] this sort of animal to the point where [the kind] animal stands [distinct] and within it [the sub-kind man] equally.')

The comparison, then, between the parallel passages in the *Physics* and the *Analytics*, particularly in light of *Met. A.* 1, supports the suggestion that Aristotle uses the same unusual language in the two passages to say the same thing. This has its pay-off, as we have just seen, in enabling us to see how to avoid certain inconsistencies and solve certain puzzles in the interpretation of the *Analytics*. Of more interest for present purposes is the pay-off for the interpretation of the *Physics*.

V *Posterior Analytics* II. 8 and the Puzzles in *Physics* I. 1

The use of the *Analytics* to understand the puzzling details of *Phys.* I. 1 is facilitated if we remember an important fact about the *Analytics* which is often overlooked. The final chapter of the *Posterior Analytics* is not the first or only place in that work where Aristotle describes in some detail how we come to know first principles. He concludes *An. Post.* II. 8 with the remark, 'We have now said how what a thing is [i.e. how a basic definition] is grasped and comes to be known' (93^b15–16). So Aristotle is talking about the same subject as in II. 19, namely, 'how the first principles come to be known' (99^b17–18), with special attention to those first principles which are definitions. This indicates that we should use II. 8 together with II. 19 to understand *Phys.* I. 1.

As we might expect, the discussion in II. 8 follows the same pattern as in the later chapter. Aristotle supposes that we start from some initial items of information about some subject and come to know the principles as the things which account for our initial information. The discussion, again, is not a piece of genetic psychology but a description of the scientist's actual method. In fact, one of the examples which Aristotle mentions in II. 8 of our initial knowledge on the basis of which a basic definition becomes known is just the one he uses prominently in II. 19, namely our initial grasp of man as 'a certain kind of animal' (93^a23–4; cf. 100^b2–3). Unfortunately Aristotle does not say more about that case in II. 8, but his other examples there are helpful. We are first familiar with a lunar eclipse, for instance, as 'a certain loss of light' by the moon (93^a23). At this stage, 'from which' we begin to learn the first principle which gives the fundamental definition of eclipse, what we are aware of, as II. 19 makes clear, is 'many cases' of loss of light by the moon (cf. I. 31, 88^a2–7). It is easy to see, moreover, how we might at this stage be only aware of a καθόλου or a συγκεχυμένον containing as yet undistinguished things, as *Phys.* I. 1 says we are. For there are other cases of loss of light by the moon which we may well not be able initially to distinguish in kind from the cases of eclipse that we have experienced. For instance, the moon loses its light when it wanes and we may not know initially whether the waning of the moon is a different kind of thing from an eclipse, or perhaps the same kind of thing. In this case the common thing which we grasp the genuine cases of eclipse *as*—namely a loss of light—contains many as yet undistinguished types of things. (In early Greek astronomy some did treat eclipses and the phases of the moon as the same kind of thing.[9]) The principles, however, as the *Physics* says, 'serve to distinguish these [previously undifferentiated] things' (184^a23). When we learn, for example, that the basic definition and cause of all the repeated familiar cases of eclipse which we grasp in experience is the blockage by the earth of the moon's light, then we can distinguish eclipses universally 'as a kind', since the causes of the other phenomena are not the same as this (See *An. Post.* I. 31, 87^b39–88^a6). This use of *An. Post.* II. 8 together with II. 19 enables us to see then, in further detail, what Aristotle has in mind in his peculiar descriptions of the starting-points, and of the causes and principles which serve to differentiate their 'jumbled up' elements, in *Phys.* I. 1.

VI Some Conclusions about Aristotle's Method in Natural Science

The correspondence between the *Physics* and the *Analytics* enables us now to draw a general conclusion about the method which Aristotle recommends in the *Physics* for natural science, namely that this method is exactly the one he describes in the *Posterior Analytics*. It is an immediate consequence of this that Aristotle's method in natural science is not dialectic; it is not the method which he describes in the *Topics*, as opposed to the one which he lays out, particularly in the second book, in the *Posterior Analytics*. The most obvious point of difference has to do with the difference in the starting-points. Dialectic starts and reasons exclusively from ἔνδοξα, that is, from the standing convictions either of everyone, or of most people, or of the wise—either all of the wise, or most of them, or the most famous (*Top.* I. 1, 100^a29 ff.). Scientific inquiry on the other hand starts from experience; and an item of experience as Aristotle describes it in *An. Post.* II. 19, and elsewhere, need not be an ἔνδοξον—a standing conviction of the many or the wise. Items of experience may be ἔνδοξα, but they need not be.

[9] See J. Burnet, *Early Greek Philosophy*, 4th edn. (London, 1930), 78, and references given there. There is another clear case of this sort, which Aristotle introduces himself, in the *Generation of Animals*. He points out there that from our initial understanding of *sperma*, as an originating source of animal generation, it is not clear to us whether the *menses* is a type of *sperma* or not (721^a30–$^b6$ with 724^a14–20 and 726^a28–$^b1$).

What is important for scientific starting-points is that they come from proper experience, not who accepts them.[10]

Since it seems quite clear in *Phys.* I. 1, even independently of the problems of interpretation which we have discussed, that the starting-point for scientific inquiry is, as Aristotle says, simply what is 'more intelligible to (or more knowable by) perception' (184^a24–5), one might well ask why interpreters have been so quick to suppose that Aristotle's recommended method in the *Physics* is dialectic. The evidence typically cited to support this claim from *Phys.* I. 1 is Aristotle's remark that scientific inquiry moves from what is more intelligible (or knowable) to us to what is more intelligible by nature.[11] It is widely supposed that this is a description of dialectic. But where does Aristotle ever describe dialectic in this way? The closest that he comes to this is in *Top.* VIII. 5 where he lays it down as a rule of dialectical reasoning that the premisses must be more intelligible or better known than the conclusion (159^b8–9). It is plausible enough that he means by this that the premisses must be more intelligible to us than the conclusion, which is then less intelligible to us than the premisses.[12]

But, clearly, just because an argument or an inquiry proceeds from what is more intelligible to us to what is less intelligible to us does not suffice to show that it proceeds from what is more intelligible to us to what is more intelligible by nature. So Aristotle's description of scientific inquiry in *Phys.* I. 1 is not as such a description of dialectic. Nor can we say that what he describes there must at least count as one type of dialectical reasoning, since it is at least one type of reasoning from what is more intelligible to us to what is less intelligible to us. Nothing in Aristotle's discussions of the method of dialectic marks out for us any *type* of dialectical reasoning whose rules guarantee or make it likely that the conclusion of some piece of reasoning is not only less intelligible to us than the premisses but also more intelligible by nature. What guarantees this result according to *Phys.* I. 1, and according to the corresponding material in the *Analytics*, is that the conclusion of our reasoning and our inquiry gives us a principle which *explains* (and gives us a firm delineation of) the perceptual phenomena

[10] For an elaboration of this point see Bolton, 'Definition and Scientific Method', 120–30.

[11] See Irwin, *First Principles*, 66.

[12] For discussion of this see R. Bolton, 'The Epistemological Basis of Aristotelian Dialectic', in D. Devereux and P. Pellegrin (eds.), *Biologie, logique et metaphysique chez Aristote* (Paris, 1990), section 7.

which we use to reach it. But no rule of general dialectic or of any type of dialectic which Aristotle discusses is designed to guarantee conclusions of this sort. So if dialectic does reach conclusions of this type it is accidental and not due to the method of dialectic itself.[13]

VII Further Discussion of Method in *Physics* I. 2

If the method recommended for inquiry in natural science in *Phys.* I. 1 is not dialectic but the method of the *Analytics*, we should expect this to be confirmed by Aristotle's later discussion, particularly in the later chapters of Book I where his preaching in chapter 1 ought most of all to be guiding his practice.[14] Since these chapters have seemed to many interpreters to present us with a paradigm of dialectical inquiry, they are all the more in need of attention here. In these chapters Aristotle is engaged in a search for at least some of the principles of nature, just as he indicates he will be in chapter 1. He begins, in chapter 2, by raising the question of how many principles there are and by pointing out that various positions on this topic have been taken by his predecessors. Before pursuing this question further, however, Aristotle stops to consider the position of the Eleatics, since according to their view there are *no* principles of nature because there is no natural world—no world of naturally changing things—at all. He prefaces his discussion of their view with some new methodological remarks:

> To inquire whether what exists is one and unchanging is not an inquiry in the science of nature. For just as there is no further proof for the geometer to give to someone who has rejected his principles, but this proof will proceed from some different type of knowledge or from knowledge which is common to everyone, so also neither is there [a proof to give to one who denies principles] for anyone [whose inquiry is] governed by principles . . .
>
> To inquire whether what exists is one in this [alternate, non-scientific] way is like discussing [dialectically] any other position put forward for the sake of discussion, such as Heraclitus' thesis . . . or it is like exposing as fallacious an eristic argument—which in fact the arguments of both Parmenides and

[13] It will become clear below that the type of reasoning which Aristotle has in mind in *Phys.* I. 1 and *An. Post.* II. 19 cannot be dialectical, even accidentally.

[14] It ought to guide his practice throughout his studies in natural science (184ᵃ14–16). Aristotle conceives of scientific inquiry quite generally as the search for principles. (See *An. Post.* II. 1–2, *Met. A.* 1–2) And this chapter prescribes the method for this search in natural science. But Aristotle is least likely to lapse from his intended method in the immediately succeeding discussion.

Melissus are ... For our part [as scientists], however, we proceed through the hypothesis that all, or some, of the things which exist by nature change, and this [hypothesis] is made evident by induction.

Nor is it proper [for the scientist] to expose every error, but only as many as someone falsely demonstrates from his principles; other errors, not. For example, it is the geometer's task to expose as fallacious the procedure of squaring [the circle] by use of segments, but not the geometer's task to expose Antiphon's procedure. Nevertheless, though these people [the Eleatics] are not doing inquiry in the science of nature, they do incidentally discuss problems in natural science. So perhaps it is proper to engage in a bit of [dialectical] discussion with them. For this investigation has use for [natural] philosophy. (184^b25–185^a20)

There are a number of points which Aristotle makes in this important passage which are directly relevant to our topic. First, Aristotle explicitly says that the study and the refutation of the position of the Eleatics is not a job for someone engaged in scientific inquiry in a natural science. Rather, it is a job for someone who argues from 'knowledge which is common to everyone'.[15] What does Aristotle mean by this? His remarks on this point are strikingly similar in every detail to remarks which he makes in *Soph. El.* 9 and 11. There Aristotle also distinguishes argument, and in particular refutation (ἔλεγχος, λύσις), which is the proper job of the scientist, from argument and refutation based on 'the common things' (κοινά), which are matters of common knowledge and not known only by specialized scientists in a given area (170^a36–$^b11$; 171^b4–7; 172^a21 ff.). He even uses the same main example there as he uses here. He contrasts the refutation of the *scientific* attempt to square the circle through segments, which is proper to the geometer and thus employs *demonstration*, with the

[15] The alternative (185^a2), that the principles are demonstrated by some special science, super-ordinate to natural science, is not relevant here. So the refutation of the Eleatics must proceed from πασῶν κοινῆς (ἐπιστήμης), 185^a2–3. This phrase might be translated 'knowledge (or science) common to all the sciences'. (So Irwin, *First Principles*, 67 and n. 52, following most commentators.) Ross (*Aristotle's Physics* (Oxford, 1936), 337, 461) takes the phrase, so understood, to mean 'universal science', i.e. metaphysics. But metaphysics is universal because it studies what is primary, not because its contents (excepting the common axioms and the like) are in general common to all sciences, or to everything (*Met. E.* 1, 1026^a29–32). Some, including Irwin, take the phrase to refer to dialectic, since dialectic is an art useful for all sciences (*Top.* 1. 2). But dialectic also does not typically prove things 'from *knowledge*' common to all sciences (or to everything). Aristotle does not describe metaphysical knowledge (or science) or dialectic as 'knowledge common to' all sciences or to everything. He does, however, describe dialectic, or a certain form of dialectic, as based on knowledge which is common *to all people* in contrast to knowledge which is based on mastery of a science, as we shall directly see.

refutation of Antiphon's procedure (171^b34–172^b4). The latter mode of argument is based on the common things in the relevant area which 'everyone, even the unlearned ... know themselves no less than the scientists' (172^a30–4), and is the province of dialectic, Aristotle says, and in particular of the special type of dialectic which he calls peirastic (172^a27–36).[16]

So the first point that Aristotle is making here in *Phys.* I. 2 is that the natural scientist cannot use a scientific, that is a demonstrative, argument to refute someone who denies that the natural world of changing things exists. In natural science it is an indemonstrable first principle that the natural world of changing things exists (cf. *Met. E.* 1, 1025^b7–28, quoted below). This principle is the basis for all demonstration in natural science, so no demonstration of it, in order to refute scientifically someone who denies it, is possible. One can only refute this denial dialectically, or peirastically.

VIII How does the Scientist know Principles?

What, however, is the epistemic status of the principle, the hypothesis that the natural world exists, for the scientist? If the dialectical procedures for dealing with the principle lie outside the bounds of strict scientific inquiry, how does the scientist as such come to know the principle? Aristotle does not claim or imply, as some have argued, that this principle is, for the scientist, self-evident or an object of epistemically 'immediate apprehension'.[17] Such a claim would in any case, of course, be quite absurd. How could it be self-evident that a natural world of changing things exists? Aristotle does not make this claim; rather he says quite specifically that 'for our part', i.e. for the natural scientist, this first principle becomes 'evident through induction' (185^a13–14). This is, of course, just what we would expect, given what Aristotle says about how we come to know first principles in *Phys.* I. 1, particularly in light of the parallel discussion in *An. Post.* II. 19. He specifically claims in the latter passage that we come to know *all* first

[16] In a parallel passage in *Rhet.* I. 1, Aristotle uses the same language and contrast which he employs in *Phys.* I. 2 (185^a2–3) and describes the basis for certain dialectical arguments as 'what it is common to everyone to know (κοινὰ ἁπάντων γνωρίζειν) and does not require any special scientific knowledge' (1354^a1–3). For further discussion of this matter see Bolton, 'Basis of Dialectic', section 10.

[17] See Irwin, *First Principles*, 68, 118 for a recent statement of this traditional view, taken also by Ross. See also n. 19.

principles 'by induction', that is, by inductive argument based on information provided by perception (100^b3–5). He repeats this claim in *Met. E.* 1 in a passage quite similar in its language to our passage in the *Physics*.

[The sciences] do not provide any proof (λόγος) of *what* [the object of their study] is. But starting from [an account of] what it is which some sciences make evident by perception and some set out as an hypothesis, they in this way demonstrate, more or less strictly, the things which belong in their own right to the kind with which they deal. Thus it is quite clear that there is no demonstration of the essence or of what something is, by such induction, but some other [non-demonstrative] way of making it evident. They similarly also say nothing [by way of demonstrative proof] as to whether the kind with which they deal exists or not, because it is the task of the same kind of intellectual reasoning [i.e. induction] to make it evident what something is and whether it exists. (1025^b10–18)

This passage says quite generally that the first principles of the special sciences, both definitions and existential principles, are typically made evident, to the scientist, by a procedure which treats them as hypotheses which are inductively confirmed (cf. *EE* 1218^b37–1219^a5). Aristotle mentions perception here as an alternate basis for coming to know principles, in addition to the hypothetical procedure which he mentions in *Phys.* I. 2 and elsewhere emphasizes (*An. Pr.* I. 30, 46^a17–27, *An.* I. 1, 402^b21–5). But in the *Posterior Analytics* he makes it clear that induction is the process at work in both kinds of cases (II. 19, 100^b3–5 with II. 2 90^a24–30 and I. 31, 88^a11–17). Aristotle's general name for the processes by which we learn in science other than by demonstration is 'induction' (I. 18).[18]

It is sometimes suggested that induction for Aristotle, as he understands it in the *Analytics* and elsewhere, is not a procedure for proof or justification at all but rather a procedure for pointing to what is otherwise evident, e.g. by being self-evident.[19] However, Aristotle standardly uses the term 'induction' for a type of argument (*Top.* I. 12, 105^a10–16), and he explicitly says in the *Analytics* that inductive arguments '*establish* (δεικνύναι) the universal through the fact that the particular is evident' (I. 1, 71^a8–9, cf. II. 7, 92^a34–8).[20] This shows clearly that

[18] The same point is made in *EN* VI. 3, 1139^b26–31.

[19] A main source for this approach is Ross, *Aristotle's Analytics*, 47–51. For a review of the more recent literature see R. McKirahan, 'Aristotelian Epagoge in *Prior Analytics* II. 21 and *Posterior Analytics* I. 1', *Journal of the History of Philosophy* 21 (1983).

[20] Cf. *EN* I. 7 (1098^b1–4) where Aristotle also mentions induction as a procedure by which principles are established. Whether, in contexts which have to do with things

Aristotle does not conceive of inductive reasoning as a procedure for discovery only and not justification. This is why inductive reasoning can serve as a procedure for learning, as Aristotle supposes it will in *An. Post.* I. 1 and II. 19. For all learning through reasoning involves coming to know something by inferring it from what is previously known which, thereby, provides an adequate justification for what is inferred (I. 1, 71^a1–17 with I. 2, 72^a25–32; cf. *EN.* VI. 3, 1139^b26–31).

IX Why the Scientist's Justification is not Dialectical

Aristotle's account of principles as reached by induction, in *Met. E.* 1, corresponds in very close detail to what he says in *Phys.* I. 2 about how natural science reaches its fundamental existential principle—that a world of naturally changing things exists (185^a12–14, cf. 1025^b18–21). However, in the *Physics*, Aristotle contrasts this procedure, used by the inquiring scientist, with the non-scientific, dialectical procedure to be used for refuting the Eleatics. This supports the result that the scientist's own procedure for confirming his principles is not dialectic. But it is worth investigating in further detail just how and why Aristotle draws the contrast he does, between scientific and dialectical procedures for dealing with principles. If there is for the scientist a sound inductive proof that a natural world of changing things exists, why does Aristotle not simply draw on that proof to refute the Eleatics rather than assigning that proof to the scientist and using other, dialectical arguments to refute the Eleatics?

As the beginning of an answer to this question we may note that an inductive proof which starts from premisses such as 'Birds fly' and 'Trees grow in the summer' and concludes that, therefore, 'A natural world of changing things exists' will not be convincing to any Eleatic, since he will be no more ready to accept the premisses of such a proof than he was to accept the conclusion. This cannot be the whole answer by itself, however, since there is also no reason to think that any Eleatic will accept the dialectical premisses which Aristotle in fact goes on to use. In general, we need to remember, it is not a requirement for a good dialectical or peirastic argument that an opponent be convinced by it or accept the premisses. What is required is that the conclusion follows from premisses of the proper sort. The premisses must

other than the learning of principles, Aristotle thinks of induction differently is beyond the scope of this paper.

be appropriate ἔνδοξα which, in the case of peirastic, are matters of common knowledge. An opponent may balk and not accept these premisses but if he does he is simply refusing to do proper dialectic. There is nothing dialectically wrong with the argument itself (*Top.* VIII. 11, 161^a16–24, $^b1$–6).

To see the full answer to our question it is necessary to remember certain crucial technical details in Aristotle's theory of refutation, and in particular of dialectical refutation, since Aristotle clearly intends his refutation here to be dialectical. To begin with, an inductive argument simply does not count as a refutation for Aristotle. A refutation is by definition a type of syllogistic, and thus deductive, argument (*Soph. El.* 165^a2–3, 167^a23–6, 171^a2–5). The premisses for a dialectical syllogistic refutation can be established inductively (*Top.* VIII. 1, 155^b17–25, 34–156^a7). But the premisses in a syllogistic argument must, of course, be different from the conclusion; and in a refutation the premisses cannot include or even immediately imply the conclusion since that would beg the question (167^a23–6). So the inductively established claim that a natural world of changing things exists cannot be used as a premiss for a syllogism that refutes the (direct or immediately implied) denial of this, because the conclusion of such a purported refutation would already be (or be immediately implied by) one of the premisses. So no inductive argument for the principle that this natural world of changing things exists can be used to refute someone who denies this.

This result can be stated more generally since every dialectical proof, like every refutation, is a syllogistic proof (*Top.* I. 1, 100^a18–23, cf. *Soph. El.* 34, 183^a38 ff.).[21] So, quite generally, the inductive procedures which the scientist uses to establish his own proper principles cannot be used to prove dialectically those principles, since what these inductive procedures establish—namely the principles themselves—cannot be used to prove syllogistically or prove without begging the question, those very principles. If there is to be any syllogistic proof of those principles, and thereby refutation of those who deny them, it must be from premisses established by other procedures than the very ones which the scientist uses to establish inductively those very principles. Dialectic, as itself a method for syllogistic proof, can sometimes provide us with such procedures, Aristotle implies (185^a2–3). But this shows again, conclusively I think, that whatever procedures dialectic

[21] Dialectic uses inductive arguments, as Aristotle says in *Top.* I. 12, but as he makes clear in the passages in VIII. 1 referred to above, this use is to obtain premisses for a syllogistic argument.

itself may draw on to prove scientific first principles they are not and do not include the very inductive procedures for establishing the principles which Aristotle ascribes to the scientist. It would be a mistake to think that this result, even though conclusive, depends merely on certain arcane technical requirements of dialectic which *we* might not think it proper to invoke in describing refutation in general. As noted above, peirastic dialectic, which is Aristotle's instrument of skilled refutation, draws for its premisses on common knowledge. Among other things this means that skilled refutation aims, quite reasonably, to draw as much as possible on common ground to the disputants on a given point. To take the case relevant here, it aims to draw on common ground both to those who are right about a given scientific principle and to those who are wrong (*Soph. El.* 11, $172^{a}21$–$^{b}1$, cf. *Rhet.* I. 1, $1355^{a}24$–9 with *Top.* I. 2, $101^{a}30$–4). But, of course, in any given case one cannot beg the question against an opponent, so one cannot use any common knowledge which would involve this. One must, rather, restrict oneself to other common knowledge. For the reasons just indicated, reliance on the items of common knowledge used in an inductive proof of the principle that a natural world of changing things exists *would* beg the question against the general denial of this. So it is quite reasonable that the technical requirements of dialectic should be geared to rule this out and to require non-inductive proof, if any, of such a principle. An opponent may, of course, in a given case, refuse to accept even the common ground provided by the relevant non-question begging common knowledge. But in this event the opponent eliminates the possibility of even the most minimal form of non-specialized rational discussion of disagreements over principles and it is, once again, reasonable for the technical requirements of dialectic to be geared to reflect this.[22]

X How are Principles 'Known Through Themselves'?

This puts us now in a position to understand a further passage in *Phys.* II. 1, in which Aristotle returns to the question of how the basic principle, that the world of naturally changing things exists, comes to be known by the scientist. He says there:

That the natural world [of changing things] exists it would be absurd to try to establish (δεικνύναι). For it is obvious (φανερόν) that there are many such

[22] I discuss more generally the issues involved here in 'Aristotle's Conception of Metaphysics as a Science', (forthcoming).

things. To establish what is obvious through what is not is one expression of the inability to distinguish what is knowable through itself from what is not knowable through itself. That it is possible to be in this state is quite clear. For someone blind from birth might reason syllogistically to conclusions about colours. So necessarily, in such cases, the proof (λόγος) is about the [colour] names only, and nothing [about colours] is understood [through the proof]. (193ᵃ1–9)

This is a passage to which some have appealed to show that scientific first principles, for Aristotle, are self-evident. But the passage does not warrant such a conclusion. Aristotle is speaking here about the requirements of proof that are incumbent on the scientist. As earlier in *Phys.* I. 2 where he says there is no *proof* (λόγος) for the scientist that there are naturally changing things, when he says here that this cannot be established, he means that there is no strict scientific proof, no demonstrative syllogistic proof through a middle term for this principle. His further remarks make this clear. The reason why scientific principles cannot be established is that they are knowable 'through themselves'. This is a point which Aristotle emphasizes elsewhere, notably in the first chapter of the *Topics* (100ᵃ27 ff.). He there explains why the principles are knowable through themselves: 'For in the case of the first principles of the sciences it is not necessary *to inquire why* (διὰ τί) they are so' (100ᵇ19–20). In *An. Post.* II. 2 Aristotle explains what he means by 'inquiring why' (διὰ τί). This is, he says, inquiring for a 'middle term' which makes the thing in question to be so, and through which it can be syllogistically demonstrated that it is so (90ᵃ1). Thus, Aristotle says here in the *Physics*, one would only try to establish, i.e. to demonstrate, principles if one was ignorant of this point and could not distinguish what is knowable through itself from what is not. That Aristotle means to be ruling out here syllogistic proof, through a middle term, of the principle that nature exists, is made further apparent by the comparison with the congenitally blind person. What this person mistakenly tries to do, is to 'reason syllogistically' to conclusions about colours which are, for sighted people, known through themselves (193ᵃ7).

This passage further suggests that Aristotle thinks not only that there is no demonstration that nature exists but that there is no no adequate syllogistic proof of this, not even a dialectical proof. He could argue this on the ground that a dialectical proof has to have premisses which are more of the nature of ἔνδοξα and more intelligible to us, i.e. more obvious to us, than the conclusion,[23] and nothing from which it could be syllogistically derived that the world of naturally changing things exists is more obvious to us than that fact. But none of this would show *how* this fact came to be so obvious to us, or to the scientist, and in particular it would be perfectly compatible with Aristotle's claim in *Phys.* I. 2 and elsewhere that for the scientist this becomes obvious through inductive proof. As we have just seen, to say that a principle is known '*through* itself' and not through something else is not, for Aristotle, to say that it is known *in* itself or by epistemically direct apprehension. It is to say simply that it is known without use of any mediation of a middle term such as would be used in a syllogistic proof, whether dialectical or demonstrative. This does not rule out inductive proof, since inductive proof does not rely on the mediation of any such middle term. So propositions known inductively *are* known '*through* themselves'; the truth of the connection between the terms of such propositions is not grasped through the mediation of any other term. The basis which Aristotle offers in *Phys.* II. 1 for the scientist's knowledge that nature—the world of naturally changing things—exists is that 'it is obvious (φανερόν) that there are many such things' (193^{a}3–4). This is just the sort of base which Aristotle requires for an inductive argument, since such arguments establish their conclusions by virtue of the fact that, in each of the cases which make up the base, 'the particular is evident' (δῆλον, *An. Post.* I. 1, 71^{a}8–9).

That the scientist's proof of his principles is inductive non-dialectical proof does not mean, of course, that dialectic is of no use in scientific inquiry. Aristotle himself says here that the dialectical investigation of the views of the Eleatics is of some use for natural philosophy. This fits with his famous remark in *Top.* I. 2 that it is necessary (he does not say it is sufficient) to use dialectic in reaching the first principles of the sciences (101^{b}1–2). There he assigns this role to dialectic, owing to its power as a method of testing or examining (ἐξεταστική 101^{b}3). This is what he goes on to employ dialectic to do in his review of the Eleatics. He first, in chapter 2, tests and refutes their claim that what exists is one and unchanging and then, in chapter 3, tests and shows their argument for the claim to be fallacious (see I. 3, 186^{a}4–6). It would be of interest to study these arguments in detail to see just how Aristotle's discussion is of use for the natural philosopher in his search for principles. That would require a good deal more

[23] See *Top.* VIII. 5 and Bolton, 'Basis of Dialectic', section 7.

space, however, than is available here. What is clear enough is that those dialectical procedures must somehow be in aid of the general inductive procedures which constitute the proper scientific method for reaching the first principles, since that is the way the scientist reaches them.

XI Aristotle's Non-Dialectical Treatment of the Principles

Of more direct interest to us here, in any case, is the question whether, in addition to his dialectical discussion, Aristotle produces, or at least indicates how he would produce, an inductive non-dialectical justification of the particular principles which he uncovers by the end of *Phys.* I. Those principles, of course, are form, privation, and substratum. Before we consider how Aristotle justifies the positing of these things as principles, it will be useful to say something about the sense in which they *are* principles.

Unlike the examples of principles which Aristotle uses in *An. Post.* II these principles are not, and are not introduced in, definitions. But, as we have seen from his remarks there and in *Met. E.* 1, Aristotle thinks that the same type of inductive procedure which is used to reach and to validate those principles that are definitions is also used in the case of those principles, called 'hypotheses' in the *Posterior Analytics*, that are existential in character. In the *Analytics* Aristotle uses the term πρῶτα, the primary things, to refer to the entities introduced in such principles (I. 10, 76^a32) and he uses the same term in *Phys.* I. 7 for the principles introduced in *Phys.* I (190^b18). An example from the *Posterior Analytics* is instructive here. The basic existential principle of geometry is that two-dimensional magnitude exists (I. 10, 76^a31–6). This corresponds to the basic principle of natural science that nature, i.e. the natural world of changing things, exists (cf. *Phys.* II. 1, 193^a1 ff.). But in the case of geometry it is possible to spell this out more fully. Aristotle does so when he indicates that the fundamental existential principle of geometry is that points and lines exist (76^b3–6). This fills out the content of the principle that two-dimensional magnitude exists because these are the basic objects which make all geometrical magnitude possible. All other geometrical objects are constructible out of points and lines, but the latter are not constructible out of, or otherwise reducible to, each other or to any more fundamental entities. Just so, in natural science the principle that nature, the realm of naturally changing things, exists can be more fully spelled out as the principle that form, privation, and substratum exist. This principle is formulated at a higher level of generality than the principle that points and lines exist because the type of form or substratum, for instance, which makes natural change possible in one natural science, say meteorology, will differ from the type which makes natural change possible in another, say biology (cf. I. 7, 189ᵇ30–2). Nevertheless, in each case there will be something which plays each of these roles and which is not reducible to anything more fundamental. In *Phys.* I. 5 Aristotle makes just this point when he says that 'the principles can come neither from each other nor from anything else, and everything else must come from them' (188ᵃ27–8). This is exactly what he would say about points and lines in geometry and it is a clear indication that Aristotle is thinking of the principles of natural science which he discusses in *Phys.* I in the way he would think of points and lines in geometry. How then does Aristotle establish the fundamental principle that form, privation, and substratum exist? Does he do so by inductive argument as we have been led to expect he will?

XII Aristotle's Proof that the Principles are Contraries: *Physics* I. 5

After the dialectical refutation of the Eleatics in *Phys.* I. 2–3, which Aristotle has told us in advance will not establish any principles for the scientist, he proceeds in I. 4 to review the opinions of his predecessors as to what the principles of natural change are. He finds that they 'all in some way make contraries principles' (I. 5, 188ᵃ26). The qualification that they do this only 'in some way' is perhaps important. But, in any case, Aristotle says he does not find in his predecessors any proof or argument (λόγος, 188ᵇ29) for this sort of conclusion, so he proposes to offer a proof or argument (λόγος, 188ᵃ31) himself. The specific conclusion which he argues for in his proof is that all natural change is change in which one thing comes to be from some contrary (either from the exact contrary, as white to black, or from some intermediate contrary, as white to grey 188ᵇ21–6). What argument does Aristotle use for this conclusion? It has been suggested by some that Aristotle's argument is deductive, or syllogistic. He starts from a general principle, which he does not justify, that random things do not act on and are not acted on by random things (188^{a}31–3). Then he assumes, without stating it, that every becoming is an action of one thing on another and concludes that random becoming, that is, becoming between randomly connected termini, is impossible. This, then, shows that the termini of becoming cannot be contradictories since there are no causal, non-random regularities between contradictories. So the termini must be contraries since there are non-random regular connections between contraries and changes can all be described with contraries as their termini. Those who offer this analysis of Aristotle's argument often then go on to criticize him both for supposing that the fact that change *can* be described in this way shows that it *should* be, and for the assumption that the termini of change are non-randomly, that is causally, connected. They argue, for example, that the fact that something lacks the shape of a statue is 'insufficient to explain' why it comes to be a statue, i.e. to 'explain why the change *happens*'.[24]

This objection, however, and the analysis of Aristotle's argument on which it is premissed, miss what it is that Aristotle is trying to show. He is not trying to establish the principle that the contrary state from which a change originates is *sufficient* to explain why the change occurs to the other *specific* contrary state. As we have seen, the contraries are existential principles required, so to speak, to construct change. They 'explain' why change occurs in that limited sense, and not because a given contrary state is sufficient to bring it about, by a causal regularity, that some resulting contrary occurs. The parallel with geometry is again instructive. A particular circle requires points and lines, at the least a point for the centre and a line for the radius, in order for the circle to be constructed. This does not mean, however, that a given point and line are sufficient to guarantee that a given circle or any circle is constructed.

So Aristotle is not guilty of the error with which he is charged. Nor does he in fact argue from the principle that things do not act randomly on other things to the conclusion that the termini of change are contraries. He uses the former claim, in the text, only to introduce the discussion of the latter (188^{a}31–4). When he comes to the latter as his reasoned conclusion, at 188^{b}21 ff., he prefaces this conclusion with the phrase: 'if then *this* is true . . .'. The 'this' in this case, must refer to the immediately preceding material (188^{a}35–^{b}21). What Aristotle offers us there, however, is simply a review of a full range of cases

[24] Irwin, *First Principles*, 70–1.

where natural change (and, in a parallel way, artificial change) occurs. What is white does not come to be from just anything, e.g. from what is musical, but from what is black or some intermediate colour, and only accidentally from what is musical in case it just happens that what is some opposite colour is musical, (188^a33 ff.). What is in tune comes to be from some opposite state of lack of attunement (188^b12–15); and so on. This is the material on which Aristotle explicitly bases his general conclusion and, given this, it seems clear that his own argument (λόγος, 188^a31) is an inductive argument.

Is it an inductive argument based on things known by experience, as Aristotle has indicated it should be? Some have wanted to argue that it is not. Rather, it has been suggested, Aristotle thinks that the claim that what is white always and only comes to be by nature from some opposite colour, and only accidentally from what is musical, is an a priori or a conceptual truth.[25] But there is no clear indication in the text of the argument itself that Aristotle believes this.[26] Moreover, it seems pretty clear that this is not in fact an a priori or a conceptual truth. If it *always* did happen that what is white came to be from what is musical, that is, if all and only musical things became white owing to something about musicality, which is not conceptually impossible, then Aristotle could not say that what is white comes to be by nature only from a contrary and otherwise accidentally. For what always happens, at least in this sort of way, for Aristotle, happens not accidentally but by nature (*Phys.* II. 8 198^b35 ff.). So even if in this case what is white *also* always came to be from some opposite colour it would be an additional natural fact, and not an accidental fact, as Aristotle claims, that it comes to be from what is musical.

In addition to this, there are conceivable situations in which white does not come to be from some opposite colour in any way. For the atomists, for instance, as Aristotle understands them, white comes to be when certain atoms are arranged in a certain way. But previous to this those atoms might have been so dispersed that they gave rise to

[25] See e.g. Charlton, *Physics I–II*, 66. The contemporary notion of a 'logical' or 'conceptual' truth is notoriously problematic. Since Aristotle uses no such terminology, this is itself a reason for avoiding its use in interpreting him. I follow some minimal accepted rules for its use in rejecting its application here, however. (The notion of a priori knowledge is not as problematic.)

[26] He says: 'How *could* what is white come to be from what is musical except where it is an accident that what is not white, or black, is musical?' (188^a35–6). As a rhetorical question this expresses the ridiculousness of what is rejected but not Aristotle's basis for so regarding it.

no colour at all, or not dispersed but so arranged that they gave rise to no colour at all (perhaps by neither reflecting nor absorbing any light at all).[27] In this case white could come to be from (what is) no colour at all. A physical theory of this type is not conceptually impossible or ruled out a priori. One might say that Aristotle is rather cavalier here in ignoring this type of view which opposes his own, given that he is aware of it. But we should remember that he is confident that he has adequate arguments, which he produces elsewhere, against this type of theory.

There is, then, no good reason to suppose that Aristotle thinks that the premisses of his inductive argument are either a priori or conceptual truths. Rather, since he indicates in *Phys.* I. 1, and in parallel passages in the *Analytics*, that the scientist's starting premisses are items of experience, or of what is 'more knowable to perception' (184^{a}24–5), we should conclude that this is the way he views the premisses of his inductive argument here. This does not mean, of course, that Aristotle's inductive argument is a good one. It may well be that something enough like the atomist's story mentioned above is not merely conceivable but really holds, so that Aristotle's argument is undermined. But if it is undermined then we will also know that fact not by any a priori reasoning but by empirical reasoning of which Aristotle himself would approve.

XIII Aristotle's Proof of the Third Principle: *Physics* I. 7

Aristotle's conclusion at the end of *Phys.* I. 5, that 'it is evident that the principles must be contraries' (189^{a}9–10), is, of course, not the end of the story. In fact, this 'evident' conclusion will turn out to be true only in a way. In the language of *Phys.* I. 1 it turns out that the notion of a contrary as it has been developed so far is a *καθόλου* in the sense of a *συγκεχυμένον*, which contains so far undistinguished things as parts. When, for instance, 'the musical' comes from 'the unmusical', the unmusical can and should be understood in two ways. In one of these ways it is a state which is strictly opposite to the musical state which

[27] In response to the first example it might be said that Aristotle doesn't intend his principle to cover the case of a new (white) thing coming into existence. But it is not clear that he would not have to count this example as involving a case of white (or a white thing) coming to be, i.e. of *λευκὸν γίγνεται* (188^{a}37). In any case, the second example avoids this problem.

comes to be. In another way it is the thing, for example, the person, in the unmusical state who is acted on when the musical thing comes to be.

Hence, in one way the principles are no more numerous than the contraries but are in a manner of speaking two; but they are not two without qualification, because of the difference in being which belongs to them, but rather three. For being a man and being unmusical are different. (190^b35–191^a2)

So change requires three principles, two contraries—form and privation—and a substratum which is acted on when the change occurs. Uncovering these three things serves, again in the language of *Phys.* I. 1, to distinguish the things which lie undistinguished in the καθόλου notion of a contrary.

What kind of argument does Aristotle use to justify his final conclusion that there are these three principles? As he did in working his way to his initial result that the principles are contraries, Aristotle begins in I. 6 with a discussion of the views of his predecessors. His discussion here is, again, perhaps largely dialectical, but the most that he draws from it is that his own view 'has some argument' in its favour but still 'is very problematic' (189^b17–18, 29). In I. 7, following his earlier pattern, Aristotle moves from this inconclusive discussion of his predecessors to produce a conclusive argument of his own (189^b30 ff.). It is important to look at the way in which he proceeds in the crucial part of this argument (I. 7, 190^a31 ff.).

What comes to be is so-called in many ways. In some cases there is no coming into being but a particular thing comes to be *something*. There is coming into being simply only in the case of substances. In the other cases it is obvious that the *thing* which comes to be something must serve as subject. For a quantity, or a quality, or a relation, or a time or a place comes to be [a quantity or a quality, etc.] of some subject; since only a substance is ascribed to nothing else as its subject, while everything else is ascribed to the appropriate substance. But that substances too, and whatever else exists without qualification, come to be from some subject is obvious on review. For in every case there is something which serves as subject, from which the thing which comes to be does come to be—as for instance plants and animals come from seed. Some of the things which come to be without qualification come to be by change of shape, for instance a statue; some by addition, for instance things which grow; some by taking away, for instance the Hermes from the piece of stone; some by composition, for instance a house; some by alteration, for instance things which change their matter. It is obvious that all things which come to be in this manner come to be from subjects. Thus it is evident, from what has been said, that *everything* which comes to be is always composite: there is something which comes to be and something which comes to be that—the latter in two ways, either the subject or what is opposed. (190^a31–$^b13$)

As with Aristotle's earlier proof that the principles are contraries, here also his proof that every case of coming into being involves three things is based on a review of particular cases. It is the fact that those cases are obvious (φανερά) cases of the general conclusion that leads Aristotle to draw that conclusion as something which is evident. So Aristotle's argument is, again, inductive in form. There has been no temptation to claim here that Aristotle thinks our knowledge of the particulars on which the generalization is based is conceptual or a priori. That plants come from seed, that things require the addition of material to grow, that houses are produced by putting building materials together; these are not conceptual or a priori truths.[28] Nevertheless, some have wanted to argue this cannot be Aristotle's real argument for his conclusion since such an empirical argument could not explain why Aristotle is so confident of his conclusion. Rather, it is suggested, he must have in mind the following a priori argument:

> Considering now the mere concept of becoming we may argue that if one thing is properly said to become another then obviously . . . there must be something which does persist throughout the change, for otherwise the change would merely consist in one thing coming to be where another had ceased to be and there would be no reason to say that the one *became* the other. This argument appears to have an *a priori* certainty.[29]

The first thing to be said about this argument is that it is not the argument which Aristotle uses. The argument which he uses is the one quoted earlier. Secondly, it is not in fact clear that the 'mere concept of becoming' does require that something persists throughout any case of becoming. It is conceivable that cases of becoming could involve no single persisting object but rather only sequences of non-persisting objects (call them, say, space-time slices) which are ordered in a certain way. It is the proper ordering of the non-persisting objects

[28] Aristotle prefaces this argument by making certain distinctions, based in part on things 'we say' (189^b32 ff.); e.g., 'A man becomes musical' or 'An unmusical man becomes a musical man'. (Here and above I follow the traditional rendering 'musical' for the word which is perhaps better translated 'cultured'.) Aristotle does clearly think that various things of this sort, which we say, are obvious truths but he can hardly think that these particular things are known a priori or that it is inconceivable that they should be false.

[29] D. Bostock, 'Aristotle on the Principles of Change in *Physics* I', in M. Schofield and M. Nussbaum (eds.), *Language and Logos* (Cambridge, 1982), 186.

rather than the presence of any single persisting object which, on this proposal, distinguishes cases of becoming from cases where things merely succeed other things. Nothing in our ordinary understanding of becoming rules it out that we might discover, or that some already have discovered, that the actual cases of becoming which we are all aware of are like this. Since Aristotle does not himself argue that this is ruled out on conceptual or a priori grounds but only that it is ruled out by a proper review of actual cases we should not attribute such an argument to him. It is easy to feel, of course, that Aristotle's review of actual cases here is rather fast and loose. But the explanation for his confidence could be, not that he has an a priori argument up his sleeve to use if challenged, but that he can call on extensive acquaintance with the phenomena with which the natural scientist deals to rebut any proposed counter-example to his principles. This is, in any case, the appropriate response for him given that as a natural scientist he reaches his own knowledge of his principles by induction.

XIV Conclusion

Aristotle's method for the natural scientist, then, as he describes and uses it in the opening book of the *Physics*, is not an a priori method or a method which is appropriate for philosophy as opposed to science. However crude his description or use of the method may be his endeavours belong, in spirit, with those which we now think of as clearly scientific.[30]

[30] An earlier version of this paper was presented at a special session of the Los Angeles Area Colloquium in Ancient Philosophy on Aristotle's *Physics*. I am indebted for very helpful comments, on that occasion and subsequently, to James Bogen, David Charles, Cynthia Freeland, Lindsay Judson, Frank Lewis, Richard McKirahan, Michael White, and Charles Young.

2

What Makes Reality Intelligible? Reflections on Aristotle's Theory of *Aitia*

JULIUS M. MORAVCSIK

To say that some parts of reality are intelligible is to say that we can understand these and can offer explanations for them. Understanding and explaining are what we do; reality provides the objects for these states or activities. This suggests that theories of reality are one thing, while accounts of explanations and understanding constitute something else. The former embraces science and metaphysics, the latter a part of epistemology and the philosophy of mind. From the point of view of a conception adopting this bifurcation Aristotle's theory of *aitia* seems puzzling. For on the one hand it is a theory about various configurations of elements of reality, and on the other hand it seems to be a theory about explanations. We can, however, remove the puzzlement if instead of asking, 'How do we explain various parts of reality?' we ask, 'What features of parts of reality make these intelligible?' In this essay Aristotle's theory of *aitia* will be interpreted as an answer to this question. The answer is, roughly, that just as certain configurations of elements of reality make corresponding statements true, so certain configurations of elements of reality make some explanations adequate, true, and insightful. Aristotle's theory of *aitia* is a correspondence theory of explanations, thus dealing at once both with the world and with the ways in which we explain the world. Furthermore, for Aristotle the intelligibility of the world depends partly on fundamental changes and processes in the world being the manifestations of powers and potentialities that have an intelligible nature of their own.

Some will object immediately to any attempt to uncover the ontological configurations underlying adequate explanations by saying that explanation is a many-splendoured thing, and thus there exists a wide variety of configurations that correspond to explanatory patterns. Is there any non-vacuous characterization of the basic relations structuring the relevant configurations? Aristotle's answer is the affirmative, and his examples show that he is casting a wide net indeed for the range of phenomena his theory is to cover. For example, his account analyses such diverse configurations as propositions arranged in the pattern of premisses and conclusion, the constituents of biological specimens and the developing specimens themselves, and the relation between human actions and their initiating antecedents.

One might still feel that explanations are in the head while parts of elements of nature and their developments are in the world. Which is the proper subject-matter for Aristotle's theory? Or is it the case that his theory does not quite fit either mind or matter?[1] Asking, however, 'What are the constituents, aspects, and antecedents of a human action?', for example, is not only to raise a sensible question about ontological configurations, but at the same time—given an assumption about correlations between structures in reality and structures in human understanding—it is also to ask a searching question about explanation and understanding. Answering the question in the 'material mode' gives us data for giving an account of what we regard as adequate explanation in this kind of case.

The subsequent sections will show how Aristotle can give a non-vacuous general account of fundamental types of explanations and their ontological correlates.[2]

I What is in the Head and what is in the World?

Before we explicate the Aristotelian position, we need to lay aside two issues. First one might raise the Kantian claim that there is a dichotomy between the noumenal and phenomenal world, and that the configurations mentioned apply only to the latter. This issue needs to be laid aside, for it never surfaces within the realist conception of the world and the human mind that is adopted by Aristotle and his predecessors. Similar considerations apply to raising Humean doubts about the reality of causal powers in contrast with mere temporal conjunctions of events. As we shall see, Aristotle interprets natural

[1] This question is raised in J. Annas, 'Aristotle on Inefficient Causes', *Philosophical Quarterly* 32 (1982), 311–26.

[2] For an extended account see J. Moravcsik, '*Aitia* as Generative Factor in Aristotle's Philosophy', *Dialogue* 14 (1975), 622–38.

changes as the realization of potentialities. His analysis of change leaves room for something like the modern notion of an event, but these entities do not have an independent status in his ontology. A realist conception of potentiality is a presupposition rather than a consequence of Aristotle's views on what makes the world intelligible. This presupposition was not challenged by his contemporaries, though some of his predecessors tried to avoid the notion of potentiality altogether in their ontologies.

Given, then, this much 'realism', what is an Aristotelian *aitia*? One of his characterizations looks suspiciously epistemological (194^b18–20). For it says that we do not really understand what something is until we know or grasp what its primary *aitia* is. The 'something' in this formulation is supposed to include also genesis and destruction; we need to understand the *aitiai* of these items as well (194^b20–2). However further evidence shows that Aristotle has in mind the grasp of ontological entities and their configurations, and not mental or linguistic items. For in 198^a14–16 he says that the different kinds of *aitiai* correspond to different kinds of items that answer *dia ti* questions. The lines following give us some examples (16–21). These items are entities in the world. For example one of the *aitiai* is a raid that prompted people to go to war. Another is health as a goal or pursuit of people engaged in certain activities. Again, the essence, or nature, of something can bring about ways in which the object in question develops or grows. For essence or nature makes an element of nature into a self-determining entity with its own schedule of growth and development. *Dia ti* questions are, then, asking on account of what entity and relationship events take place, actions are performed, or substances are the way they are. The answer singles out the appropriate entity and relationship. This shows that it is misleading to characterize the *aitiai* as 'answers to why-questions' or 'the content of because-clauses'. For both of those items are mental or linguistic, whereas the examples show us that the *aitiai* are not.

There is nothing strange about Aristotle's way of looking at these things. We often point to elements of reality that are in some ways responsible for other elements. Here are a few simple examples. In our effort to understand in some context why people are dying we single out a drought, and say, 'That explains it'. In our attempt to understand why the car overheated, we point to a leak in the radiator and say, 'That accounts for it'; and the plumber pulls out something from the pipe and says, 'That accounts for the blockage'. In a medical school the research team looks not for linguistic items but for virus or bacteria that will account for certain illnesses.

In all of these cases our search is prima facie for elements of reality that stand in certain relations to other elements of reality. We may be able to analyse these efforts in what is called in contemporary philosophy the 'formal mode'. But even if such attempts were successful, they would not destroy the thesis that in many of our endeavours to understand nature we are in search for elements of reality and their interrelations rather than for explanations interpreted as mental or linguistic items. Just as some philosophers assume that corresponding to a simple true assertion of subject-predicate form such as 'Socrates is wise' there is a particular that has the attribute of wisdom, so Aristotle thinks that certain configurations of elements of reality underlie and correspond to certain explanations.

In proposing what might be called a correspondence theory of explanation Aristotle's motives are not linguistic. Quine once suggested that just as we do not posit 'sakes' in order to explain the meaning and truth of sentences involving the 'for the sake of' construction, we need not posit other—from his point of view suspicious—entities in our semantic analysis. Aristotle posits configurations of elements of reality corresponding to adequate explanations not in order to assign reference to certain linguistic expressions, but to account in a realist way for that in virtue of which some explanations are adequate and yield insight.

Intelligibility contrasts not only with unintelligibility, but also with mere describability. For example, a world consisting of independent singular facts of the '*a* is *F*' type would be describable but not subject to explanations.

Let us suppose that one could sum up all of the truths about the world in a book containing only a series of singular propositions of subject-predicate form. (We shall bypass the issue of identity, and take that for granted.) We could then also form, on the basis of the book, generalizations of the form 'all *A*s are *B*s'. Both science and common sense are not satisfied with such generalizations. They seek generalizations that have law-like qualities, i.e. that admit of counterfactual formulation as well. For example, we are not satisfied with noting that all leaves at a certain time are green. We claim that even if the leaf population were not identical with the one that happens to exist in our world, the statement would still be true. The reason for this is that we think that the colour of leaves is related to important processes that take place in plants such as photosynthesis. But if we treat this generalization as law-like, it makes sense to ask what in the world makes it a true law-like generalization? In response to this question we posit dispositions.

Some dispositions may be acquired as a result of interaction with the environment, while others are in the nature of the entity in question. The latter form the core of what Aristotle calls potentialities. We can explain changes as the realization of potentialities, and then account for the potentialities in terms of what initiates these, what their completion amounts to, and what structures they require. These layers of entities and structures renders the world intelligible in Aristotle's view. Adequate explanations must reflect the nature and effect of potentialities.

Sections 194^b23–5 and 195^a15–26 present us with an account of the basic *aitiai*. One kind of *aitia* is constituency. As we shall see, this notion as used by Aristotle is much wider than the modern notion of material. Another *aitia* is the motion- or process-initiating factor. This has been identified at times with the modern notion of efficient cause. I have argued elsewhere that this identification is misleading at best and false at its worst.[3] Within Aristotle's non-mechanistic teleological conception of reality there are many factors that initiate change and process that cannot be described by the modern term 'cause'. We see causal links primarily as links between events. But Aristotelian process-initiating factors can include the skill of a craftsman, a substance, as well as a realization of a potentiality. Another *aitia* is the 'form' of an element. We shall explore this difficult notion subsequently, and show that its proper interpretation is in line with the overall interpretation of *aitia* presented here. Finally, function and goal also serve as *aitia*, since these too contribute to bringing about processes and activities.

Aristotle does not say that every part of reality must have in its nature all four of these *aitiai* in order to be intelligible. Numbers do not have motion-initiating factors or goals, for example. But to the extent that things are intelligible, the conditions of such intelligibility will involve some of these four factors.

This brief review of the *aitiai* shows that Aristotle's theory of these is about intelligibility and hence both about the structure of the objects as well as the structure of accounts that humans need to attain.

[3] For a discussion see J. Moravcsik, 'Aristotle on Adequate Explanations', *Synthese* 28 (1974) 9.

II Processes and their Descriptions

We saw that according to our interpretation Aristotle's candidates for adequate explanations should not be description-dependent. They should simply single out the element responsible for another element in the way specified by the explanation. In view of this we need to pay attention to one of Aristotle's examples of *aitia*-giving in which a man called Polyclitus is said to be responsible for a certain statue coming into being ($195^{a}33$–$^{b}3$). On Aristotle's account, to say that Polyclitus was the process-initiating factor is less illuminating than to point out that Polyclitus the sculptor brought the statue into being. Does this not show that *aitia*-giving is description-dependent, since presumably Polyclitus = Polyclitus the sculptor? This objection assumes that Aristotle would agree to the modern versions of the semantic theory of identity.[4]

To understand Aristotle's approach let us look at a substance like Polyclitus from a metaphysical point of view. Within that perspective the deceptively simple proper name 'hides' a highly complex entity with many parts and potentialities and attributes. Different parts and different attributes account for different abilities and activities that Polyclitus performs or possesses. Modern common sense, science, and much of philosophy would surely agree with Aristotle.

Presumably moderns and Aristotle would also agree that referring to an entity is not the same as referring to it and all or some of its parts. In referring to a person Jones *simpliciter* I am not also referring to his arm. This accounts for the contrast between

(1) Jones was responsible for the cricket victory.

and

(2) Jones and his strong arm were responsible for the cricket victory.

The two statements are clearly not equivalent. Sentence (2) entails (1) but not the other way around. We can see this also by reflecting on

(3) Jones and his keen eye were responsible for the cricket victory.

[4] For further discussion of this issue, see Cynthia Freeland's essay, 'Accidental Causes and Real Explanations', Chapter 3 in this volume, section I; for some of the recent literature on the subject, see nn. 15 and 16 of her essay.

Sentences (2) and (3) are not equivalent. They point to different factors leading to a victory. Aristotle would say that the two statements propose different *aitiai* for the same outcome.

Aristotle's approach to examples like the one involving Polyclitus can be understood if we think of him as treating attributes and powers in the same way as we would treat parts. Indeed such a move is not unnatural. Let us consider

> (4) Jones and his shrewdness were responsible for the cricket victory.

It seems entirely natural to treat the reference to an attribute here in the same way as the references to parts in (2) and (3).

Thus I suggest that we read:

> (5) Polyclitus the sculptor was responsible for the statue.

as equivalent to:

> (6) Polyclitus and his sculpting were responsible for the statue.

These reflections show that Aristotle's *aitia*-giving is not description-dependent, though it can be described as 'feature-dependent' in the sense that the *aitia* will single out not only an element but in some cases also a part or feature of the element that played a vital role in the generation or destruction to be accounted for.

One might ask why we should not simply say that Polyclitus' art was responsible for the statue? Such a view would admit of two interpretations. According to one of these the art referred to is something general that can be shared among many practitioners. According to the other reading the art in question is something unique to Polyclitus; a specific non-substantial particular that only he possesses. (On non-substantial particulars see Strawson.[5]) While the second reading would please a typical modern audience obsessed with the importance of individual talent, genius, etc., the first is much more likely to be what Aristotle, or earlier Plato, would have had in mind.

Fortunately we need not decide the issue. For on either reading we would miss an essential component of Aristotle's conception of this kind of *aitia*. Arts—universal or particular—cannot by themselves initiate processes. They can do that only in so far as they are embodied in substances. Thus a complete characterization of the 'source' ('*archē*') of the statue involves both the art of sculpting and the agent practising it. Thus 'Polyclitus the sculptor'—and also the more direct 'a sculptor'—

[5] P. F. Strawson, *Individuals* (London, 1959).

must be interpreted so as to introduce both the relevant agent and his art.

These considerations show that *aitia*-giving even in this case remains ontological and is feature-dependent. How Aristotle's insight might or might not be reconcilable with modern semantic notions of reference and intensionality is beyond the scope of this paper.

What was said here about the creating agent applies also to the object produced. The statue produced may be the ugliest object in the room, but for the purposes of explicating the creative process and generation in question, it is the object being a statue that is relevant. Under normal circumstances the sculpting ability is exercised to produce a sculpture. This is the part of the nature of the entity that the art in the artist produces. Being the ugliest entity in the room is merely an accidental feature of the object and is not directly related to the exercise of the sculpting ability.

Distinguishing the object as a whole from its parts, properties, potentialities, etc., does not introduce an ontological jungle into Aristotle's philosophy. If we analyse an opera into acts, arias, chorus, etc., and realize that to refer to the opera is not *eo ipso* a reference to its parts, aspects, etc., this is not to lapse into ontological promiscuity, but simply to realize that operas are complex entities, with different constituents playing different roles. Aristotle does not wish to sweep this fact under the semantic rug of simply naming things.

III A Role for Events in a Metaphysics of Power and Potentiality?

As noted above, in modern philosophy accounts of changes are given typically as relations between events. Does Aristotle share this view? If not, what is his 'realist' account of changes? Does he have a notion similar to causal relations between events? Julia Annas suggests a negative answer to this question.[6] The closest item among *aitiai* to efficient cause is the motion-initiating factor. We saw, however, how problematic any identification between these notions is.

We shall leave efficient causes out of the picture, and concentrate on the question of what the closest analogue to an event is in Aristotle's ontology. Let us lay aside conceptions of events as mere passings of

[6] J. Annas, unpublished paper on Aristotle and causation.

time, or as the instantiations of attributes over time even in an unchanging world, and consider for our purposes events as changes of location or quality over time. In this sense Aristotle does acknowledge events, since he acknowledges the existence of changes and aims at an account of these. However, there are no 'mere' changes in Aristotle's ontology and the analogues to changes do not have an independent ontological status.

Changes for Aristotle are parts of processes. A process, in turn—such as the development of an animal, the building of a house, or the returning of foliage in the spring—is given a partly teleological or functional interpretation. Shall we say, then, that Aristotle has events in his ontology, but that these have necessarily teleological orientations? This would still not be satisfactory. One could have a conception of a world in which events cause events, and the whole historical process tends towards realizing a cosmic purpose, e.g. the will of God. Such a conception would be markedly different from that of Aristotle. For he interprets the process as constituted by sequences of changes in which potentialities that are unique to species or their genera are realized or brought to fulfilment. Substances constitute—ultimately—these kinds. Hence changes in the normal course of history are the result of self-determined schedules of potentiality realizations of substances and interactions between substances. Thus it is misleading to think of Aristotelian changes as causing each other. The changes have as one of their sources potentialities, and these in turn are ingredients of other ontological kinds.

Still, there is a link between modern causal explanations and Aristotelian explanations of change and locomotion. For if we add up in any given case the potentialities that are to be realized in that context and the conditions that result from the potentialities not being hindered, then this set of conditions is equivalent to what in at least some modern schemes would be as causally necessitating some changes. Thus one might say that there is 'extensional equivalence' between some versions of modern causal theories and the Aristotelian account of substance, activity, and potentiality that bring about a change, but in terms of analysis the two conceptions differ.

These reflections shed light also on why for Aristotle there should be non-eventlike entities as motion-initiating factors. For though his conception of nature posits chains of changes, there are additional factors that contribute to the coming into being of any one member of such a chain. This is so because in the normal case a change is a partial realization of a potentiality, and different potentialities have different kinds of sources. In the case of procreation, which is a part of essential human functioning, the key motion-initiating factor is the father, since—according to Aristotle—he is the substance whose active power is being realized. In the case of processes that are not essential to human functioning, such as sculpting, the motion-initiating factor is an art, since that is the unique element needed to combine with human agency—a condition for producing artefacts in general—to bring about that particular kind of artefact.

Some passages suggest, however, that there is no room in Aristotle's ontology even for processes. In *Phys.* $200^{b}32$–$201^{a}2$ Aristotle says that there are no processes 'in addition' to the things (*pragmata*) that undergo these. This, however, does not mean that Aristotle denies the existence of changes and *kineseis*. Let us return to the point made earlier about a substance being a complex entity, made up of parts, attributes, and potentialities. The potentialities of a natural substance are realized according to a self-determined schedule. Within this conception of nature, once we obtained a complete description of every active substance there would be nothing left over, waiting to be accounted for. The full account of any substance would include its life history and its interactions with other substances. Thus Aristotle does not mean to adopt what in modern parlance would be a reductionistic programme for changes. His view is, rather, that once we have a full account of substances, giving an additional account of all of the natural *kineseis* would be to count some elements of reality twice. Ross suggests that the *pragmata* in this passage are not the things undergoing change, but respects in which change occurs, thus referring forward to the reference to four categories in the following lines. But on that reading—as Ross admits—it is difficult to find a motivation for the inclusion of this statement. The reading proposed in this paper motivates the statement, since it places it within the framework of Aristotle's substance-oriented ontology. Since all changes are related to various elements in some of the categories, there are no processes over and above the things undergoing change; all changes and processes are accounted for in a 'full' description of other elements in Aristotle's ontology.[7]

[7] See W. D. Ross, *Aristotle's Physics: A Revised Text with Introduction and Commentary* (Oxford, 1936), 536. My interpretation assumes the admittedly controversial thesis that Aristotle regards artefacts not as genuine substances but only analogues of these. They are like real substances except that their 'soul' is the artist or in the artist, and thus

Further evidence for this interpretation is provided by Aristotle's locating changes and processes in the categories of substance, quantity, quality, and place.[8] Processes too have essences and are re-describable in ways analogous to the ways in which items in other categories are—although, to be sure, they do not have a full essence, since their nature and being depends on the nature and being of substances.

These considerations show that there is a place for event-like entities in Aristotle's ontology. For at least some events are changes, and Aristotle not only acknowledges these as having an ontological status, but also has an account of them in terms of sequences of potentiality realizations. These changes and processes can be accounted for—in the sense explained in this paper—by other elements of reality. They, in turn, have also an explanatory role, helping to account for the nature of the potentialities that are—in the normal case—their sources.

IV The Ontology of Explanatory Chains

According to the interpretation presented one entity can account for, or be responsible for, another entity. But if entity *B* explains entity *A*, do we not need entity *C* to explain the nature of entity *B*? Aristotle's conception seems to lead to explanatory chains within which a given

outside the entity. This deprives artefacts of having a self-determining nature and a developmental schedule. But it is still true of them that they are brought about by the four familiar factors, and that these are also responsible for their maintenance. They are brought about by an agent—the instantiated artistry of the artist—and have a form and function, though this is designed by the artist and not by forces and structures within the object. The matter of these objects consists of the realizations of capacities to endure. These must include their weight, hardness, etc. Without the constant realization of potentiality resulting in constancy in stuff, weight, etc., the artefact would not survive. But since the 'soul' is outside, maintenance and persistence are not the dynamic processes that they are in the case of biological substances. In the biological case constant motion and activity are required for persistence; in the case of artefacts stability is required for persistence, and this is achieved not by constant self-determining activity within the object, but by the static realizations of the capacity to survive, resulting in at most minor changes in hardness, weight, etc.

The analogy with artefacts is used by Aristotle in order to give intuitively simple and persuasive introductions to the four *aitiai*. It is unlikely to be a matter of different developmental stages, since all the basic Aristotelian ontological constructions already have their ancestors in Plato's thought. On this point see J. Driscoll, 'The Platonic Ancestry of Primary Substance', *Phronesis* 24 (1979) 253–69.

[8] *Cat.* ch. 14.

explanandum has another entity as its explanans and this in turn another one. Where do these chains end? Sooner or later some entity needs to be taken as ultimate or self-explanatory. Within modern non-entitative conceptions of explanations this does not cause a problem. Explanations that humans give can end for a variety of reasons. These can be conventional delineations of departmental boundaries, lack of interest in further pursuits, or historically and culturally limited notions of what is fundamental. There may be more to it than that, but there need not be. However, any such termination of Aristotelian entitative explanatory chains is unacceptable: entities in the world cannot cease to exist or become ultimate and self-explanatory on account of barriers between disciplines invented by humans. We need metaphysical grounds for terminating entitative explanatory chains.

Plato's account of what makes the world intelligible (e.g. *Phaedo* 96–106) is a striking example of entitative explanations. He interprets the Forms as underlying whatever order and stability the world has, in virtue of the unique natures the Forms have and the relationships that some elements of reality can have with them. The unique nature of the Forms is discussed in recent literature under the heading of self-predication. This is misleading, as it suggests semantic considerations. A better interpretation is suggested by reflection on Plato's characterization of these natures as something complete that can be at most be reflected partly by other elements of reality. Our grasp of these natures enables us to see how and why a Form is ultimate and self-explanatory. The question 'What makes a Form ultimate and self-explanatory?' admits of possible answers varying greatly in terms of subtlety. We need not assume that whatever accounts for the equality of two things must have in a quantitative sense 'a lot of equality', or be 'very equal'. We do owe, however, some metaphysical account of why 'the buck stops here', and Plato attempts several times to address this issue. A detailed examination of his efforts is beyond the scope of this paper.

Plato regards the soul as self-moving (see, e.g. *Phaedrus* 245c–e). Why should the soul not be another end of an explanatory chain? If Plato had agreed, he would have had more than one kind of explanatory chain. But Plato construed the soul as dependent on the Forms, both for the objects of cognition and for the harmony and order that its functioning requires. Thus his remained a unitarian theory of explanatory chains.

Aristotle differs from Plato in two main respects. First, his theory of *aitia* is a pluralistic theory of explanatory chains. Second, he separated the question of what is ontologically fundamental from the question of what is ultimate in terms of explanatory chains. This required that he reconsider the notion of self-sufficiency, and not share Plato's simple and austere view on that. However, this topic too would take us too far from our main subject.

The different *aitiai* constitute for Aristotle different entitative explanatory chains. For example, flesh with its dynamic structure stands in the constitutive relation to the human body, helping to account for its functioning. A male human stands in the motion-initiating relation to his offspring, contributing to the full account of generating a new substance. Health stands in the function or aim relation to humans and some of their activities. Yet none of these entities need be for Aristotle ontologically fundamental.

In this way Aristotle can have four different explanatory chains, and yet insist that substances are what is ontologically fundamental. All else depends on substances in a way in which substances do not depend on other things. Spelling out this asymmetry, and listing the various proposals as to what Aristotle really thought substances are, cannot be done here. But this interpretation of *aitia* is compatible with the main lines of interpretation concerning these other topics.[9]

Aristotle's scheme also admits hierarchies of, for example, constitutive relations. The constituency of an element *E* has in turn other constituents, and these have constituents as well. But being the final item in such a chain is not what endows an entity with being ontologically fundamental.

Someone might ask for explanations of notions like constituency or goal and function. For Aristotle, however, these remain basic, non-definable notions. He explains explanations in terms of these and not the other way around. A Kantian might posit these as basic innate human ideas in terms of which we organize experience. But given Aristotle's 'in-put' theory of concept formation he would locate these elements not in the mind but in the world.

V Form and Matter as *Aitiai*

In this section we will see how form and matter fit into the roles assigned by this interpretation to *aitiai*. This might seem puzzling at

[9] For more discussion see J. Moravcsik, 'Essences, Meanings, and Generic Propositions', forthcoming.

first. How could bits of stuff do any explaining? Why should attributes account for natural processes and functioning? Finally, how do attributes and material fit into an ontology that according to our interpretation centres on power and potentiality?

We shall begin with matter. Is Aristotelian matter something like material in the modern sense?[10] Some examples might support such a view. In *Phys.* I. 7 bronze is given as an example of matter. On the other hand, in *Met.* 1045^{a}20–5 matter is identified with potentiality. How are we to reconcile this evidence? Assuming as we have all along that artefacts are not real substances, the way out is not difficult. The second passage applies to natural entities and their aspects while the first only to artefacts that are brought into the picture by Aristotle only to illuminate some aspects of his concept of a substance. With respect to the notion of matter the analogy is infelicitous. For a bronze statue does not contain a self-determined schedule of development and maintenance as the biological elements do. The analogy is good only for pointing out that just as not anything can develop into this or that animal, so not anything can serve as matter for a statue. The dynamic aspect of existence and persistence, so vital to living things, is absent in the cases of artefacts.

Others have noted already that Aristotle's notion of matter is much wider than our notion of material.[11] For it includes also abstract elements, such as premisses in an argument, and letters in a syllable (*Phys.* 195^{a}15–20). Even in these abstract cases we see the work that potentiality does. Letters have no independent status in spoken language. Their role is simply that of potential contributors to syllables.

In the core cases, namely those of biology, the identification of matter and potentiality is easy to see as long as we keep in mind that biological entities can be and remain what they are only by constant activity, and that these activities constitute the sequential—or gradual—realizations of potentialities whose source is in these entities themselves. Frozen at a moment we can see a substance as informed potentiality, i.e. an actual so-and-so. But over stretches of time the substance is a series of potentiality-realizations, hierarchically arranged.

[10] For a defence of a negative answer to this question see A. Code, 'Soul as Efficient Cause in Aristotle's Embryology', *Philosophical Topics* 15 (1987) 51–9.

[11] Ross, *Aristotle's Physics*, 293.

Matter for animals, such as flesh, seems to be just what we call stuff or material. But underlying this appearance there is constant activity needed to keep flesh a living tissue. The flesh as informed matter and the potentiality-realizations coincide, as a modern semanticist might put it.

Given this interpretation of matter as potentiality we can see how the entitative explanatory theory applies to this *aitia*. We understand what a specimen of a biological species is by understanding the powers and potentialities that are realized in its existence and life-span.

Needless to say, a full account of Aristotelian matter would require much more to be said. These sketchy remarks, and the ones following on 'form', are meant to show only how adequate interpretations of these notions make it possible to fit these elements of Aristotle's metaphysics into the entitative account of *aitia* presented here.

The conception of Aristotelian matter sketched helps in the understanding of the Stagirite's notion of 'form'. Again, some considerations pose initial difficulties. Shape is given as one of the examples, and if one thinks of 'form' as attributes in general, it is difficult to see how these are necessarily related to potentialities and how they could fit into explanatory roles related to these. But shape is given as an example only in connection with artefacts like statues. As we saw, their 'matter' contains potentialities only in a very weak sense. Furthermore, as we shall see, being an attribute is not what is essential to an Aristotelian form.

In *Met.* 1041^b7–8 and 27–8 form is characterized as that by reason of which this kind of matter becomes this definite thing. This is hardly a characterization of attributes in general. Rather, it links certain items closely to potentialities. Forms are invoked not to answer a nonsensical question like 'What makes this actual thing actual?', but rather to answer the question 'What is responsible for the realization of this potentiality as a part of something coming into being or persisting?'

In the light of this passage, we can understand the description we find in *Phys.* II. 7, where form is described as whatever accounts for an entity being what it is or for its functioning. If forms are invoked to answer the question 'On account of what can this thing be what it is and do what it does?', then surely items from a variety of modern ontological categories could fill the bill. Different kinds of items will be required in different contexts.

For example, in *An.* 412^a19–22 the soul is described as the form of a natural body potentially having life. This is, then, what makes the potential living thing into an actual one. Thus it is not surprising to find in 413ᵃ25–8 that the soul is partly a power, partly activity, and partly a structure. These are the items that make something potential into an actual living thing. (The same point is made at 424ᵇ1–3.)

Structures, activities, and powers all satisfy the formula for 'form'. Furthermore, there is no good reason to suppose that Aristotle thought of all of these as universals or attributes. In the case of powers and activities it is just as easy to think of these as 'mass-terms', i.e. the sum of what is denoted by 'activity', rather than what can be divided into universal and particular.

Form, then, is whatever actualizes a potentiality, thus leading to the development and flourishing of an element in nature in the primary cases, and of items that are related by virtue of various dependency relations to parts of nature. So understood, Aristotelian form can easily play the role required of it within an entitative explanatory theory. One can point to an activity or structure and say, 'This is what makes this a flourishing human, beaver, tree, etc.'.

VI Conclusion

Once we understand potentiality, actualization, and the priority of what is self-determined in Aristotle's scheme of things, we can make sense of his theory of *aitia* as an entitative explanatory theory, differing from Plato's in its pluralism, and its separation of what is ontologically prior from what is prior in terms of explanatory role.

If we try the entitative explanatory theory on a collection of notions including thing, event, and causality in the modern sense of these expressions, we run into trouble. Why should events account for each other? How do events account for things? Causal links presuppose events and things rather than serve as a constituent helping to account for them.

If we exchange things and events for powers, potentialities, and their realizations, then we can understand Aristotle's *aitiai* as the basic ingredients in potentiality realizations. There must be a source for the potentiality, there must be something that brings it to actuality, and there must be a 'goal', or some final state which is the culmination of the realizations. Within this dynamic picture, the different elements of reality discussed in this paper do serve in explanatory and accounting roles. Aristotle's dynamic conception of nature turns out to be more like Heraclitus' dynamic ontology than the 'thing-metaphysics' of Descartes, Locke, or in our times Strawson.[12]

[12] I wish to thank Nancy Cartwright and Susan Levin for useful suggestions.

3

Accidental Causes and Real Explanations

CYNTHIA A. FREELAND

Introduction

Among Aristotle scholars current opinion is virtually unanimous that the theory of the four causes is really a theory of explanation. Julius Moravcsik, for example, writes that the theory, 'properly understood', 'is the Stagirite's account of what constitutes an adequate explanation'.[1] Julia Annas remarks in a similar vein, 'It is a great improvement to cease thinking of an *aitia* as a cause and to treat it instead as an explanation, a "because".'[2] A host of other Aristotelian scholars have joined in, as has the philosopher of science Bas van Fraassen, who sees in Aristotle's texts a theory of causes emphasizing, like his own, 'the pragmatics of explanation'.[3]

* An earlier version of this paper was read at the Central Division meetings of the APA in 1986; I am grateful to David Charles, my commentator on that occasion, for his criticisms, and to others present, in particular Dan Graham, for their comments. A later version was read at a conference on Aristotle's *Physics* co-sponsored by Rutgers University and USC in 1989; I would also like to thank the participants at that conference for their questions and suggestions. Finally, I am indebted to Lindsay Judson for a number of helpful suggestions about improvements in the structure, style, and substance of this essay.

[1] Julius Moravcsik, 'Aristotle on Adequate Explanations', *Synthese* 28 (1974), 3. Moravcsik's view is presented in a somewhat modified form in 'What Makes Reality Intelligible? Reflections on Aristotle's Theory of *Aitia*', in this volume. I will refer to this more recent paper in more detail below.

[2] Julia Annas, 'Aristotle on Inefficient Causes', *Philosophical Quarterly* 32 (1982), 319.

[3] Jonathan Barnes, *Aristotle's Posterior Analytics*, Clarendon Aristotle Series (Oxford, 1975), esp. p. 96; William Charlton, *Aristotle's Physics I, II* (Clarendon Aristotle Series; Oxford, 1970); Martha Nussbaum, *Aristotle's De Motu Animalium* (Princeton, 1978); Richard Sorabji, *Necessity, Cause and Blame: Perspectives on Aristotle's Theory* (London and Ithaca, NY, 1980), esp. chs. 1, 2; and Bas C. van Fraassen, *The Scientific Image* (Oxford, 1980), esp. ch. 5; and also 'A Re-examination of Aristotle's Philosophy of Science', *Dialogue* 19 (1980), 20–45. I should note that these writers (Moravcsik in particular) do not assume it is at all evident what is meant by saying that the four αἰτίαι are explanations, or what form explanation takes in Aristotle. For an especially clear answer to these

The motivation for the explanation approach seems mainly negative. Advocates of this approach deny that Aristotelian αἰτίαι are causes because they take it as obvious that (at least) three of the Aristotelian αἰτίαι differ quite radically from causes as 'we' conceive them (necessary or sufficient conditions, events constantly conjoined to effects, events probabilistically conjoined to effects, or others). On the explanation approach we can preserve some parallels between Aristotle's and 'our' accounts of causes, provided we reserve use of the term 'cause' for what he calls moving or efficient causes. Then presumably the other αἰτίαι (formal, final, and material) fall into place in this picture by being interpreted as explanatory factors relative to this one 'real' (i.e. efficient) sort of causal relation that seems most to resemble our own.[4]

I believe that several serious problems for the explanation approach have not been well thought out in the literature. Three puzzles in particular concerning the positive commitments of this approach deserve further critical attention. The first problem for the explanation approach concerns its understanding of Aristotle's views on the metaphysics of causal relations. There is no clear, well-defended consensus about whether Aristotle should be understood to be a realist about any or all of the four types of causal relations. Further questions can be raised about the sorts of entities he takes to be involved in causal relations, and about whether he understands causal contexts as intensional or extensional.

The second problem with interpreting the canonical presentation of the doctrine of the four cases in *Phys.* II as a theory of explanation is that Aristotle elsewhere offers his own theory of explanation: I mean his account of scientific explanation, framed in terms of the syllogism, in the *Posterior Analytics*. On this account, scientific explanation requires the provision of a definition that links the explanandum to pre-eminently knowable first principles via a proper deductive framework.[5] One would suppose that this theory of scientific explanation is

questions see Gail Fine, 'Forms as Causes: Plato and Aristotle', in Andreas Graeser (ed.), *Mathematics and Metaphysics in Aristotle* (Bern, 1986). For a dissenting, or at least more tentative, opinion about the merits of the explanation approach see David Charles, *Aristotle's Philosophy of Action* (London and Ithaca, 1984), esp. p. 47.

[4] This motivation seems fairly close to the surface in Fine's article.

[5] For an account of the relevant sort of explanation as co-ordinate with 'understanding', see Myles Burnyeat, 'Aristotle on Understanding Knowledge', in E. Berti (ed.), *Aristotle on Science: The Posterior Analytics*, Proceedings of the Eighth Symposium Aristotelicum (Padua, 1981), 97–139. Also see Barnes, *Posterior Analytics*, pp. ix–xvii;

distinct from the *Physics'* account of causes or of 'causal explanations', though of course somehow related to it. Advocates of the explanation approach ought to supply a clear account of this relationship.

The third, most pressing problem for the explanation approach is that explanation is not itself a transparently clear notion. There is a great deal of controversy about the nature of explanation within discussions of contemporary metaphysics and philosophy of science. Settling questions about whether Aristotle is a causal realist will not suffice to settle questions about whether he is an explanatory realist: these positions map on to one another in complex ways.[6] Scholars discussing Aristotelian explanation have concentrated on the formal model of scientific explanation in the *Posterior Analytics*, but this could well differ from his theory of the four causes in the *Physics*. Perhaps in the latter theory, though not in the former, Aristotle emphasizes that explanation is a 'pragmatic rather than syntactic or semantic concept.'[7] Explanations are relative to inquirers, situations, contexts, presuppositions, and interests; van Fraassen has argued that precisely these sorts of factors are the basis for Aristotle's distinguishing the four types of causes in the *Physics*. That is, he sorted out the four kinds of *αἰτίαι* he did as the result of classifying different sorts of why-questions that arise in different contexts for inquirers with distinct sorts of interests.[8] This pragmatic variation on the more standard-explanation approach raises further questions: To what extent *does* Aristotle recognize the role of pragmatic factors in explanation? Is this a fruitful line to follow in describing differences between the *Physics* and the *Posterior Analytics*?

In what follows I will discuss all these problems more fully. In section I, which provides a reading of Aristotle's account of causation in *Phys.* II, I will defend the claim that Aristotle is a causal realist of an

and Barnes, 'Aristotle's Theory of Demonstration', in Jonathan Barnes, Malcolm Schofield, and Richard Sorabji (eds.) *Articles on Aristotle*, I, *Science*, (London, 1975), 65–87).

[6] As explained by Jaegwon Kim in 'Explanatory Realism, Causal Realism, and Explanatory Exclusion', in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (eds.), *Realism and Antirealism*, (Midwest Studies in Philosophy 12) (Minneapolis, 1988), 225–39, esp. 231–2). (I am grateful to Christopher Shields for drawing my attention to Kim's article and its relevance for my project in this essay.)

[7] Michael Scriven, 'Causation as Explanation', *Noûs* 9 (1975), 4.

[8] Van Fraassen, 'A Re-examination', 32–4. This is in contrast to a version of the explanation approach proposed by other scholars who argue instead that Aristotle distinguishes the four causes on the basis of a kind of linguistic analysis of different types of why-questions. Such a view is defended by Dan Graham in *Aristotle's Two Systems* (Oxford, 1987).

unusual sort who believed that there are four types of objective causal relations in the world. Next, in section II, I will discuss Aristotle's views on the nature of explanation; here I will argue that he is an explanatory realist. I will consider differences between his accounts of explanatory knowledge in the *Posterior Analytics* and the *Physics*, paying particular attention to the pragmatist version of the standard-explanation approach. Though in the end I deny that the pragmatist approach affords a satisfactory interpretation of Aristotle's doctrine of the four causes, I show that it improves upon the more standard-explanation account by prompting us to recognize and reflect upon differences between Aristotle's discussions of *αἰτίαι* in the *Physics* and in the *Posterior Analytics*.

In the final section of this essay I will turn to an odd but intriguing aspect of Aristotle's picture of causes, namely, his thesis that there are, in addition to real or essential causes of various types, also accidental causes. I will explore the particular use Aristotle makes of the notion of accidental causation in his treatment of chance. Because such pragmatic factors as contexts, presuppositions, and interests seem particularly prominent in Aristotle's discussions of this sort of causation, my claims about Aristotle's causal and explanatory realism might well seem to break down for chance and other accidental causes. Still, I want to show that Aristotle's causal and explanatory realism are at play here also.

I Aristotle's Causal Realism in *Physics* II

If causation is 'really' explanation, according to Aristotle, then presumably causal statements involve certain descriptions of events or states of affairs. Although there may be some temptation on the 'explanation' interpretation to regard explanations as linguistic entities, and so to ascribe to Aristotle a Davidsonian-style concern with the logical form of causal statements,[9] most of the interpreters mentioned earlier have resisted this temptation, recognizing that Aristotle's attention is on things in the world (or that he writes in the material not the formal mode). So they call his theory a classification of 'explanatory factors', or, in Sorabji's case, of 'what provides an explanation'.[10]

[9] See Donald Davidson, 'Causal Relations', in *Essays on Actions and Events* (Oxford, 1980), opening sentence.

[10] Sorabji, *Necessity, Cause and Blame*, 40.

Gail Fine has offered an especially clear and detailed interpretation along these lines. She attempts to regiment Aristotle's theory of the four causes by suggesting that each type of cause could figure as an explanatory factor within an event cited as a cause:

[E]ven if we restrict causes to events productive of change, still, each of the four *aitiai* can be—if not a cause, at least—causally relevant, by being a factor one cites in explaining change; they will still be referred to in, be parts of, causal explanations.[11]

Several questions arise for any proposal which interprets Aristotelian αἰτίαι as explanatory factors. One concerns whether causal contexts, understood thus in terms of explanatory factors, are intensional. Notoriously, both events and states of affairs may be described in varying ways, some more or less explanatory than others. Did Aristotle regard causal contexts as intensional—i.e. is something an explanatory factor only under a certain description? Moravcsik, after some discussion, seems to answer this question 'yes'; Charles, similarly after discussion, 'no'.[12] Annas writes that Aristotle 'does not envisage that a statement might truly pick out a causal relation, but in a quite unexplanatory way'.[13] Fine comments in a criticism of Annas, 'I take it . . . that explanation, for Aristotle, can be construed extensionally'.[14]

But in my view, both the question just raised and these varying responses to it assume too much about Aristotle's theory of reference—and seeing this will show up a problem about the explanatory-factor interpretation. Specifically, what is assumed is that Aristotle could have formulated and considered the question of whether co-referential expressions are intersubstitutable in causal contexts. Such a question presupposes that Aristotle roughly shares our views on identity and reference. But a number of articles published in the last fifteen years have challenged the assumption that Aristotle grasped the modern (post-Leibnizian) concept of identity.[15] Furthermore, it has been argued that Aristotle would not have regarded such phrases

[11] Fine, 'Forms as Causes', 72.

[12] Moravcsik, 'Aristotle on Adequate Explanations', 12–13; Charles, *Philosophy of Action*, 47.

[13] Annas, 'In efficient Causes', 320.

[14] Fine, 'Forms as Causes', 71, n. 7.

[15] For the discussion that first raised doubts, see Nicholas White, 'Aristotle on Sameness and Oneness', *Philosophical Review* 80 (1971), 177–97. See also F. J. Pelletier, 'Sameness and Referential Opacity in Aristotle', *Noûs* 13 (1979), 283–311. Both White and Pelletier think that Aristotle had, but abandoned or lost his grip on, the concept of identity as we know it. For related discussions see also Alan Code, 'Aristotle's Response

as 'Polyclitus' and 'the sculptor' as co-referential, because his ontology includes, in addition to substances and accidents, certain entities (dubbed 'kooky objects' by Gareth Matthews) which are conjunctions of the two.[16] These are accidental unities, such as Socrates seated or the pale man. It turns out, on this interpretation, that Aristotle holds that the entity identified as an intrinsic cause of a statue—a sculptor—would be a kooky object dependent upon, but not identical with, the true substance—a man in this case. Thus the terms 'Polyclitus', 'the man', and 'the sculptor' are not, strictly speaking, co-referential. Still, Aristotle could well say that to call Polyclitus the cause of a statue is (*pace* Annas) to pick out the cause in an unexplanatory way, for it is to pick out something bearing a certain relationship—namely, accidental unity—with the cause of the statue, properly speaking. But because of his distinct views about identity and the ontology of accidents, this is different from saying that Aristotle envisaged either causal or explanatory contexts as intensional.

I have raised doubts about the appropriateness of directly comparing Aristotle's views of identity and reference to contemporary ones. Further doubts can be raised about whether this is a well-grounded approach to Aristotle's picture of the nature and ontology of causal relations. The explanatory-factor interpretation maps Aristotle's ontology on to a contemporary view in which causal relations obtain between events.[17] But Aristotle does not seem especially concerned to specify a uniform category of entities as causes. Sometimes he speaks of a property, like the art of building, as a cause; at other times he mentions a substance, like a man. Another difficulty is that Aristotle says that all the four causes can be either actual or potential. This means that he is perfectly willing to describe even the moving cause—the one that seems most similar to 'our' notion of causes as events—as a substance (i.e. when it is potential, rather than actual).[18]

to Quine's Objections to Modal Logic', *Journal of Philosophical Logic* 5 (1976), 159–86, and Fred Miller, Jr., 'Did Aristotle Have the Concept of Identity?' *Philosophical Review* 82 (1973), 483–90.

[16] Frank Lewis, 'Accidental Sameness in Aristotle', *Philosophical Studies* 42 (1982), 1–36; Gareth Matthews, 'Accidental Unities', in Malcolm Schofield and Martha Nussbaum (eds.), *Language and Logos* (Cambridge, 1982), 223–40. I should stress that, while their topics and views are related, these two authors reached their conclusions independently.

[17] This is especially true in Fine, 'Forms as Causes', 71–2; she suggests that this is more plausible for Aristotelian actual, rather than potential, causes.

[18] On this point, see Fine, 'Forms as Causes', 71.

Let us look more closely at Aristotle's discussion to pursue this question about the ontology of causal relations. It will come as no surprise to the reader familiar with Aristotle that he writes, in his treatment of causation in *Phys.* II. 3, that cause is said in many ways: 'Since many different things are called causes, it follows that many different things can all be causes . . . of the same thing' (195^{a}3–5). Like a number of parallel claims in the corpus—about being, goodness, health, etc.—this statement need not be interpreted as a claim about the meaning of the word 'cause' (*αἰτία* or *αἴτιον*), but can be read more straightforwardly as an assertion about how a variety of things are said to *be* causes.[19] Aristotle lists these things at 195^{b}12–16:

> All this comes to six things, which may each be spoken of in two ways. There is the particular and the general, the accidental and the genus to which the accidental belongs, and these may be given singly or in combination. (Charlton's translation, altered)

Now presumably the first-mentioned 'particular and general' causes are meant to be proper, as opposed to accidental, causes. Aristotle does not himself supply a designation here for this contrasting concept; I shall proceed to describe such causes as 'intrinsic'.[20] We may review examples of what he has in mind, as follows:

Intrinsic causes of a statue, taken alone:

particular: this sculptor
generic: an artist

Accidental causes of a statue, taken alone:

particular: a man (Polyclitus)
generic: an animal

Combined causes:

particular: Polyclitus the sculptor
generic: an artistic animal

[19] I have in mind G. E. L. Owen's influential discussion of Aristotle on 'focal meaning'; his view is criticized by T. H. Irwin in 'Aristotle's Concept of Signification', in Schofield and Nussbaum (eds.), *Language and Logos*.

[20] I borrow this terminology from Susan Sauvé, 'Self-Movement, Character, and Moral Responsibility in Aristotle's Ethics', unpublished paper, 1988; 'intrinsic' seems preferable to 'essential' here, since the latter has misleading connotations (that there's just one cause of something; that being a cause is essential to something). Annas does use the term 'essential' here; see Annas, 'Inefficient Causes', 322, and also her discussion of 195^{b}13, 322 n. 33. Charles also uses this way of speaking, *Philosophy of Action*, 46 ff.

Aristotle's notion of an accidental cause in this chapter appears to be grounded in two other notions: (1) some concept of essential or intrinsic causes, and (2) the metaphysical relation of accidental unity. Properly speaking, he tells us, a sculptor is the cause of a statue. Suppose that this sculptor is pale, musical, or named 'Polyclitus'; in Aristotle's metaphysics, the sculptor, pale man, and Polyclitus are related in a relation called 'accidental unity'. This permits us to describe the pale man, and so on, as causes accidentally (κατὰ συμβεβηκός) of the statue.[21] He tells us at $195^{b}6$–10 that there may be accidental effects as well. There is an illustration in *Met. E* 2:

> Again a confectioner, aiming at giving pleasure, may make something wholesome, but not in virtue of the confectioner's art; and therefore we say 'it was an accident', and while there is a sense in which he makes it, in the unqualified sense he does not. ($1027^{a}3$–5)

Suppose that the tasty treat made by the confectioner's art happens to be something wholesome, say, an oat-bran muffin with raisins. What has the confectioner's art actually made?—something tasty and tempting. It just so happens that this is (1) a bran muffin and (2) something good for you. The relation of accidental causation devolves upon first, an intrinsic causal tie between the confectioner and the tasty treat, and second, the accidental unity on the effect side, between the tasty treat and the wholesome bran muffin.

Aristotle's account of intrinsic or essential causal relations does not require that either the cause or the effect side of a relationship be described by referring to what we would call the essence of the thing itself. To be a sculptor is not the essence of the man named 'Polyclitus'; nevertheless, it is as a sculptor that this man is in an intrinsic causal relation to the statue he makes. Neither is being a tasty treat the essence of an oat-bran muffin. Further, what Aristotle's metaphysics counts as an accidental conjunction may not be an 'accident' in our ordinary sense of the term; there may be a perfectly good explanation for Polyclitus' being a sculptor, as for the tasty treat's being a bran muffin. The relation between the accidental and the random or unpredictable will occupy me in section III below.

The key feature of the accidental is that it is not regular or predictable. Aristotle characterizes it in *Met. E*, 2 as 'what is neither always

[21] In what follows I will interchange various ways of describing this relationship, assuming they can be unproblematically transformed: '*x* is the accidental effect of *y*'; '*y* causes *x* accidentally'; and '*x* and *y* stand in a relation of accidental causation'.

nor for the most part' (1026^b31–3). This remark provides a valuable clue about the nature of intrinsic or καθ' αὑτό causal relations; they *do* obtain either always or for the most part. Such regular and orderly connections, even merely usual, not exceptionless ones, depend upon the natures of the connected substances. For example, trees are said 'usually' to shed their leaves if their sap coagulates at the leaf-stems (*An. Post.* II. 16). Aristotle seems to believe that even exceptions themselves must admit of some explanation.[22] We would say that such καθ' αὑτό relations, by contrast to accidental relations, are natural and necessary relations. They hold in the real world, and statements about them are entailments—they support counterfactuals. These relations in the world obtain between universals of some sort, probably properties or dispositions, rather than between individuals (although of course within Aristotle's ontology such properties do not exist apart from individuals).[23] That Aristotle himself seems to think along these lines is suggested by the fact that when he does identify an individual (e.g. Polyclitus) as the intrinsic cause of something, it is *as* the possessor of a certain sort of power or skill. He writes,

> As elsewhere, so here, we should always look for the first [ἀκρότατον] cause of each thing. Thus a man builds because he is a builder, and a builder builds in accordance with the art of building; the art of building is, then, the prior cause, and similarly in all cases. (195^b21–5; Charlton's translation, altered)[24]

This advice appears to direct us to look at objective relations among properties in the world in identifying the most precise cause of something. It would apply concerning various of the four causes, for example, my frying pan conducts heat because it's made of iron, and iron conducts heat.

We can attempt to find some clarification of Aristotle's views by studying what he says elsewhere about καθ' αὑτό connections. However, this leads into a significant circularity. In his explication of the notion of καθ' αὑτό connections in the *Posterior Analytics*, Aristotle

[22] This is what I take to be the point of his example at 1027^a21–6; but the example is obscure, and so I may not be right. There are important difficulties about the relevant nomic connections that obtain within the Aristotelian universe in cases where a causal relation obtains usually ('for the most part') but not necessarily; I cannot pursue these here.

[23] For arguments in support of this presumption, see D. M. Armstrong, *What Is a Law of Nature?* (Cambridge, 1983), II.

[24] Presumably he means here that the builder builds in virtue of his possession of the art of building; he possesses an instance of this property. The universal art of building builds nothing, nor does it exist independently, according to Aristotle.

actually refers to causal relations. Thus, he says, while on the one hand a triangle is καθ' αὑτό a plane figure, and white is καθ' αὑτό a colour,

Again, in another way what belongs to something *because of itself* belongs to it in itself [καθ' αὑτό]; and what does not [belong] because of itself is accidental—e.g. if it lightened when he was walking, that was accidental; for it was not because of his walking that it lightened, but that, we say, was accidental. But if because of itself, then in itself—e.g. if something dies while being sacrificed, it died in the sacrifice since it died because of being sacrificed, and it was not accidental that it died while being sacrificed. ($73^{b}11$–16; Barnes's translation, altered; emphasis mine)[25]

That is, in the *Posterior Analytics*, Aristotle explicates the καθ' αὑτό relation in terms of causation, whereas in the *Physics* he does the reverse, explicating causation in terms of the καθ' αὑτό relation. I conclude that these two notions must be jointly illuminated.

The καθ' αὑτό relations Aristotle points to in analysing intrinsic causation must be various types of objective nomic connections. To see what these are like it is useful to consider more examples. Recall that accidentalness may apply to any of the four causes; for instance, if the iron in my frying pan is from West Virginia, then it is accidentally the case that my frying pan conducts heat *because* it's made of material from West Virginia. Accidental causation involving the final cause strikes us as more strange, but it clearly is possible within Aristotle's framework.[26] For example, Aristotle thinks that camels have four stomachs *for the sake of* digesting thorny food.[27] If I have a greenhouse with a desert room including a variety of thorny plants, then I might say that camels have four stomachs 'accidentally' *in order to* digest the sort of plants I have in my greenhouse.

The natural entailments or καθ' αὑτό relations missing in these cases of accidental causation are present in cases of intrinsic causation. There are natural and necessary connections between being a sculptor and producing statues, being made of iron and conducting heat, being round and being able to roll, or having four stomachs and digesting

[25] A substantive issue can be raised concerning the translation of σφαττόμεν in this passage: is Aristotle describing an intrinsic connection between being sacrificed and dying that should be understood as definitional, or is it (what we would call) a causal link between having one's throat cut and dying? I interpret it as the latter.

[26] Some of the most interesting examples involve human actions in which the agent is said not to know the act's real purpose. In these cases, what the agent knowingly does is not the same as what actually—and accidentally—happens. See my discussion in 'Aristotelian Actions', *Noûs* 19 (1985), 400–1.

[27] See *PA* III. 14.

rough food. In each case, of course, Aristotle's own analysis of the relevant καθ' αὑτό connection in the world is in terms of natures or powers. Iron has a nature comprised of some particular combination of the four fundamental powers, the hot, cold, moist, and dry. Hence, whatever is made of material having this particular nature is such that it will necessarily behave in certain ways given certain circumstances: it will conduct heat, it will sink in water, it will be attracted to magnets, etc. Similarly, even though it is not in the nature of a person to be a sculptor, sculpting is a sort of quasi-nature in virtue of which a person has the power to make statues given appropriate conditions, including desires and aims. Aristotle even speaks in this way of doctors or builders acting 'by nature' (πέφυκε), in *Met. E.* 2 (1027^{a}1). And again, it is in the nature of a stomach that it serves the purpose of digesting food, as it is in the nature of a horn that it is a means of self-defence.[28]

This is not the place to launch a full-scale investigation of Aristotle's concepts of powers and natures.[29] What is important to keep in mind is that these concepts play a role in his thinking about causation by grounding objective entailment relations of quite distinct sorts—not just ones that we think of as causal (i.e. efficient-causal) relations. He describes these grounding relations as various types of καθ' αὑτό connections. We on the other hand might express this by saying that in Aristotle's view the world contains four broadly different types of causal laws. Some natural laws concern how things with certain shapes will behave in particular circumstances (circular wounds will heal more slowly).[30] Other natural laws concern how things with certain material compositions and constituents will behave. Perhaps most strangely to our minds, Aristotle also must believe that there are natural laws articulating the necessary links between specific features of substances and certain purposes. For example, camels have four stomachs for the sake of concoction or good digestion. This is an instance of a more general law: natural creatures with digestion as an end will possess some means to this end. An even more general connection underlies this one: nature dispenses to living beings the means to achieve their ends.

[28] Aristotle discusses horns at some length in *PA* III. 2. Some animals have appendages resembling horns, but these are not horns because they do not serve the function of self-defence.

[29] I have discussed these notions in 'Aristotle on Possibilities and Capacities', *Ancient Philosophy* 6 (1986), 69–89.

[30] This is an example I borrow from van Fraassen; see 'A Re-examination', 40.

I believe that the evidence I have just been considering is enough to show that Aristotle is a causal realist, that is, he holds that causality is an objective feature of the world. Actually, I doubt that defenders of the explanation approach ever meant to deny this point—that they ever meant to argue that, understood in terms of explanations, Aristotelian *αἰτίαι* are merely subjective factors of our experience. Where my interpretation differs significantly from what has become the standard view is in my insistence that Aristotle is a causal realist about four distinct kinds of causal relations. This is an outcome that is fudged or resisted by adherents of the explanation approach, even those who speak of explanatory factors as real features of the world.[31] Moravscik, for example, after explicitly recognizing that there are (at least) four distinct types of explanatory factors, further inquires about what, in the Aristotelian ontology, grounds this belief. His answer is that this typology reflects aspects of Aristotelian substances.[32] This does not seem helpful, for it simply displaces the problem on to substances. My own view is that it is not especially illuminating to refer to explanatory factors, even if they are aspects of substances; such factors are explanatory because they are involved in explanatory relations that are themselves real. As I will argue in more detail immediately below, Aristotle is a realist about explanations as well as about causes. That is, differences in types of explanation are objectively grounded in different types of relation in the world. Specifically, I take it that they are grounded in the four kinds of causal relations identified in *Phys.* II. I have further suggested that these determinate objective relations are distinct types of *καθ' αὑτό* relations between properties.

II Explanatory Realism and Explanatory Knowledge

I have just tried to show that Aristotle believes there are different kinds of intrinsic causation precisely because there are different types of *καθ' αὑτό* connections in the real world. My interpretation imputes

[31] Fine writes that '[Aristotle's] explanatory factors are not (usually) sentences describing things, but real features of the world', 'Forms as Causes', 71.

[32] Moravcsik, 'Aristotle on Adequate Explanations', 5. This position is developed in complex ways in Moravcsik's 'What Makes Reality Intelligible?', this volume; more specifically Moravcsik now talks about the role played in causation by attributes, powers, and activities.

causal realism to Aristotle in the specific sense of saying that there are in reality four distinct kinds of causal ties. It is a short step to showing that Aristotle is also an explanatory realist. Explanatory realism is a position that holds that

> *C* is an explanans for *E in virtue of the fact* that *C* bears to *E* some determinate objective relation *R*. Let us call *R*, whatever it is, an 'explanatory relation'.[33]

As Jaegwon Kim has pointed out, 'the obvious first thought' about the identity of the relevant objective relation is that it is the causal relation.[34] Since Aristotle holds that there are four varieties of determinate objective causal relations, it is natural to interpret him as an explanatory realist who also holds that explanation statements refer to these objective relations. What is significant is that explanations cite real features of the world rather than psychological or doxastic states. The four kinds of καθ' αὑτό relations in virtue of which causal explanations are true are themselves 'out there' in the world. I do not think that proponents of the explanation approach (whether in the explanatory factors or the pragmatist version) have explicitly ascribed this form of explanatory realism to Aristotle; and, given my account of their negative motivation, this would be unlikely, for they are intent instead upon showing that the four kinds of explanation ultimately all refer to diverse aspects of events related by one real ('determinate objective') relation in the world, that is, efficient causation, understood however 'we' understand causation now.[35]

Two major problems arise for my interpretation of Aristotle as an explanatory realist. First (as I noted in my introduction) it will be important to offer some view of how the theory of causal explanation in *Phys.* II fits with Aristotle's theory of scientific explanation in the *Posterior Analytics*. And second, it is necessary to get clearer about Aristotle's views on the nature of explanatory knowledge (what form

[33] Kim, 'Explanatory Realism', 226 (author's emphasis). Kim thinks it likely that there are different kinds of explanation, and hence that an explanatory realist is committed to maintaining that there are different kinds of objective correlates to the explanation relations. (He mentions 'the relation of supervenience' and 'the micro-reductive relation' as examples; 238 n. 2) I think Kim, like most of us today, would find it extremely odd to hold that there could be different objective correlates to explanation relations that all count as causal relations.

[34] Kim, 'Explanatory Realism' 226.

[35] Moravcsik's version of the explanation view is different, and closer to my own (he speaks of four 'different entitative explanatory chains'; see 'What Makes Reality Intelligible?', this volume, p. 43).

does it take? do pragmatic factors enter in? etc.). The pragmatist version of the explanation approach offers attractive responses to both questions. As van Fraassen in particular has argued, Aristotle's reasons for distinguishing the four causes may reflect the significance of human interests: the distinction among four kinds of causal explanation in the *Physics* is 'pragmatic' in the sense that diversity among these sorts of causal explanation reflects facts about human contexts and inquiries. On this version of the explanation approach, Aristotle could be described as an explanatory realist in the *Posterior Analytics* without the commitment I ascribe to him—to saying that there are four various kinds of objective explanatory relations in the world. Instead, diverse explanations which arise from the human point of view in various situations all ultimately may be mapped on to one sort of objective correlate in the real world. Thus the pragmatic line seems to have some prima-facie plausibility. I want to examine it further as I seek to articulate Aristotle's views on the nature of explanatory knowledge. I think it will become clear that the pragmatic approach presupposes an account of scientific explanation that is significantly different from Aristotle's.

It is helpful to begin by locating the pragmatic approach within a framework of competing contemporary approaches to the nature of scientific explanation:

> Philosophical theories of scientific explanation can be roughly divided into three types: purely logical theories, which analyze explanation solely in terms of logical relations and truth conditions; theories with extra objective structure, which impose objective conditions on explanation beyond those of truth and logical structure; and theories with extra subjective structure, which impose on explanations psychological conditions of belief, interest, and so forth.[36]

The pragmatic approach belongs to the third, or subjectivist, category. Leading scholarly accounts of Aristotle's notion of scientific explanation have tended to interpret it as representative instead of the second, 'objectivist' group.[37] However, these accounts have focused primarily on the *Posterior Analytics* rather than the *Physics* or the

[36] Clark Glymour, 'Explanation and Realism', in Paul M. Churchland and Clifford A. Hooker (eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen* (Chicago, 1985), 104.

[37] Burnyeat's interpretation of Aristotelian explanation was explicitly based on a view articulated in Michael Friedman's article 'Explanation and Scientific Understanding', *Journal of Philosophy* 71 (1974), 5–19. Friedman is cited as a representative of Glymour's second category.

scientific treatises. Notoriously, Aristotle's scientific treatises do not exhibit the syllogisms we would expect of ideal sciences. Instead, in these treatises causal explanations are typically offered and clearly marked according to category or type. It might well be true, then, that in scientific practice, as in *Phys.* II, where Aristotle emphasizes the plurality of causes, he is utilizing a notion of explanation distinct from that set forth in the *Posterior Analytics*.

Aristotle's formal model in the *Posterior Analytics* provides a unified account of scientific explanation within which all four types of causal relation are to be represented by one specific deductive format. There is some debate about whether his formal machinery actually permits such reductive simplification; but for my purposes here what is important is that this was indeed his intention.[38] This formal model of explanation permits us to gloss over differences in types of causal or καθ' αὑτό connections. That is, though syllogisms are the proper means to represent any of the four different types of causal explanation, all types will be cited as middle terms in syllogisms having identical formal features. It will not be obvious on the face of a syllogism, so to speak, which kind of causation it describes. If scientific syllogisms differ at all, it will be in the grounds for asserting that the connections involved in each premiss are indeed necessary, i.e. in an epistemic (or perhaps pragmatic), not a syntactic feature.

What are the implications, then, for Aristotle's theory of explanation? On the one hand, it seems right to assert that the scientist who possesses a syllogistic demonstration of some proposition of the form 'All C is A' has in hand an explanation of C's being A—because B is A and C is B. The scientist knows these propositions, and knows in addition that they are appropriately linked in an organized body of knowledge to evident first principles. We need not refer to any extra propositions known, in addition to these, to explain what is meant in claiming that the scientist has an explanation of why C is A. Yet on the other hand, this account seems to leave out something very important: what *kind* of explanation does one have in hand, in having a scientific syllogism? That is, the scientist will know that C is A *because* of some middle term B; but the syllogism alone will not reveal what sort of 'because' this is.

The pragmatist line proposes an account of the difference between science in its idealized form and science put to use in explanation by

[38] Barnes points out his difficulties in trying to accomplish this end; see Barnes, *Posterior Analytics*, 214–23.

stressing a sharp distinction between the descriptive and explanatory roles played by scientific statements. Whether a given statement is part of science or explanatory depends not upon its form or content but upon its context. A statement belonging to science proper can be cited in a certain context where it suffices in order to give an explanation. Van Fraassen gives the following example:

> [I]f you ask a scientist to explain something to you, the information he gives you is not different in kind (and does not sound or look different) from the information he gives you when you ask for a description. Similarly in 'ordinary' explanations: the information I adduce to explain the rise in oil prices, is information I would have given you to a battery of requests for description of oil supplies, oil producers, and oil consumption. To call an explanation scientific, is to say nothing about its form or the sort of information adduced, but only that the explanation draws on science to get this information (at least to some extent) and, more importantly, that the criteria of evaluation of how good an explanation it is, are being applied using a scientific theory.[39]

Explanation is always extra-scientific, in the sense that it involves the application of a part of science to answering specific questions that have arisen in a particular context outside of science.[40]

Van Fraassen's pragmatic line offers a plausible suggestion about why Aristotle distinguished the four causes, namely, at least in part to provide a solution to certain notorious problems about the asymmetries of explanation.[41] A statement, for example, that 'walking after dinner contributes to non-regurgitation of food' might well be the middle term of two distinct scientific syllogisms. This proposition is a bit of scientific knowledge in either case, but it can only be cited as an explanation when the context clarifies the nature of the inquiry, and hence the nature of the causation revealed by this relevant middle term. In one context this statement could be cited as the explanation which answers a person's question about why Socrates has good digestion (because he takes a walk after dinner), whereas in another context it could instead be the answer to the question about why Socrates is taking a walk after dinner (because he wants good digestion). In the first context this premiss supplies the efficient cause, the second, the final cause—but in either case the proposition functions in the syllogism in the same way. Instances of each of the four types of

39 van Fraasen, *The Scientific Image*, 155–6.
40 In discussing this particular topic I am indebted to my colleague William Austin.
41 van Fraassen, 'A Re-Examination', 28–31.

scientific causal explanations may be cited to respond to diverse questions having particular presuppositions, contexts, contrast-classes, etc.

On the pragmatist interpretation, Aristotle argues that explanations that answer the four types of causal why-questions involve the application of bits of science (perhaps in the form of scientific demonstration) to answer specific questions in particular contexts: 'To sum up, explanation is an application of scientific knowledge; and it is the "telling" application, for, Aristotle maintains, you do not have scientific knowledge of something unless you know the "why" of it'.[42]

But van Fraassen's way of distinguishing between the intra- and extra-scientific is plainly not Aristotle's. This is an extremely important point. I think it is clear that in his *Phys.* II discussion Aristotle aims to highlight certain differences among the four causes, while in the *Posterior Analytics* he attends to common features of their formal (i.e. syllogistic) representability. But even though in his formal model of science Aristotle does not develop adequate machinery to emphasize and clarify differences among the four kinds of causal explanation, in the *Posterior Analytics* he does clearly take it to be a part of science to reveal causes—and he must mean by this to reveal the right *sort* of cause. Otherwise one would not have a 'telling' explanation, to use van Fraassen's term, even when one could supply the relevant syllogism. Further, in his actual scientific practice Aristotle insists very clearly on separating and categorizing explanations citing various from among the four causes, as recommended in the *Physics*. This is especially evident in the biological works, where he emphasizes differences between purposive accounts and efficient or material causal explanations. There is no reason to believe that he regards recognition of such differentiations among types of causal statements as 'extra-scientific'.

Part of the difficulty here, I take it, involves the difference between the *Posterior Analytics* syllogistic model of science and the actual Aristotelian sciences we have. The *Posterior Analytics* envisages explanation as, in effect, identical with intra-scientific citation of scientific demonstrations. Explanation requires knowing that certain first principles are appropriately linked with syllogistic deductions. This account obscures differences among the four types of causal explanation, envisaging an ideal possibility of formulating accounts of all four types of causes by citing them indifferently as middle terms. To know which type of

[42] Ibid. 28.

explanation a given syllogism affords might then seem to be something 'extra', a subjective not an objective feature. But it is doubtful that Aristotle would see it this way. I suspect instead he imagines that when scientists grasp middle terms they can tell which sort of causes they thereby capture. So although van Fraassen's reading is helpful to some extent, in that it fastens on to significant differences between Aristotelian sciences and their formal model, it does not get things right about either. For Aristotle, our diverse interests are generated by the world itself, and diverse kinds of causal statements are about relations that obtain in the world itself. Neither the interests nor the relations can be explained as an artefact of a specifically or idiosyncratically human curiosity, as van Fraassen's pragmatic line would propose. Furthermore, even if these different interests are not readily visible in the actual structure of the body of fully formulated scientific knowledge itself (that is, in the set of syllogisms that would constitute a completed science, if there were such a thing), they must be understood to be built into its truth-conditions—and by this I intend something semantic, not pragmatic.

III Accidental Causes

So far I have been discussing Aristotle's views about causation and causal explanation involving genuine or intrinsic causes. I wish now to focus instead on accidental causes. I have argued that Aristotle is both a causal and an explanatory realist, but this interpretation might seem to be stretched beyond the limits of plausibility when we come to accidental causes. The range of the accidental is indefinite, disorderly, seemingly chaotic; could Aristotle nevertheless really have believed that accidental causal relations are an objective feature of the world?

The most important role played by accidental causes in Aristotle's philosophy is in grounding his metaphysical account of chance and related phenomena. In *Phys.* II Aristotle moves directly from discussing the four causes in chapter 3 to a discussion of spontaneity and chance, in chapter 4. This is not surprising, for he regards chance as itself a kind of cause. (You can, after all, explain something's happening as due to chance: 'Why did the stone urn fall off the roof just now? By chance or because someone pushed it?') More specifically, chance is a certain kind of accidental cause:

> Things do, in a way, occur by chance, for they occur accidentally and chance is an accidental cause. But strictly it is not the cause—without qualification—of anything (197^a12–14).

This last remark is misleading in two ways. That chance doesn't 'strictly speaking' cause anything doesn't mean that there aren't things that happen by chance:[43] a horse that runs away happens thereby to escape a fire; Oedipus, in trying to defend himself, kills his own father. Aristotle's remark also misleadingly suggests that chance events are uncaused. If chance is a special type of cause that accounts for events happening when there is otherwise no underlying (i.e. no essential) causation at all, then a defence of chance would be related to a defence of indeterminism. But I want to argue that chance does not involve causal indeterminism any more than does accidental causation generally. Recall, for example, that if a doctor builds a house, there is a cause for the house's being built, though the doctor builds it, as it were, by chance.

However, Aristotle himself seems to link chance to indeterminism in an argument in *Met. E*. 3, where he supplies an example involving chance to argue that it is not the case that all the consequences of our actions are determined in advance. His example involves a chain of events leading up to a man's accidental death: the man eats spicy food, gets thirsty, goes out to a well to drink, encounters a band of ruffians there, and is murdered. The man is murdered by chance; eating spicy food was the accidental cause of his being murdered. (There is no intrinsic causal link between eating spicy food and being murdered.) Aristotle appears to argue that chance events like this one involving accidental causes are undetermined: it is (fortunately) not a foregone conclusion that if one eats spicy food one will be murdered.

This example and its implications are important to consider, for they seem recalcitrant to analysis either in terms of the explanation approach to causation or on my own favoured approach, which insists on Aristotle's explanatory and causal realism. I will offer an interpretation of this passage; but first I want to examine an alternative interpretation of this text that proceeds on the assumption that for Aristotle causes are explanations.

[43] Strictly speaking, chance (τύχη) applies only to certain accidents in the realm of human endeavours, and the spontaneous (τὸ αὐτόματον) is a broader notion not confined to human agents. I use the former term more broadly here for expository convenience. See Lindsay Judson's paper in this volume for further clarification of these notions in relation to Aristotle's account of causes.

Richard Sorabji offers such an interpretation in *Necessity, Cause, and Blame*, reconstructing Aristotle's argument as follows: Aristotle reasoned that, since causation is explanation, then all events that have causes (αἰτίαι) can be explained. Since coincidences lack explanations, they also lack causes. Hence, Aristotle concluded, coincidences are not determined. Sorabji faults this argument, pointing out that if causation is really determination rather than explanation, to say that an event lacks a bona fide explanation does not suffice to show it is undetermined:

> ... [A] determinist could rightly protest that necessitating, i.e. sufficient, conditions are unlike explanations ... [If there are sufficient conditions for each arrival at the well] then it must be conceded that there are sufficient conditions for what is entailed by this, namely, their being at the well simultaneously.[44]

Clearly Sorabji's critique of this argument depends on at least two points: (1) that it is intended as an argument against causal determism, and (2) that it relies on an understanding of causes as explanations. I have argued earlier in this paper against (2), but I have also said things that render (1) problematic. Sorabji assumes that Aristotle is primarily worried about threats to human freedom posed by a determinism of mechanical (or efficient) causes. But if I am right that he is a causal realist about four distinct types of causal relation, then his worry might just as well be about a forward-looking intrinsic causality, or teleological determinism. Indeed, this makes more sense in relation to his example: he's saying it would be absurd to think one ate food or went to a well *in order to be* murdered.[45]

If chance is like other instances of accidental causation—there are often intrinsic causes for events Aristotle describes as due to chance—then it does not figure into the world by intervening in the causal order; chance is not a source of causal indeterminism.[46] But what is it

[44] Sorabji, *Necessity, Cause and Blame*, 13.

[45] My suggestion is in accord with a rather different account of this chapter that was proposed by Dorothea Frede, who also faults Sorabji for focusing on accidents' interruptions of chains of efficient or moving causes. On her view, Aristotle's worry is not about 'the hegemony of necessitating efficient causes but rather the totalitarian regime of the τέλος'. See Dorothea Frede, 'Aristotle on the Limits of Determinism: Accidental Causes in *Metaphysics E* 3', in Allan Gotthelf (ed.), *Aristotle on Nature and Living Things* (Pittsburgh and Bristol, 1985) 207–25.

[46] Cf. Frede's comment: 'It would be a strange oversight on Aristotle's side had it escaped his notice that, according to his own example in the *reductio*, it is not just nature and the natural, as one would expect, that is "saved" from strict universal determinism by the existence of accidental causes but the contingency of human actions too! How does

exactly? Chance outcomes are accidents or exceptions that don't exemplify regular types of final causation; they involve 'as-if' purposes, processes which reach a bad or good end without really aiming at it.[47] Why not follow the pragmatic line, saying that there are certain of these ends that we attend to because they appear especially interesting, poignant, fateful, etc.? Getting murdered when you go to a well is one such case; similarly, we pity poor Oedipus, who chanced to kill his father when he killed a man. This question is especially acute given my interpretation of Aristotle's explanatory realism: in saying something occurred by chance we cite an explanation that is somehow odd or funny, and so it would seem especially forced to insist that there is a corresponding objective, determinate relationship out there in the world.

At this point we need to think more generally about accidental causes. When I explicated the Aristotelian notion of accidental causation earlier, I said that such a relation always depends upon two things; a proper or intrinsic causal relation, and an accidental unity. Aristotle emphasizes the former feature in *Phys.* II. 6, when he writes that 'Since nothing which is accidental is prior to what is καθ' αὑτό, it is clear that no accidental cause can be prior to a cause καθ' αὑτό' (198ᵃ6–8). In the *Metaphysics* Aristotle describes the case of a builder who accidentally produces health. Imagine this sort of case in more detail, modernized. At the scene of an emergency near a construction site, a man dressed in hard-hat and work clothes rushes out to perform cardio-pulmonary resuscitation (CPR) on a heart-attack victim. A reporter might cover the story, and headlines would read 'Construction worker saves life!' Now, the pragmatist emphasizes that this accidental relation in effect comes into existence when we pick it out. It simply seems noteworthy (has 'human interest') in a way others might not. The newspaper reporter regards the accident of a hard-hatted labourer's knowing CPR and being able to save a life as more newsworthy than numerous other accidental effects which he ignores—e.g. the heart-attack victim's squashed lunch, the labourer's mislaid hammer, etc.

Still, accidental causes are really 'out there' in the world, since

this square with his tenet that choice and the ensuing actions are up to us?' (Frede, 'Limits of Determinism', 207).

[47] '[I]f it were not for such interferences [viz.: by accidental causes] one could not maintain a distinction between the *intended* end and the end *actually* achieved'. Frede, 'Limits of Determinism', 221 (author's emphasis).

intrinsic causal relations and accidental unities in which they are grounded are objective features of reality. There must be some basis in reality for the assertion of accidental causation in a case like this: it was after all the construction worker and not someone else who saved the person's life. Presumably he did so because he happened to have learned CPR; it was in virtue of this that he saved a life, not because he was a construction worker. That is, there is an intrinsic causal relation here between an agent who knows a medical procedure and an act of healing, and there is also an accidental unity, between the construction worker and the possessor of medical knowledge.

My rejection of the pragmatist's interpretation of accidental causes has a parallel in contemporary disputes among philosophers of science. In discussing a variety of people's interests in and views about the causes of a carriage accident, N. R. Hanson commented:

> There are as many causes of *x* as there are explanations of *x*. Consider how the cause of death might have been set out by a physician as 'multiple haemorrhage', by the barrister as 'negligence on the part of the driver', by a carriage-builder as 'a defect in the brakeblock construction', by a civic planner as 'the presence of tall shrubbery at that turning'.[48]

In this example, extra-scientific interests determine the part of the causal network someone is concerned about. The civic planner worries about obstructions to views on the highway, whereas the manufacturer worries about brake failure. Not surprisingly, van Fraassen finds here grist for his pragmatic mill. The pragmatist maintains that explanations of the accident are relative to human presuppositions and inquiries.

On the pragmatist line, there are a variety of explanations of the carriage accident that do not necessarily refer to determinate relations in reality. And this might similarly seem to hold for Aristotle's treatment of accidental, as opposed to intrinsic, causation. But, as should be clear from the carriage accident example, the physician or the manufacturer *are* looking for real explanations in terms of genuine or intrinsic causes.[49] Although their interests in diverse aspects of the

[48] N. R. Hanson, quoted by van Fraassen in *The Scientific Image*, 125.

[49] In criticizing van Fraassen's account of this example, Wesley Salmon has argued as follows: 'We can see readily how contextual factors are involved in the correct identification of the why-question that is used to call forth a scientific explanation. It obviously does not follow ... that once we have ascertained by reference to pragmatic and contextual factors what explanation is being sought, the explanation itself must embody pragmatic features.' Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the*

accident may seem, like the newspaper reporter's, to reflect pragmatic considerations, the causal stories they point to are genuine.[50]

I want to say that when we focus on accidental effects and causes such as these, we are simply selecting from an indefinite number of accidental causes and effects the ones which we find especially striking or which answer to particular interests of ours. But the ones we do not pick out are equally objective. When confectioners make tasty treats, they also produce items that may be wholesome, an accident we notice because it is important to us. But they also produce things having a certain colour, weight, size, and shape, a particular price and weight, and a specific number of raisins. They make things that my father dislikes, though he knows they're good for him.

In the world at any time an indefinite number of such accidental causal relations really exist; perhaps this is just what Aristotle means when he speaks of accidental causes as 'indefinite' (197^a9). Hence there can be no Aristotelian science of either chance or accidental causes. Still, Aristotle recognizes the important role of the accidental in his account of causation, as in his metaphysics in general, and in so doing he allows for the more quirky aspects of human interests in the world. What is strangest about this doctrine, I believe, is that these are features of the world that are only noticed and significant for us humans, creatures who make our way about in the world by observing, classifying, and discoursing on accidental unities (kooky objects). Aristotle's God, of course, thinks only of itself; but even the lesser intellects of the star-souls would surely not conceive of the world or of its connective fabric via the accidental. They would have no use for a notion of accidental causation. In a way, humans, who pay more attention to chance and the accidental, know more about the universe than God.[51] But what we know is not fit to be deemed science; in other

World (Princeton, 1984), 130. (I am grateful to William Austin for drawing this criticism to my attention.)

[50] This has an important consequence for the practice of science. As scientists develop better accounts of the causes of a phenomenon, they acquire better explanations. For further discussion of the possibility of launching a scientific inquiry from a merely accidental grasp of the explanandum, see Robert Bolton, 'Essentialism and Semantic Theory in Aristotle: *Posterior Analytics* II. 7–10', *Philosophical Review* 85, 514–44. Problems parallel to those I have been discussing here about Aristotle's notion of explanation have been raised in modern discussions of evidential relations in science. See Helen Longino, 'Evidence and Hypothesis: An Analysis of Evidential Relations', *Philosophy of Science* 46 (1979), 35–56.

[51] Cf. van Fraassen: 'It is sometimes said that an Omniscient Being would have a complete explanation, whereas these contextual factors only bespeak our limitations due

words, it is not really knowledge. The interests of people, in so far as we strive to be truly rational animals, will presumably direct our inquiries to proper causal connections, supplanting the accidental ones which remain, after all, 'inscrutable to man' (197^a9–10).[52]

to which we can grasp only one part or aspect of the complete explanation at any given time. But this is a mistake. If the Omniscient Being has no specific interests . . . and does not abstract . . . then no why-questions ever arise for him in any way at all—and he does not have an explanation in the sense that we have explanations' (*The Scientific Image*, 130).

[52] This quotation is from a passage in which Aristotle is only discussing 'what men think', but he seems to agree.

4

Chance and 'Always or For the Most Part' in Aristotle

LINDSAY JUDSON

This paper has two aims. One is to examine Aristotle's analysis of chance in *Phys.* II. 4–6; the other is to try to elucidate a concept which has been very much neglected by commentators despite its central importance to a number of Aristotle's doctrines and theories—his concept of 'happening always or for the most part'. These two aims go hand in hand, because the analysis of chance invokes the concept of happening always or for the most part, and also provides important evidence as to what this concept is. I shall begin with the analysis of chance (sections I–II), then turn to 'happening always or for the most part' (sections III–VI), considering along the way what light the former sheds on the latter. In pursuit of these broad questions a number of points of detail raised by the analysis in II. 4–6 will have to be ignored.

I

'Chance' is intended as a translation of τὸ αὐτόματον, the main *analysandum* of the *Physics* discussion. It does not capture the idea of τὸ αὐτόματον perfectly, since for us 'chance' applies to a much wider range of cases: events could be ascribed to chance when they satisfy Aristotle's other conditions for being ἀπὸ τοῦ αὐτομάτου, but fail to be ἕνεκά του in the sense in which the products of Aristotelian chance must be—that is, when they do not benefit anyone (see section II below).[1] None the less 'chance' comes closer to Aristotle's concept than any other translation.[2] Although elsewhere Aristotle frequently

[1] Although he does not discuss the point, Aristotle of course has a classification at his disposal which would cover these cases, namely that of accidents (συμβεβηκότα).

[2] Recommending the translation 'spontaneous', Guthrie writes: '*Tyche* was the common word for chance, luck or fortune ... *Automaton* had much in common with its

uses τύχη in the same sense as τὸ αὐτόματον, in *Phys.* II. 4–6 he generally gives it a much narrower sense, so that 'things due to τύχη' form a subset of 'things due to τὸ αὐτόματον'; I shall translate τύχη as 'luck'.[3]

Aristotle has a twofold purpose in introducing chance at this point in *Phys.* II. One purpose is forward-looking: Aristotle wants to go on to discuss the claims of other physicists that certain things are due to chance—the atomists' claim that the ordered cosmos is due to chance (II. 6, 198^{a}5–13), and Empedocles' (supposed) claim that the arrangements of the parts of animals are due to chance (II. 8). Commentators tend to see this forward-looking purpose as paramount—and sometimes berate Aristotle for tailoring his analysis to suit this 'hidden agenda'.[4] But although the forward-looking aim is clearly important to Aristotle, these complaints are unwarranted. Aristotle's analysis is

modern derivative 'automatic', as used e.g. in Homer of the tripods invented by Hephaestus, which ran of themselves' (*A History of Greek Philosophy* vi (Cambridge, 1981), 236 n. 1). Now things which happen 'of themselves' do share a feature with τὸ αὐτόματον as Aristotle analyses it in the *Physics*: they imitate the workings of some natural or deliberative subject (see sections II and IV, below). But there the resemblance ends, for they do this in virtue of some mechanism (usually internal to their subject), which operates in a regular, repeatable, and predictable way—this is the case for Hephaestus' tripods and the αὐτόματα described by Aristotle at *MA* 7, 701^{b}1–10. Events which happen ἀπὸ τοῦ αὐτομάτου, on the other hand, are on Aristotle's analysis anomalous, irregular, and opaque to science. So I take αὐτόματος to be ambiguous: it simply does not mean the same in *MA* 7 and in *Phys.* II. This point is worth stressing for another reason. It is a familiar complaint that Aristotle's account of spontaneous generation in the biological works is inconsistent with his account of spontaneity in the *Physics* (see D. M. Balme, 'Development of Biology in Aristotle and Theophrastus: Theory of Spontaneous Generation', *Phronesis* 7 (1962) 91–104; cf. Augustin Mansion, *Introduction à la physique aristotélicienne*, 2nd edn. (Louvain and Paris, 1946), 310 and n. 54). If αὐτόματος is ambiguous, as I claim, the problem of wholesale inconsistency simply does not arise; indeed, the arguments for the inconsistency are just further arguments for the view that αὐτόματος is ambiguous. (Note, however, that generation ἀπὸ τοῦ αὐτομάτου (καὶ ἀπο τύχης) in *Met.* Z (7, 1032^{a}28–32 and 9, 1034^{b}4–6) seems to be *chance* generation—exceptional cases in which an animal normally generated from seed is generated without seed—and not the regular spontaneous generation discussed in the biological works.)

3 I shall concentrate on Aristotle's analysis of τὸ αὐτόματον in general, and ignore the interesting question of his distinction between τὸ αὐτόματον and τύχη. Like τύχη, 'luck' has a narrower extension than 'chance', and is most appropriate when used in connection with a special class of subjects—persons—which is reasonably close to the class to which Aristotle's τύχη is confined (subjects capable of προαίρεσις). Within II. 4–6, Aristotle uses τὸ αὐτόματον in two ways: in a wider sense which includes luck, and in a narrow sense which covers chance things which are not due to luck. Throughout this paper I use 'chance' for τὸ αὐτόματον in the wide sense; in nn. 50 and 52 below I use 'chance$_N$' for the narrower sense.

4 See D. M. Balme, 'Greek Science and Mechanism I: Aristotle on Nature and Chance', *Classical Quarterly* 33 (1939), 129–38; Guthrie, *History of Greek Philosophy* vi. 241–2.

actually guided by his other purpose, and quite naturally arises from the discussion of the four causes in chapter 3: Aristotle wishes to explain how chance fits into his account of causation.

That this is his main concern in chapters 4–6 is evident from the *endoxa* he selects. He does mention the physicists' claim about the cosmos, the refutation of which forms the final section of the discussion; but most of the *endoxa* relate to the question of the *causal role* of chance:

> Both luck and chance are said to be among the causes, and many things are said both to be and to come to be because of luck and because of chance. So we must consider in what way luck and chance are among these causes, and whether luck and chance are the same thing or different, and the whole question as to what luck and chance are. (195^b31–6)

The earlier physicists, Aristotle says, adopt two very different attitudes towards chance, each of which is problematic. Some say that nothing happens by chance; but this is absurd, as many things clearly do. But those who agree that some things happen by chance give no account of what it is; it is clear that most of them do not regard it as a type of cause on a par with 'such things as love, strife, mind, fire, or anything else of that kind' (196^a18–19), but if it is not, how can things happen as a result of chance? We seem driven back to the first arm of the dilemma.

Aristotle does not attempt to detect truth in all the *endoxa* which he describes: as I have mentioned, he thinks that those who hold that the cosmos is due to chance are simply wrong. Nevertheless his main purpose is the characteristic one of seeking a middle way which will show how most of the *endoxa* are partially true. In effect, his aim is to do justice to the idea that chance is *something* and to the idea that it is *nothing*. He does this by denying that chance is a causal force in its own right (or a fifth type of cause co-ordinate with the other four), while maintaining that there is none the less a sense in which events do happen by chance. In the process he demonstrates what is right about other *endoxa*, especially the sense in which chance is obscure: on Aristotle's analysis, chance *is* obscure, but not in virtue of being a mysterious or supernatural force (196^b5–7, 197^a8–21).

II

Aristotle begins his positive account of luck and chance in chapter 5 by pointing to the existence of two overlapping classes. The first is the class of things which come to be neither always nor for the most part. I shall speak of members of this class as happening *rarely*;[5] and for convenience, I shall use the term 'event' to cover the whole range of things which might be said to come to be either rarely or regularly—events, processes, states of affairs, activities. It is the existence of rare events which demonstrates the existence of chance:

> First, then, since we see some things always coming to be in the same way, and some for the most part, it is evident that neither luck nor what is due to luck[6] is said to be the cause of either of these—either of what is of necessity and always, or of what is for the most part. But since there are also things which happen in addition to these, and all say that these happen from luck, it is evident that luck and chance are something. For we know that such things are from luck and that the things due to luck are of this kind. (196^{b}10–17)[7]

The second class is that of events which happen 'for the sake of something'. We might be forgiven for taking this to be the class of events which have final causes; but as we shall see in a moment, Aristotle has a much wider class in mind.

Aristotle is now ready to give what we may call his first definition of chance:

> Now things of this sort [i.e. events which happen for the sake of something], whenever they come to be incidentally, we say are from luck. (For just as in the case of being also, there is being *per se* and being incidentally, so it is possible for something to be a cause.) . . . As has been said, therefore, whenever this comes to be [i.e. when something comes to be incidentally] among the things which come to be for the sake of something, then it is said ⟨to be⟩ from chance or from luck. (196^{b}23–31)

This gives two conditions which an event must satisfy to be the product of luck or chance: (1) it is 'for the sake of something', and (2) it 'comes to be incidentally'.

[5] Strictly speaking this class includes things which happen exactly half the time as well as those which are less frequent, and I intend such events to be covered by my term 'rare'.

[6] Simplicius' remark on 196^{b}10–17 applies to most of ch. 5: 'It is clear that, having not yet distinguished luck and chance, his argument concerns both, even though he only mentions luck' (*in Phys.* 334.35–335.1).

[7] For rarity as a necessary condition for chance, see also (e.g.) *An. Post.* I. 30, 87^{b}19–27, *Cael.* I. 12, 283^{a}32–^{b}1, *GC* II. 6, 333^{b}4–7, *EE* VIII. 2, 1247^{a}31–3, *Rhet.* I. 10, 1369^{a}32–^{b}5.

The first condition raises an obvious difficulty, since Aristotle standardly *opposes* 'by chance' and 'for the sake of something'.[8] It would be possible to interpret ἐν τοῖς ἕνεκά του γιγνομένοις (196^b29–30) merely as 'in the general class of things which are normally ⟨but need not be⟩ for the sake of something'; but this solution will not do. At 196^b34–6 he clearly refers to a chance event as being for the sake of something: ἦλθε δ' οὐ τούτου ἕνεκα, ἀλλὰ συνέβη αὐτῷ ἐλθεῖν καὶ ποιῆσαι τοῦτο τὸ τοῦ κομίσασθαι ἕνεκα;[9] similarly at 197^a1 he describes the benefit produced by the chance event as 'the τέλος'. Ἂν πραχθείη at 196^b22 also suggests that here ἕνεκά του applies to more events than it does in the standard sense, while in one and the same sentence at 197^b18–21, Aristotle says both that chance events are ἕνεκά του and that they do not come to be for the sake of what results. It is hard to resist the conclusion that Aristotle does wish to say that chance events are ἕνεκά του, and that he is using the phrase not in the familiar, explanatory sense, but in an unfamiliar, non-explanatory one.[10]

If this is so, what is the non-explanatory sense? Simplicius' interpretation is generally accepted: he takes '*E* is ἕνεκά του in the non-explanatory sense' to mean '*E* is such that it might have been ἕνεκά του in the explanatory sense' (*in Phys.* 335.33–336.5).[11] I prefer to take it to mean '*E* actually confers some benefit', i.e. '*E* is the *per se* cause of some benefit'.[12] There is little direct evidence bearing on the question,

[8] See e.g. *An. Post.* II. 11, 95^a8–9 (ἀπὸ τύχης δ' οὐδὲν ἕνεκά του γίνεται), *Phys.* II. 8, *PA* I. 5, 645^a24–5, *Rhet.* I. 10, 1369^a32–$^b5$.

[9] I follow William Charlton in reading τοῦτο τὸ τοῦ κομίσασθαι, since this makes the sentence rather less awkward by attaching the non-explanatory sense of ἕνεκά του to a description of the event, and not to the agent's performance (*Aristotle's Physics, Books I and II*, Oxford, 1970, 48–9). but the sentence would still associate chance with ἕνεκά του even if we retained the reading of the MSS. Deleting τοῦ κομίσασθαι ἕνεκα to avoid this (so Bonitz and the Oxford translators) is a counsel of desperation, as emendation would be needed in many other places in chs. 5–6 as well.

[10] This is convincingly argued by Jim Lennox in 'Aristotle on Chance', *Archiv für Geschichte der Philosophie* 66 (1984), at 52–6.

[11] W. D. Ross (*Aristotle's Physics*, (Oxford, 1936), 517–19), Charlton (*Physics I and II*, 106), and Lennox ('Aristotle on Chance', 56–60) all endorse some version of this view.

[12] There are three points to note here. (1) 'Benefit' is intended in the most general sense: *E* produces a benefit if it serves someone's interests, or satisfies a desire or need. (2) *E* will normally be an *efficient per se* cause of the benefit; but Cynthia Freeland has pointed out to me that there are cases in which it would be a formal cause. For simplicity's sake I ignore these cases in what follows. (3) Apart from the remarks at 197^a25–30, Aristotle's discussion ignores the idea of *bad* luck entirely. Mansion claims that Aristotle's analysis of luck does not leave room for bad luck, and that this is a sign of a substantial change of doctrine which he thinks occurred between an earlier discussion of chance and a later revision (*Introduction à la physique*, 307–8; cf. Balme, 'Greek Science and Mechanism I', 129). The first of these claims is formally correct, but the second does

but my reason for preferring this interpretation to Simplicius' is that in the only other context in which Aristotle appears to use ἕνεκά του in a non-explanatory sense, he does mean 'actually produces an end' and not 'is suitable for producing an end'. This context is the list of things ignorance of which prevents an action from being ἑκούσιον—a list which includes ignorance of the ἕνεκά του.[13] Ignorance of what end the action is *suitable* for producing is irrelevant to whether or not it is ἑκούσιον; what is relevant is ignorance of what end it will actually produce, and the *EN* v. 8 and *EE* passages confirm that this is Aristotle's meaning:

[Injuries] done in ignorance are *errors* when the person acted on, the act, the instrument or the ἕνεκά του is not what the agent understood; he thought . . . that he was not hitting . . . with this end, but ⟨the end⟩ which resulted was not what he thought the ἕνεκά του was—e.g. he threw not in order to wound but only to graze. (*EN* v. 8, 1135^b12–16)[14]

What does condition (2) in the definition of chance involve? Aristotle's remark about incidental causation at 196^b24–9 indicates that coming to be incidentally is a matter of being incidentally caused in some way, and that being incidentally caused is at least a necessary condition of coming to be incidentally. This idea is supported by the explanation Aristotle gives of the sense in which chance is ἀόριστον (see p. 79 below), and is confirmed both by the example of the lucky creditor at 196^b33 ff. and by the subsequent descriptions of luck and chance as incidental causes (197^a12–13 and 32–5; cf. 198^a5–10).

The concept of incidental causation raises many questions which I shall not attempt to pursue here.[15] Briefly, if we use 'E_i' as a place-

not follow. Clearly what Aristotle needs in order to be able to include bad luck is the idea of 'producing a benefit *or harm*'; and while the association of ἕνεκά του even in its non-explanatory sense with the notion of a τέλος certainly does not suggest that idea, it is easy enough to extend Aristotle's analysis in this way so as to cover bad luck as well as good. (Ross, *Aristotle's Physics*, 41, offers a different account of how bad luck fits into the Aristotelian analysis.)

[13] *EN* III. 1, 1111^a2–6 and 18–19; v. 8, 1135^b12–16; and *EE* II. 9, 1225^b1–5.

[14] For a discussion of these passages in slightly different terms, see Cynthia A. Freeland, 'Aristotelian Actions', *Noûs* 19 (1985), at 400–1. Note that producing an actual benefit seems to be a condition of being a lucky event in *our* sense: a man who stumbles upon treasure is neither lucky nor unlucky if he is indifferent to riches. But of course this is not true of our notion of chance, as I said at the beginning; and in any case it does not prove that it is Aristotle's condition also.

[15] For a discussion of some of the issues see section III of Freeland's paper 'Accidental Causes and Real Explanations' in this volume. Ascriptions of *per se* and incidental causation can be treated as highly intentional, so that E_1 may be a *per se* cause of E_2 under one set of descriptions of E_1 and E_2, but an incidental cause of E_2 under another; or as extensional but presupposing a very fine-grained ontology, so that these different descriptions

holder for events, '+' to indicate accidental concurrence in a subject, or 'accidental unity', and '⇒' to signify *per se* causation, then E_1 can be an incidental cause of E_3 in two different ways:

Schema I	Schema II
E_1	$E_1 \Rightarrow E_2$
$+$	$+$
$E_2 \Rightarrow E_3$	E_3

Aristotle's illustrations at $196^{b}26$–7 and $197^{a}14$–15 are cases of type I; in *Met. E.* 2 he gives examples of both types—a cook (aiming at giving pleasure) produces health in someone, a housebuilder (who happens to be a doctor) cures someone ($1026^{b}37$–$1027^{a}5$).

We can now see why chance is ἄδηλον and ἀόριστον ($196^{b}27$–9 and $197^{a}8$–14): its causes are ἀόριστα (cf. $197^{a}9$–10). Take any given event E_1 in schema I; since an unlimited number of types of event can attach to the subject of this event, an unlimited number of types of event can stand in the position of E_2, and hence an unlimited number of types of event can stand in the position of E_3 (and a similar argument holds for E_3, E_2, and E_1 respectively in schema II). Thus there is no finite specification of the possible incidental causes which a given type of event may have, and there is equally no finite specification of the possible incidental 'effects' which a given type of event may have. For this reason any putative 'laws' which connect types of events under incidental causation would simply be constituted by infinitely long disjunctions, and these could not in Aristotle's view be objects of knowledge.[16]

The *per se*/incidental distinction applies to all four causes (*Phys.* II. 3, $195^{a}3$ ff.); does Aristotle have a particular mode of causation in mind when he speaks of chance events coming to be incidentally? It is sometimes held that Aristotle takes luck and chance to be incidental *final* causes, but this is not so. Aristotle himself raises the question at $195^{b}33$–4 and $196^{b}8$–9, and says that it is *efficient* causation which is

in fact pick out different events. (And, of course, there are intermediate positions too.) How fine-grained Aristotle's ontology of events, etc. actually is makes no difference to the substance of his account of chance events as incidentally caused. Simply for convenience of presentation, I shall use the language of a very fine-grained ontology (in which, for example, 'the man's going to the agora' and 'the man's going to where his debtor is' in the story of the lucky creditor pick out different events which unite accidentally in the man).

[16] Note that nothing in this account rules out our being able to know *in a particular case* that a concurrence of such-and-such a sort is likely, and hence being able to predict the chance event itself. See further pp. 91–2 below.

involved: 'of the modes of cause, each of these [luck and chance] is in the sources of change' (198ᵃ2–3). Since luck and chance are elsewhere described as incidental causes, as we noted above, it is clear that the mode of incidental causation in question is incidental efficient causation. This is not to say that one cannot identify incidental causes of the other modes in cases of chance: in the example of the lucky creditor, for instance, the *per se* final cause of the man's going to the agora—seeing a spectacle, or whatever it might be—will stand as an incidental final cause of his meeting his debtor. But Aristotle makes it plain that 'coming to be incidentally' in the analysis of chance involves—and only involves—incidental efficient causation.[17]

It is an interesting feature of the way we talk about luck and chance that events which lead to a chance result—and the benefit which this chance result produces—can also be ascribed, in a derivative way, to chance or luck, and can themselves be called chance events (cf. Ross, *Aristotle's Physics*, 516–17). For this reason, let us distinguish the events which satisfy Aristotle's definition—i.e. which are chance events in the non-derivative sense—by calling them 'chance outcomes' (or 'lucky outcomes'). As we have seen, it is a necessary condition of an event's being a chance outcome that it have an incidental efficient cause. Employing the two schemata for incidental causation which I gave above, we can derive two causal structures in which chance outcomes can be embedded; in both, the chance outcome has E_1 as an incidental cause:

Schema I*	Schema II*
E_1	$E_1 \Rightarrow E_2$
$+$	$+$
$E_2 \Rightarrow \text{CO} \Rightarrow$ *telos*	$\text{CO} \Rightarrow$ *telos*[18]

[17] Holders of the view that chance is an incidental *final* cause include Lennox, 'Teleology, Chance, and Aristotle's Theory of Spontaneous Generation', *Journal of the History of Philosophy* 20 (1982) 219–38, e.g. at p. 233. Aristotle's explicit statement directly contradicts this. The only passage which might seem to favour the final-cause interpretation is Aristotle's remark at 197ᵃ1–2: ἔστι δὲ τὸ τέλος, ἡ κομιδή, οὐ τῶν ἐν αὐτῷ αἰτίων. I agree with Lennox that Aristotle is explaining why the lucky event satisfies the 'coming to be incidentally' condition. Clearly the word τέλος is suggestive of final causation, but τῶν ἐν αὐτῷ αἰτίων is equally suggestive of efficient causation; thus Aristotle could just as well mean that the *desire* for the end was not a *per se* efficient cause as that the end was not a *per se* final cause. What is more, at 197ᵃ17–18 Aristotle describes the possible incidental causes of the lucky event as desires for various ends.

[18] 'CO' is a place-holder for the chance outcome, and '*telos*' for the benefit; as before, '+' signifies concurrence in a subject, or accidental unity, while '⇒' now signifies *per se efficient* causation.

The analysis of chance as incidentally caused enables Aristotle to avoid denying the existence of chance or having to take chance events to be altogether random (i.e. with only forwards connections with the causal order), and yet also enables him to avoid taking chance to be a causal force in its own right.

Now that we have a sketch of incidental causation, we can ask, what is it for an event *E* to come to be incidentally? Given the way in which Aristotle links this notion with that of incidental causation, three possible answers suggest themselves:

(1) *E* has at least one incidental cause.
(2) *E* is incidentally caused, but has no *per se* cause.
(3) *E* is incidentally caused by at least one, and is *per se* caused by none, of some favoured set of events.

In section IV I shall defend the view that (3) is Aristotle's answer.

The final question to be asked in connection with the first definition of chance is: what has happened to the idea that chance events happen neither always nor for the most part? Aristotle began the chapter by outlining two conditions: being rare and being 'for the sake of something'. His obvious intention is to define chance events in terms of the events which satisfy both of these conditions. In the account of chance and luck which follows, and which culminates in the first definition, the second of these conditions is straightforwardly carried over, as we have seen. The rarity condition, on the other hand, disappears, and is replaced by 'coming to be incidentally'.[19] Yet that he intends rarity to be part of his analysis of chance is clear both from his strategy at the start of chapter 5,[20] and by the fact that the rarity condition reappears later in the chapter (197^a3–5, 18–20, 30–2).

Clearly Aristotle intends the condition that the chance event 'comes to be incidentally' to be the *same* as the condition that it happens rarely. How can this be? Coming to be incidentally is a matter of being incidentally caused in some way, and hence of being incidentally connected in some way; so these incidental connections must be *rare* connections. That what is incidental is rare is, of course, confirmed by *Met. Δ*. 30 (quoted below, pp. 86–7) and *E*. 2.[21] The most natural way

[19] At 197^a5–6 Aristotle gives a definition of τύχη in which the rarity condition is likewise replaced by the condition that luck is an incidental cause.

[20] Note also that his remarks there must imply that rarity is a necessary condition of being a chance event (196^b10–17, quoted at the beginning of section II above). For rarity as a necessary condition for chance, see also the references listed in n. 7.

[21] I do not of course include in this claim τὰ καθ' αὑτὰ συμβεβηκότα (see *An. Post.* I. 4–6; *Top.* V. 3–5, VII. 5; *Met. Δ*. 18 and 30).

to cash out the idea of rarity in the case of incidental causation is to take Aristotle to hold that tokens of a given type of incidental 'effect' only occur rarely within the range of occurrences of tokens of the incidental cause. I shall return to this idea in the next section.

III

I shall leave the analysis of chance to one side for the moment and ask in this section and the next what Aristotle means by saying that an event happens always, or for the most part, or rarely. These expressions are judgements of frequency—judgements as to how often a given event happens or a state of affairs holds.[22] This point may seem too obvious to mention, but I do so for two reasons. First, it is a familiar fact that Aristotle very closely associates judgements of these three types with others whose reduction to judgements of frequency is highly problematic if not downright impossible: 'always' is associated with '(by nature and) of necessity', 'for the most part' with 'by nature (but falling short of necessity)', and 'rarely' with 'incidentally' (κατὰ συμβεβηκός).[23] How to give a good account of these various associations has long been a vexed question; but clearly the right place to start is by asking precisely what sort of judgements of frequency figure in them.

Second, while commentators have tried to give accounts of one or other of these problematic associations, and have discussed the frequency-judgement relevant to the particular association in ques-

[22] 'Frequency' is not meant in any technical sense, of course; nor do I mean to suggest that Aristotle developed any sort of probability theory.

[23] 'Always' and 'of necessity': *Int.* 9, 19^{a}9–18; *An. Pr.* I. 13, 32^{b}4–10; *Top.* II. 6, 112^{b}1–20; *Phys.* II. 5, 196^{b}10–20; *Cael.* I. 12; *GC* II. 9, 335^{a}33–^{b}3 and II. 11; *PA* I. 1, 639^{b}23–4; *Met. E.* 2, 1026^{b}27–30. 'For the most part' and 'by nature': *An. Pr.* I. 13, 32^{b}4–13; *Cael.* I. 12, 283^{a}2–3 and III. 2, 301^{a}7–9; *Phys.* II. 8–9, especially 198^{b}34–6; *PA* III. 2, 663^{b}28–9; *GA* IV. 8, 777^{a}17–21; *EE* VIII. 2, 1247^{a}31–3; *Rhet.* I. 10, 1369^{a}32–^{b}2. 'Rarely' and 'incidentally': *Met. Δ*. 30, 1025^{a}14–19 (quoted on pp. 86–7), *E.* 2, 1026^{b}27–1027^{a}28, as well as a number of the passages just cited. The difficulty of giving a good account of these associations might lead one to think that Aristotle uses the expression 'always or for the most part' simply to *mean* 'by nature', and 'rarely' to mean 'not by nature' (Henry Mandell tried to convince me of this idea at the conference mentioned in n. 60 below). But although it is beyond doubt that it is for Aristotle an a priori truth of some sort, when he says that what happens always or for the most part happens by nature, he clearly does not intend it as a banal analytic truth or as a stipulative definition. *Mutatis mutandis*, the same points hold for the association of 'rarely' with 'incidentally'.

tion,[24] none, to my knowledge, has considered all three together. Yet it is clear that the three notions of 'always', 'for the most part', and 'rarely' must be examined and understood together, not piecemeal; and if this is so, the problematic associations should not be examined piecemeal either.

There are two types of frequency-judgement to which Aristotle's use of 'always', etc. might refer; we may call these judgements of 'absolute' and 'conditional' frequency. If the judgements are of absolute frequency, then to say that an event happens always, for the most part, or rarely means that events of this type[25] happen *all the time*, or *most of the time*, or neither. The judgement of frequency is made without reference to the occurrences of other events (except by way of comparison: 'events of this type happen less often than ones of that type'). It is this sort of frequency which is involved in saying that visits by the Queen to Christ Church are rare occurrences—she visits the college perhaps once in ten years—but diligent academic study is going on in Christ Church most of the time (there are more hours when someone or other in the college is studying than there are hours when no one is), and breathing is going on there all the time. Conditional frequency, by contrast, concerns *ordered pairs* of events: thus E_1 is usual/rare relative to E_2 iff E_2 is usually/rarely accompanied by E_1 *when* E_2 *occurs*. We make judgements of conditional frequency if we say that the Queen usually comes to Christ Church *when she is invited to open a new building*, that academic study rarely takes place in the college on Christmas Day, or that an eclipse always takes place when the sun and moon are situated thus-and-so relative to the earth. Whereas on the absolute frequency view 'always', etc. can be represented as quantifiers ranging over periods or instants of time, on the conditional frequency view they range over certain *cases*—that is, over cases in which E_2 occurs. The terminology of 'always/for the most part' is in itself neutral

[24] The association of 'always' and 'of necessity' is discussed in Jaakko Hintikka, *Time and Necessity: Studies in Aristotle's Theory of Modality* (Oxford, 1973), and Sarah Waterlow, *Passage and Possibility: A Study of Aristotle's Modal Concepts* (Oxford, 1982). 'For the most part' and 'by nature' are discussed in Gisela Striker, 'Notwendigkeit mit Lücken: Aristoteles über die Kontingenz der Naturvorgänge', *Neu Hefte für Philosophie* 24–25 (1985), 146–64, and the nature of 'for the most part' judgements in Mario Mignucci, '῾Ως ἐπὶ τὸ πολύ et Nécessaire dans la Conception Aristotélicienne de la Science', in E. Berti (ed.), *Aristotle on Science: The Posterior Analytics* (Padua, 1981). I discuss Mignucci's paper in n. 26 and on p. 87.

[25] Obviously this categorization of events must be primarily a categorization of event-types, since individual event-tokens happen only once; event-tokens can none the less be classified in a derivative way, as being tokens of a type which happens always/etc.

between these two interpretations, because even 'temporal' words such as ἀεί can be understood (as can 'often'—and indeed 'sometimes', 'more times', etc.—in English) either temporally, or non-temporally in terms of numbers of cases.[26]

It is not hard to see which of these Aristotle must have in mind. If we classify events in terms of absolute frequency, the results are quite at variance with those of Aristotle's classification. Many of his standard examples of events which happen 'for the most part'—such as hot weather in summer, or confectionery-makers producing sweets—do not happen most of the time in this sense: there are fewer times when confectioners are making sweets or when there is hot summer weather than when no sweet-making is taking place or when it is not hot summer weather. Again, there are many celestial events, such as the conjunctions of planets, the completion of a Great Year, and so forth, which happen only very rarely in terms of absolute frequency; but Aristotle would classify them with the perpetual motion of the celestial spheres (and regard them as happening with the same necessity), rather than with royal visits to Christ Church.

Moreover, a number of Aristotelian doctrines involving the threefold classification have little prospect of making sense on the absolute-frequency interpretation. (1) Aristotle holds that there is no λόγος or ἐπιστήμη of rare events;[27] yet how could the timing of solar eclipses, or of the appearance of Halley's comet, be studied and calculated by astronomers if that were the case?[28] (2) As I have said, he associates

[26] As Mignucci points out in ''Ὡς ἐπὶ τὸ πολύ', when Aristotle does add some qualifying phrase to his talk of always or for the most part, it is sometimes a 'temporal' phrase, and sometimes a 'cases' phrase. Mignucci himself views the temporal usage as primary, and the 'cases' usage as merely reflecting the flexibility of ordinary language (pp. 185–7). But if anything the reverse seems to be true. He cites three passages in which 'temporal' words appear in connection with ὡς ἐπὶ τὸ πολύ (ποτε at *HA* vi. 4, 562^{b}3–9, ὀλιγάκις at *GA* i. 19, 727^{b}26–30, and ἀεί at *EN* ix. 2, 1164^{b}30–4); in all three it is obvious that Aristotle is concerned with the frequency of cases, and not with the frequency of times. On the other hand, the passages which Mignucci sees as referring to cases (*Cat.* 7, 7^{b}24–7; *HA* v. 20, 553^{a}4–6 and vi. 29, 578^{b}15–16; *PA* iii. 2, 663^{b}28–9; *Rhet.* ii. 15, 1390^{b}23–4) cannot easily be construed as referring to times.

[27] *Phys.* ii. 5, 197^{a}19–20; *Met. E.* 2, 1027^{a}19–23; cf. *An. Post.* i. 30, 87^{b}19–27.

[28] As Richard Sorabji has pointed out, there are some things which could come under the heading 'the timing of eclipses' but which are incidental events (*Necessity, Cause, and Blame: Perspectives on Aristotle's Theory* (London, 1980), 9–11 and n. 16); Aristotelian examples include an eclipse occurring just when the Athenians are about to withdraw, and, perhaps, an eclipse occurring *tomorrow* (although what Aristotle means at *Met. K.* 8, 1065^{a}16 is unclear). In these cases the incidental event is the concomitance of the eclipse with some other event (if the eclipse occurring tomorrow is incidental, 'tomorrow' must be construed as an indexical: what is incidental is the connection between the timing of

'for the most part' with 'by nature', and opposes 'rare' and 'by nature';[29] but a proposition's being true at more times than it is false is simply irrelevant to the issue of whether the event which it describes occurs in accordance with nature. If the number of goats in the world were sufficiently large, for instance, 'Some goats are hermaphrodites' could be true at most or all times (cf. *GA* IV. 4, 770^{a}35–6 and ^{b}33–6), yet this condition could still be, in Aristotelian terms, an unnatural and anomalous one for goats. Likewise, there are fewer times when a birth of a normal goat is taking place than when one is not; this hardly makes such births contrary to nature. (3) The same point holds for the doctrine that rarity is part of the analysis of chance. Thus, for instance, at 197^{a}2–5 Aristotle says:

> It is in these circumstances that he is said to go [to the market place] from luck, but if he went having chosen and for the sake of this, or frequenting the place either always or for the most part when collecting subscriptions,[30] he would not be said to go from luck.

Aristotle can hardly be saying, 'this type of case is due to luck—unless the man spends more than half his whole life collecting subscriptions in the market'.

(196^{b}34–197^{a}1 might seem to support the absolute frequency interpretation: '[The lucky creditor] went not for the sake of this [getting the money], but it came about for him that he went and did what was for the sake of collecting [in the non-explanatory sense: i.e. 'did what produced the beneficial result that he collected his money']—and he did this not frequenting the place either for the most part or of necessity.' It might seem as if Aristotle is saying that the creditor's getting the money would not be due to luck if his going to the agora were not itself unusual. But this would be an absurd claim for Aristotle to make: how usual it is for the creditor to go to the agora has nothing to do with whether his meeting the debtor on this occasion is a coincidence or a piece of good planning. Ross rightly suggests that we are meant to understand τούτου ἕνεκα with φοιτῶν (this is made easier by the fact that Aristotle several times glosses ἐλθεῖν with '(for the sake of) getting the money'): if the creditor usually goes to the

the eclipse and what our position in the time series happens to be). Such incidental conjunctions are not the subjects of astronomical study, whereas the interval between two particular eclipses, for instance, is.

[29] For references, see n. 23 above.

[30] There are no good reasons for deleting κομιζόμενος here, as Ross does; but my point does not depend on its retention (see below).

agora *to get his money*, we would not have a case of luck at all.[31] Now 'going there to get his money always or for the most part' still admits of interpretation in terms of conditional frequencies; but what is relevant to whether the case is one of luck is whether the man regularly goes to the agora *when collecting*—a point about conditional frequencies. The precise way in which conditional regularities are relevant to luck is discussed further in section V below.)

If Aristotle has in mind judgements of conditional frequency, all these difficulties disappear. First, the right events turn out to happen rarely, for the most part, and always; second, the various doctrines involving the threefold classification which were so problematic on the absolute-frequency interpretation promise to make sense.[32] Thus the association of 'always or for the most part' and 'by nature', and the claim that there is no ἐπιστήμη of what happens rarely, have some plausibility if the frequencies in question are conditional: the workings of nature are causal workings (operative in all four Aristotelian causes), and hence the obvious regularities to see as 'natural' are regular *connections* between things. As we have seen, the same is true for Aristotle's account of chance as rare.[33]

Further evidence supports the conditional-frequency interpretation. (1) As we saw in section II, Aristotle treats the condition that chance events are rare as equivalent to the condition that they come to be incidentally. Since coming to be incidentally is a matter of being incidentally caused in some way or other, it would be natural to construe the rarity of chance events as rarity relative to the occurrences of the relevant incidental cause(s). I shall return to this point in the next section. (2) In two important passages Aristotle applies the threefold classification in a way which is very suggestive of conditional frequency and which would be hard to square with absolute frequency. In *Met. Δ*. 30, he connects rarity with the incidental:

We call *incidental* that which attaches to something and can be truly asserted, but neither of necessity nor for the most part; for instance, if someone in digging a hole for a plant found treasure. This—the finding of treasure—is incidental for the man who digs the hole; for neither does the one come of

[31] Ross, *Aristotle's Physics*, 520; Ross presumably thinks that we should take 197^{a}2–5 in the same way.

[32] Though of course much more needs to be said to explain what Aristotle means by these claims and why he thinks that they are true. I return to these issues in sections IV–V below.

[33] On the doctrine that there is no ἐπιστήμη of rare events, see pp. 91–2 below.

necessity from the other or after the other, nor, if a man plants, does he for the most part find treasure. (1025^a14–19)

This is echoed in the discussion of teleology in *Phys.* II. 8:

For these things, and everything which is by nature, happen in a given way either always or for the most part, whereas none of the products of luck or chance do. For we do not take the frequent rains in winter to be due to luck or to coincidence, but frequent rains in summer we do; nor heat in the summer but only if it happens in winter. (198^b34–199^a3; cf. *Met.* E. 2, 1026^b30–6)[34]

I shall end this section by mentioning two dissenting views. In '*Ὡς ἐπὶ τὸ πολύ*', Mignucci defends a view of 'for the most part' which is essentially the same as the absolute-frequency interpretation, though he does not formulate the distinction between absolute and conditional frequency. He rejects the idea that 'for the most part' is a quantifier ranging over cases, and argues that it ranges over instants of time. His main argument[35] is that Aristotle twice associates being for the most part with being *εἰκός* (*An. Pr.* II. 27, 70^a2–6, *Rhet.* I. 2, 1357^a34–$^b1$), and that this association can only be understood if 'for the most part' ranges over times. But there is nothing in this association which favours Mignucci's times-interpretation over the cases-interpretation; in fact, it favours cases over times for reasons analogous to those given above. Mignucci thinks that, if 'for the most part' refers to cases rather than to times, then for-the-most-part propositions would be capable of being always true, and hence, via the association of 'always' and 'of necessity', of being necessary and certain rather than merely probable. But this line of argument in effect begs the question, since it assumes that the relevant sense of 'always' is 'at all *times*', rather than 'in all *cases*'.[36]

John Cooper argues that 'always' and 'for the most part' carry a much stronger sense for Aristotle than the one which I wish to ascribe

[34] A further passage, *An. Post.* II. 12, 96^a8–11, might appear to support my interpretation, since Aristotle there associates the 'always/for the most part' classification with propositions of the form 'All/most *A*s are *B*s': 'some things are not always, but for the most part; e.g. not every male human has hair on his chin, but for the most part they do.' But since Aristotle's interest here is in *syllogistic propositions*, the use of propositions of the form 'All/most *A*s are *B*s' is only to be expected, and does not tell us whether 'always' and 'for the most part' propositions are essentially of this form.

[35] See also n. 26 above.

[36] Mignucci also appeals to *Rhet.* II. 25, 1402^b34–1403^a2 in support of his view of the link with *εἰκός* ('*Ὡς ἐπὶ τὸ πολύ*', 197 and n. 51); but, as he agrees, 'le passage est difficile', and it can be interpreted so as to support either view.

to him.[37] Cooper seems to agree that the terms involve what I have called conditional frequency; but he insists that they involve the idea of *eternal instantiation* as well:

> When Aristotle speaks . . . of some arrangement as being so 'always or for the most part' . . . he means to say that that arrangement is found existing eternally or recurring regularly throughout all time, with only the occasional exceptions implied in the 'for the most part' rider. . . . [N]or would he count animal generation as happening 'always or mostly' in some particular way if animals were found in the universe only in a certain finite period of its existence, even though *when* they exist they are always or mostly generated in that way. (p. 203, n. 4)

The basis for Cooper's view is his interpretation of two arguments against 'materialism' in *Phys.* II. 8–9 which, he thinks, have as their starting points premisses concerning what happens always or for the most part. He claims that these arguments make best sense if we construe their respective initial premisses as saying that the formation of animal parts in ways which benefit the animal, and that the ordered growth of complex organisms from a given kind of seed, are *eternal* features of the world. It is this eternality which Aristotle thinks cannot be explained by his materialist opponents.

But on this interpretation Aristotle would be very blatantly begging the question: precisely because of the features of their theories which are in question, none of his opponents will agree that these processes are eternally occurring ones.[38] If Aristotle's premisses that these things happen 'always or for the most part' have this loaded sense, then he merely assumes at the outset that the theories of Empedocles and the atomists are inadequate, and presents no argument which they are obliged to consider. So we should take the initial premisses about happening always or for the most part to be making the weaker claims, neutral between the materialists' theories and Aristotle's, that observable regularities of these sorts exist (though not necessarily eternally).[39]

Furthermore, although Aristotle does believe that most if not all natural phenomena which 'happen always or for the most part' *are*

[37] J. M. Cooper, 'Aristotle on Natural Teleology', in Malcolm Schofield and Martha Nussbaum (eds.), *Language and Logos* (Cambridge, 1982).

[38] This is true not just for Empedocles, the opponent Aristotle mentions by name, but also for his other obvious target, the atomists.

[39] I wish to thank David Charles for drawing my attention to a similar criticism of Cooper by Charlotte Witt: see n. 17 of Charles's paper in this volume. Cooper's interpretation deserves a fuller treatment than I can give it here; I hope to discuss it elsewhere.

eternally instantiated, there is good evidence that he does not intend the phrase to mean that. When he uses it in II. 5 in connection with the creditor going to the agora,[40] Aristotle is clearly not considering the supposition that the man goes to the market throughout all time![41] Again, at *Met. E.* 2, 1027^{a}22–6 Aristotle speaks of honey water curing fever for the most part: there is no suggestion that honey water's effect on fever need be regularly instantiated throughout the whole of time, or even that honey water itself regularly exists throughout eternity. Aristotle simply points to a natural regularity without regard to the question of its eternal instantiation.

IV

There is good reason, then, to take Aristotle's threefold classification as based on judgements of conditional frequency—that is, on judgements of the frequency of tokens of E_1 given the occurrence of tokens of E_2. The analysis of 'always', 'for the most part', and 'rarely' is not yet over, however, as we can see by considering an objection to this interpretation.

Let us call the event-type which occupies the 'E_2' position in the conditional schema the 'event-condition' for E_1. There is a striking difference between some 'always/for the most part/rarely' judgements and others. In some judgements of this sort, the event-condition is explicitly stated (rain *in winter*, finding treasure *when planting*), or it is obvious from the immediate context (honey water cures fever—sc. when it is given to fever victims). But in many cases, Aristotle's judgements contain no overt or contextual reference to any event-condition: are these judgements of conditional frequency?—and if so, what is the intended event-condition? These questions are especially urgent because when Aristotle speaks of his threefold classification of the totality of events—those which happen always, those which happen for the most part, and those which happen rarely—no event-condition is specified.

[40] 196^{b}36–197^{a}5; see pp. 85–6 above.

[41] It might be thought that Aristotle's usage is simply different when it comes to regularities in human affairs; but the passage is intended to illustrate the condition in Aristotle's general analysis of chance events that chance events happen neither always nor for the most part—the analysis to which the argument in II. 8 adverts—and there is no hint that this condition is radically different when it applies to πράξεις and when it applies to natural phenomena.

To answer these questions we must return to the analysis of chance. As I have said, the condition in that analysis that *E* is rare turns out to be equivalent to the condition that *E* 'comes to be incidentally', and three possible meanings for this phrase suggest themselves: *E* is incidentally caused by something or other; *E* has an incidental cause but lacks any *per se* efficient cause; or *E* is incidentally caused by at least one, but is *per se* caused by none, of some favoured set of events.

Aristotle's meaning cannot be the first of these. The condition that *E* comes to be incidentally is meant to distinguish chance events from other beneficial events; but every beneficial event will be rare relative to *something* and will be incidentally caused by something. Coming to be incidentally must be a stronger notion, or every beneficial event would be a chance event. The same conclusion can be drawn directly from the fact that when Aristotle applies the threefold classification to the totality of πράγματα, the resulting classes are clearly meant to be exclusive: but they would not be exclusive if being rare were simply a matter of happening infrequently relative to something or other, since events which happen always or for the most part also satisfy that condition. Likewise one and the same event could happen 'always' relative to one thing, and 'for the most part' relative to another.

The second of the three possible senses of 'coming to be incidentally' is indeed stronger in the requisite way. But it would be very surprising if Aristotle's intention here were that what distinguishes chance events from other beneficial events is that the former lack any *per se* efficient causes. The examples of incidental causation which Aristotle gives when explaining the notion of 'coming to be incidentally' ($196^{b}26$–7 and $197^{a}14$–15) are all instances of my schema I (p. 79 above); and instances of schema I actually *require* a *per se* cause.[42] If Aristotle intended 'comes to be incidentally' to mean 'has an incidental cause but no *per se* one', he could hardly have introduced the notion in this way.

Now it is true that being an incidentally caused event of the type represented in the other schema (schema II) is compatible with having no *per se* cause; and it is also true that many chance events which fit schema I also contain somewhere a causal structure which fits schema II, and arise because of some coincidence—perhaps even all do so. Hence it would not be impossible for someone to hold that events

[42] This point is made by Charlton (*Physics I and II*, 108). He believes that it is true that chance events have no *per se* causes, and regards it as an error on Aristotle's part that he fails to recognize this truth.

which fit schema I owe their status as chance events to their connection with some event(s) which lacks *per se* causes. But again there is no hint of any of this in Aristotle's account in *Phys.* II.[43] Moreover, being incidentally caused in the schema II mode is, of course, also compatible with *having* a *per se* cause; and in the cases we are considering—coincidences—there is an obvious candidate for the *per se* cause, namely the conjunction of the *per se* cause of each of the event's co-inciding 'elements'. Thus the *per se* cause of the coincidental encounter in the agora is the conjunction of the *per se* cause of the creditor's going to the agora and that of the debtor's going there. This can be represented by the following schema:

$$\begin{array}{ccc} A & \Rightarrow & B \\ & & + \\ (A \text{ and } C) & \Rightarrow & (B \text{ and } D) \\ & & + \\ C & \Rightarrow & D^{44} \end{array}$$

While decisions on the part of the creditor to see a spectacle are in this case only rarely (and hence incidentally) connected with meeting his debtor, instances of this conjunction of causes would regularly result in their meeting—and would do so in a perfectly understandable and predictable way. Someone might object that the existence of conjunctive *per se* causes of coincidental events is incompatible with Aristotle's claim that there is no ἐπιστήμη of the incidental. But in fact Aristotle can consistently admit the existence of such causes while denying that they are objects of knowledge. Although they will produce their effects in a law-like fashion, they cannot be the objects of theoretical, scientific knowledge, since scientific understanding as Aristotle conceives it is restricted to the natures of species taken individually, and so excludes the conjunctive operations of more than one subject. And as for practical knowledge, the indefiniteness of incidental connections explains why there are no τέχναι of incidental causes (*Met. E.* 2,

[43] Nor can I see any evidence elsewhere in Aristotle's writings that he held the view that chance events lack *per se* causes. The best attempt to argue the case (for coincidences, at least) is Sorabji's interpretation of *Met. E.* 3 (*Necessity, Cause, and Blame*, ch. 1); but I am not convinced by it (for some criticisms, see Gail Fine, 'Aristotle on Determinism', *Philosophical Review* 90 (1981), at 563–4 and 576).

[44] '(*B* and *D*)' denotes the co-presence in the agora of creditor and debtor, while '(*A* and *C*)' denotes the conjunctive event which, on this view, is the *per se* cause of their co-presence. When, as in this example, *B* and *D* have different subjects, the relation denoted by '+' will have to be weaker than 'accidental unity'; but it will still be a relation of concomitance.

1027^{a}19–23). Getting from incidental cause to incidental effect will involve bringing together two chains of causation; but there may be any number of ways of doing this, and no systematic way of ordering these as more and less efficient, for which means are the most efficient may depend entirely on the circumstances of the particular case. Keeping a man in a dark cellar will make him pale: but how does one get the man into the dark cellar? Violence may be appropriate in one case, the mention of Amontillado in another.[45]

How else could we interpret coming to be incidentally so as to distinguish chance events from other beneficial ones? I suggest that Aristotle has in mind the third of the possible senses of 'coming to be incidentally', and that his doctrine is this: a chance event is incidental *relative to all the natural and deliberative processes taking place in the subject*;[46] in other words, *E* comes to be incidentally if neither the φύσις nor the προαίρεσις of its subject is the *per se* cause of *E*. What marks out chance outcomes is that they are beneficial events which are not reliably connected with the natural workings or the choices of the subject who benefits from them.

It is easy to see how an account of chance along these lines could be motivated. Our aim of understanding the world about us—of making sense of the operations of nature and the strategies of rational agents—requires us to distinguish pieces of behaviour which are reliably connected with those operations and strategies from those which, even if they appear to be, are not. And Aristotle is certainly committed to something like this line of thought. For he thinks that chance events lack *per se* final causes: 'whenever things come to be not for the sake of what results . . . we say that they come to be from chance' (197^{b}19–20, to be discussed below).[47] It is central to his account of chance that chance outcomes involve the appearance of final causation (since they

[45] Note that one cannot maintain that conjunctive *per se* causes *as such* are impossible; for without them we could not give a sensible account of *planned* conjunctions—as when I arrange things so that the creditor and the debtor unwittingly arrive at the agora at the same time.

[46] By this I mean relative to exercises of these natural and deliberative processes by any (relevantly similar) subject, and not merely relative to *this* subject's exercise of them. In *Phys.* II. 5, it is true, Aristotle illustrates the rarity condition in terms of the same subject's activities (197^{a}2–5); but this illustrative procedure is really compatible with either view, and if these frequency judgements are to be straightforwardly useful in explanatory theories—as they are clearly meant to be—they should be relativized to the activities of any relevantly similar subject.

[47] See n. 8 above for references to Aristotle's opposition of 'by chance' and 'for the sake of something'.

are beneficial outcomes) while having no actual *per se* final cause;[48] and the distinction between merely apparent and actual final causation is fundamental to his natural science. Why is it that chance involves merely the appearance of final causation? Precisely because chance outcomes are only rarely connected with the natural capacities and choices being exercised in the subject, and hence the ends towards which those processes are directed cannot explain the occurrence of these outcomes. So Aristotle is committed to the view that a chance event is rare not merely relative to something or other, but relative to all the then-active workings of 'mind and nature' in the subject.

Of course, to say that Aristotle is committed to a certain view is not to say that he realizes this, still less that it lies behind a particular text. But there is in fact good reason to suppose that Aristotle is aware of the line of thought which I have sketched, and that he utilizes it in his treatment of chance; this is to be found in his *second* definition of chance.

This definition occurs at 197^b18–20, part of which I have already quoted:

> Consequently, it is clear that among the things which come to be for the sake of something without qualification, whenever things come to be not for the sake of what results, things whose cause is external, we say that they come to be from chance.

Thus E is the outcome of chance iff (1) E is among the things which come to be for the sake of something 'without qualification', (2) E does not come to be for the sake of what results, and (3) the cause of E is external.

'Without qualification' (ἁπλῶς) in clause (1) seems to mean 'in the most general sense',[49] and so indicates the broad, 'non-explanatory' sense of ἕνεκά του which characterizes chance outcomes. If this is right, clause (1) is the same as the first clause of the first definition. The first definition's second clause (the condition that E comes to be incidentally) has been replaced in the second definition by clauses (2) and (3). As with the replacement in the earlier definition of the rarity

[48] I ignore here the question of bad luck, which does not materially affect the present point. (On bad luck, see n. 12 above.)

[49] So Ross, *Aristotle's Physics*, 522. Lennox takes Ross to be saying that the phrase 'things which come to be for the sake of something without qualification' means 'things which generally (i.e. normally) come to be for the sake of something' (Lennox, 'Teleology, Chance', 233 and n. 30). But this is a doubtful reading of Ross, and a virtually impossible reading of ἁπλῶς.

condition by coming to be incidentally, the natural inference is that these new clauses are meant to be equivalent to the condition that *E* come to be incidentally—and hence equivalent to the original condition that *E* happen rarely.[50] Yet clause (2) by itself appears to be equivalent to 'coming to be incidentally' construed as I have suggested: for there will be a teleological explanation of the occurrence of a beneficial outcome *E* if and only if *E* has a *per se* cause in one or more of the natural and deliberative processes going on in the subject.[51] What about clause (3), the 'external cause condition'? Unless Aristotle is introducing out of the blue a wholly new condition on chance events (see n. 50), the best interpretation is to take it as spelling out the clause (2) condition in a different way; that is, to take 'things whose cause is external' to mean 'things whose cause is incidental to the operations of the subject's φύσις and νοῦς'.[52] Aristotle uses two descriptions of the same condition because they focus on different aspects of it. The first is concerned with (*per se*) final causation in chance events—or rather with the lack of it—while the second focuses on the structure of *per se* efficient causation which explains the absence of any final cause: the cause of the chance outcome is *external* precisely in the sense of being external to the natural and rational processes occurring in its subject.

If this is right, Aristotle is prepared to treat as equivalent the condition that *E* is rare and the condition that *E* is only incidentally related to its subject's natural and deliberative processes, and not susceptible of final-cause explanation—in terms of the τέλη of those processes. We have good reason, then, to think that Aristotle defines a

[50] It would be all the more surprising if clauses (2) and (3) were meant as *new* conditions, given that the definition of chance in general was the business of ch. 5, and ch. 6 is largely concerned with distinguishing between luck and chance$_N$ (chance in the narrow sense: see n. 3 above). Nor can we suppose that these conditions only apply to chance$_N$; for 197^{b}20–2, and in particular τούτων at ^{b}21, makes it clear that the definition at ^{b}18–20 applies to chance in the wide sense—the sense which includes luck. Aristotle says here that lucky events are those which satisfy the definition at ^{b}18–20 and which also satisfy two further conditions: so ^{b}20–2 describes luck as a subset of the 'definiendum' of ^{b}18–20. Conditions (2) and (3), then, apply both to chance$_N$ and to luck.

[51] Note that *E* must have at least an incidental connection with the natural and/or deliberative processes of its subject *S*, or *S* would not be its subject.

[52] This interpretation of ὧν ἔξω τὸ αἴτιον is suggested by Charlton (*Physics I and II*, 110). Note the comparable expression at 197^{a}1–2: 'the τέλος—collecting the money—is not one of the causes *in him*'. Aristotle uses 'external cause' again in a puzzling passage later in the chapter, 197^{b}32–7. I take the relevant sentence (^{b}35–7) to mean 'but (between chance$_N$ events and events *which happen by nature*) there is this difference: of the former the cause is external, while of the latter it is internal'. Although this squares with my interpretation of ἔξω at 197^{b}20, the later passage is too obscure to rest a case on it.

chance outcome as a beneficial event which is rare relative to the natural and deliberative processes taking place in its subject.[53]

In his analysis of chance Aristotle never adverts to—let alone discusses—this relativization of rarity to the subject's natural and deliberative processes. On the contrary: he introduces in a very formal way the two classifications 'beneficial/non-beneficial' and 'always/for the most part/rarely', and appears to think that it is enough to define the class of chance outcomes as the intersection of the classes of beneficial and rare events. This suggests that when no event-condition is given, Aristotle's 'always/for the most part/rarely' classification is meant to *incorporate* this relativization to the nature and choices of the relevant subject(s): to say that E happens neither always nor for the most part just is to say that it occurs only rarely on the occasions when the same natural capacities are being exercised or the same choices have been made—and likewise, *mutatis mutandis*, with 'always' and 'for the most part'.[54] Indeed, it is only by taking the threefold classification this way that we can understand it as involving conditional rather than absolute frequency, and as generating the classes of events which Aristotle takes it to generate.

V

In this section I wish to distinguish two possible views which Aristotle might hold of the relationship of 'rarely' to 'incidentally', and of 'always or for the most part' to 'by nature'.[55] This distinction is pertinent here because these associations are involved in the analysis of chance. The obvious question is, which of these views does Aristotle actually hold?; but answering this question involves more issues and more controversial texts than could be dealt with here, and is a task for another paper.

[53] This will be explored further in section V below.

[54] The case of judgements that something is or happens *incidentally* presents an analogous feature. These too are essentially relational judgements. As such they can be made relative to any condition, providing that this is specified or made obvious by the context: cold in the dog-days happens incidentally, a stumble may be incidentally worse than pleurisy (*EN* v. 11. 1138^{b}2–5). But these judgements can also have the particular function of signifying what is incidental relative to the φύσις of the subject (as they do when Aristotle employs the distinction between essential and accidental predication). In the nature of the case, however—since 'incidental' here usually modifies a predicate—the analogue of the 'event-condition' is specified via the designation of the subject of the predication.

[55] For simplicity of exposition I shall for the most part ignore *mind* here, and speak solely in terms of nature.

(1) Aristotle might take *E*'s happening rarely as by itself both necessary and sufficient for its happening incidentally to its subject, and likewise *E*'s happening always or for the most part as necessary and sufficient for its happening in accordance with the nature of that subject. On the interpretation outlined in sections III–IV, this would amount to the view that events which normally or always accompany the operations of the nature of a thing are due to its nature, and events which only rarely do so are not.

These straightforward 'equations'[56] leave no metaphysical room for regularities which are incidental to their subjects, or for rare happenings which are none the less in accordance with φύσις. We can distinguish two types of 'incidental regularity'. (i) Suppose a given species of animal, *S*, never or only rarely interbreeds with another species—that is, members of *S* refrain always or for the most part from interbreeding in this way. It seems that we ought to be able to distinguish at least in principle between cases in which this failure to interbreed is due (in part) to the φύσις of *S* in some way, and cases in which interbreeding is perfectly possible physically, and the failure is due to something incidental to *S*'s nature, such as geographical distribution.[57] (ii) An incidental regularity can also be *completely accidental*, that is, although each instance of the event has an explanation, there is no overarching explanation of why it occurs regularly. If Aristotle takes *E*'s regularity to be sufficient for *E* to belong *per se* to its subjects, he cannot distinguish either of these types from natural regularities; nor can he distinguish the corresponding types of 'non-incidental rarity' from the genuinely incidental.

This is relevant to Aristotle's analysis of chance because if he does hold this 'simple view' (as we might call it) of the connection between regularity/rarity and nature/the incidental, his analysis faces counterexamples involving all these types of case. Thus suppose the creditor knows that his debtor is always in the agora at a particular time, and regularly intends to go the agora on subscription-day to meet him; but suppose also that the creditor has a fairly bad memory, and so usually forgets to go. When he does go, his meeting the debtor, though rare

[56] That is, equations of the extensions of the relevant terms, and not equations of their meaning: see n. 23 above.

[57] Note the explanation which Aristotle reports of why interbreeding is more common in Libya than elsewhere, at *GA* II. 7, 746^b9–11: 'the scarcity of water means that they all meet together at the few places where there are springs, and in consequence mating takes place even between animals of different species'.

relative to the creditor's intentions to meet, is not a piece of luck (the creditor's memory is not so bad that remembering the trip to the agora is a fluke). This counter-example involves a non-incidental rarity of type (i). Again, the creditor might have a run of very good luck, and, without any overarching explanation, just happen to bump into his debtor on the right day more often than not—an incidental regularity of type (ii).

(2) Alternatively, Aristotle might regard the equation of 'always or for the most part' with 'by nature' as reflecting a fundamental truth about the connection between regularity and nature, but as being itself only very nearly true: in exceptional cases the inference from regularity to happening by nature is defeasible by considerations of the nature of the relevant subject (and perhaps by other parts of our scientific theory). On this view an event's happening always or for the most part constitutes highly privileged but defeasible evidence for the existence of a *per se* connection with its subject's nature: it is sufficient for ascribing such a connection *providing that* there are no countervailing considerations concerning the nature of the subject. This position would leave conditional frequency relative to the φύσις of the subject at the centre of the picture, while recognizing that the *relatum*—φύσις—has a role to play in the relevant explanatory theory in its own right.[58] The idea of defeasibility by considerations of the nature of the subject would make it possible to draw the distinction between the two cases of regular failure to interbreed described above; and extending the defeasibility condition to include defeasibility by other parts of the theory (i.e. parts which concerned other natures) would make it possible to acknowledge incidental regularities of type (ii) as well.[59]

The considerations adduced above show that what I called the simple view—view (1)—is false, and that an analysis of chance which relies on

[58] It is also to these considerations of the subject's φύσις that we must appeal when the actual occurrences of the 'event-condition' are too few to make a scientifically respectable judgement of an event's conditional frequency. Of course, the roles of frequency judgements and of judgements about the subject's φύσις are not as easily isolated from each other as I have pretended for the sake of exposition. This is especially true of a developing science; but Aristotle's interest in the question of how we arrive at a conception of something's φύσις on the basis of observed regularities is not itself well-developed. See *An. Post.* II. 19; Robert Bolton's paper in this volume discusses many of the questions which that chapter poses.

[59] And thus to acknowledge consistent runs of good or bad luck, once νοῦς was also taken into account. Aristotle discusses whether such runs of luck are possible in *EE* VIII. 2.

it is likewise flawed. But the simple view, and the consequent account of chance, are not absurd views, and it is not hard to imagine a philosopher holding them. Moreover, as I argued in section III, even the simple view makes much better sense than the doctrines to which Aristotle is committed on the absolute-frequency interpretation. So I think that even if it turns out that, given my interpretation of the meaning of the 'always/for the most part/rarely' classification, Aristotle has to be seen as holding the simple view, this would not be a serious objection to the interpretation: it would just be that Aristotle failed to see the need for (or consciously rejected) the more sophisticated view.

I shall explore view (2) further elsewhere, and (to let the cat out of the bag) attempt to defend the claim that Aristotle embraces it (or something like it) and rejects the simple view. But the interpretation of his analysis of chance and of the meaning of the 'always/for the most part/rarely' classification which I have defended in this paper does not depend on this claim, for the reasons which I have just given.

VI

It is a commonplace that individual substances occupy a central position in Aristotle's natural science and his philosophy alike. The interpretation advanced here of Aristotle's threefold classification of the totality of events reveals yet another way in which this is true: this classification hinges on the natures and choices of individual substances, and its application to the empirical world is entirely dependent upon what natural science discovers about the workings of those substances. This suggests that the centrality of individual substances will itself be central to understanding Aristotelian positions which invoke the threefold classification. The conditional-frequency account of 'happening always', for instance, suggests that the doctrine that what happens always does so of necessity is rooted not in his conceptions of time and modality, but in his conceptions of ἐπιστήμη and φύσις. Likewise, the 'equation' (whichever way we take it) of *E*'s happening rarely and its happening incidentally relative to all the natural processes of its subject should lead us to expect that the centrality of individual substances is itself central to other issues that involve chance and related notions; and I think that this is in fact the case. The defence of teleology against Empedocles in II. 8, the idea of *regular* spontaneous generation in the biological works, and the argument in *Met. E.* 3, I would argue, all turn on the assumption of the centrality of substances. But these too are stories for another day.[60]

[60] An earlier version of this paper was read at a conference on the *Physics* held at the University of Southern California in December, 1989; I am grateful to the participants at that conference for their comments. I am also grateful to John Ackrill, Jonathan Barnes, David Charles, Malcolm Schofield, and Jenny Tyler for their helpful suggestions at various points in the genesis of the paper.

5

Teleological Causation in the *Physics*

DAVID CHARLES

I

Aristotle's discussion of the fourth cause is introduced in *Phys.* II. 3 as follows.

> Again a mode of cause is the goal [τὸ τέλος]. This is that for the sake of which [τὸ οὗ ἕνεκα]: as health is that for the sake of which there is walking. ('Why is he walking?' We say 'in order to be healthy', and saying this we think we have given the cause.) The same applies to all things that come to be on the way to the goal [τὸ τέλος] through the action of someone else: reduction of flesh, purging, drugs or surgical instruments in the case of health. For all these are for the sake of the goal, even though they differ as some are activities, others instruments. (*Phys.* 194^b32–195^a3)

Later in the same chapter Aristotle wrote,

> the others are causes of the goal [τὸ τέλος], that is the good [ἀγαθόν] of the other things. That for the sake of which [τὸ οὗ ἕνεκα] means that which is the best, that is the goal of the things that lead up to it. And it makes no difference whether we say the good or the apparent good. (195^a23–6)

The fourth cause—the teleological cause—plays a distinctive role in Aristotle's discussion of physics and of metaphysics. He held that in these areas, as in biology and psychology, one must employ teleological causation in addition to material and efficient causation. Many of his predecessors erred, in his view, in thinking that such phenomena could be explained in terms of material causation alone. Thus he (famously) wrote, 'It is not likely that fire or earth or any such element should be the cause of things manifesting goodness and beauty both in their being and their coming to be' (*Met.* 984^b11–14). In most recent discussions of Aristotle's account of teleology, the major interest has been to discover why he thought that teleological causation was indispensable in studying natural processes, and why he thought that his predecessors had erred in attempting to avoid using it. Some have argued that this was because he believed that it would never be possible to characterize conditions sufficient for the existence of an organism or its parts if one employed only the resources of material causation and efficient causation (involving the matter alone).[1] Others have suggested that Aristotle believed that because the essences of certain psychological or biological phenomena are determined by their role in a teleological system they could not be reductively defined, even if there was a complete account in terms of material and efficient causation of the conditions under which they existed. On this view, Aristotle held that teleological causation was a type of causation which could not be understood on the basis of material and efficient causation alone.[2]

These discussions are not without interest. But there is a prior question which easily can be overlooked. Is Aristotle's basic account of what teleological causation is defensible? This is a major issue in assessing central areas of his philosophy. In this essay, my aim is to address only some aspects of this question:

(1) to seek to understand how Aristotle conceived of teleological causation, and
(2) to consider some of the resources he had to justify his employment of teleological causation.

Further, since my present focus is on *Phys.* II, I will, in the main, ignore considerations which, although relevant to these questions, are drawn from his other works. This will allow for further discussion elsewhere of whether (and to what extent) his conception of teleological causation remained unchanged throughout his writings.

If we are to achieve goal (1) we must consider at least three sub-issues.

(1*a*) In the examples cited, there is some diversity among teleological causes. Health (a state) and being healthy (a non-intentional activity) are both mentioned in 194^{b}32 ff. Elsewhere in *Phys.* II, there is an even wider variety of teleological causes:

organisms: men (200^{b}3);
artefacts: houses, axes (200^{a}10, 24);

[1] For a clear statement of this view see Allan Gotthelf, 'Aristotle's Conception of Final Causality', *Review of Metaphysics* 30 (1976–7), 226–54.

[2] See D. Charles, 'Aristotle on Hypothetical Necessity and Irreducibility', *Pacific Philosophy Quarterly* 69 (1988), 1–53; F. Lewis, 'Teleology and Material/Efficient Causes in Aristotle', *Pacific Philosophical Quarterly* 69 (1988), 54–98.

intentional actions: cutting with an axe (200^b5), obtaining money (196^b33);
natures of things: (198^b17, 194^a28).

So what, if anything, unifies this class? Do they form one ontological category? Or is there some further way in which they are unified?[3]

(1*b*) In the examples cited, there is also some variety among what is caused teleologically. Some effects are actions (walking, purging), others are instruments (drugs, surgical instruments). Elsewhere in *Phys.* II, there is an even greater range of (possible) teleological effects:

processes: raining (198^b20 ff.);
parts of organisms: teeth (198^b25);
things that come into being: (199^a8);
things being as they are: (199^a8).

How is this class to be unified? Do they form one ontological category? If not, what makes them all teleological *effects*? Is there a consistent pattern present in this apparent plethora of teleological causes and effects? What, if anything, constrains what can be included in the relevant lists?

(1*c*) What is the relation between teleological cause and effect when the former is the cause of the latter? When Aristotle writes 'walking is for the sake of health', is the relation 'for the sake of' primitive? Or did he think that it could be further explained? If so, how?

Consideration of these issues raises in a sharp form the question of how far Aristotle succeeded in justifying his employment of teleological causation. In particular, it gives rise to the following sub-issues.

(2*a*) What grounds does Aristotle offer for thinking that teleological causation is a genuine type of causation which is distinct from (e.g.) efficient causation? And this involves consideration of what, for Aristotle, makes something a genuine type of causation.

[3] There is considerable variety also in Aristotle's examples of efficient causes. Some are objects (e.g. the doctor (195^a21, 30), the builder (195^b5, 21, 23), the seed (195^a21), Polyclitus and the sculptor (195^b10)), some are objects in activity (e.g. the builder building (195^b6), this man building (195^b19), this man doctoring (195^b19)), while others appear to be states or attitudes of agents (e.g. the art of building (195^b23), wishing to see something (197^a16)). However, some of these will only be potential, not actual, causes. The latter appear to be *substances in activity* (195^b16 ff.), while the former are the substances themselves which are potentially active in the way required to bring about the relevant change. Further, the connection between the builder actively building and his exercising his skill as a builder is a close one. By contrast, the examples of teleological causation are too diverse to be ordered in this way.

(2*b*) To what extent does Aristotle succeed in establishing that teleological causation is a genuine type of causation? Or, more minimally, does he succeed in establishing that it is as defensible a mode of causation as efficient causation?

These five sub-issues are distinct from a further question which I will attempt to exclude (as far as possible) from the present discussion.

(3) Does Aristotle give good grounds for taking teleological causation to be *indispensable* in his account of biological or psychological phenomena?

Favourable answers to (2*a*) and (2*b*) would not, by themselves, lead to an affirmative answer to (3). For Aristotle might have taken teleological causation to be a genuine type of causation, but have thought that was not essential for an adequate explanation of biological or psychological phenomena.

II Teleological Causes and Effects

The *relata* in *Phys.* II. 8 fall into two broad categories. The first concern agency, the second natural organisms. They may be classified (somewhat roughly) as follows:

(1) *Cases of agency*
- (i) Actions done for the sake of something, which may be either a state *or* a further action;
- (ii) Objects (or features of objects) which exist for the sake of actions done for the sake of something, which may be either a state or a further action.

(2) *Cases involving natural organisms* (and not involving purposive agency)
- (iii) Processes of formation of organisms which occur for the sake of the whole organism;
- (iv) Parts of organisms (or features of those parts) which are present for the sake of the whole organism.

In the *Physics*, the first category includes cases both of human (e.g. craftsmen: 199^a12 ff.) and of animal agency (e.g. spiders: 199^a22). The second (as I define it) includes cases such as the formation of teeth (198^b25 ff.) or the location of the roots (199^a27 ff.) which do not involve imagination, *nous*, or anything (psychological) of this type. Aristotle says little about the cases of sub-human animal action in the *Physics*. In what follows, therefore, I will compare his account of human agency with that of natural phenomena not governed by imagination, or anything (psychological) of this type. Aristotle's discussion of spiders, swallows and ants raises important issues, but they lie outside the scope of the present paper.

In the *Physics*, Aristotle does not seem concerned to organize these cases further. Thus, he treats (i) and (ii) as on a par (195^a1–3) as both occur 'for the sake of a goal'. He does not even attempt to argue in the cases involving *agency* that (i) is more basic than (ii) in the order of teleological causation. While he notes that both (iii) and (iv) obtain in cases of organisms (199^a6–8, 30), he makes no attempt to argue that either (iii) or (iv) is the more basic (in the *Physics*). It is true that in *Phys.* II. 8 he focuses most insistently on (i) (the case of the craftsman), and (iii) (the case of natural genesis). But this seems to be only because he is pursuing the analogy between artistic and natural production, rather than because he is taking these to be the central cases of teleological causation in terms of which (ii) and (iv) are to be explained. When the point of the analogy is completed, he alludes to cases where the explanandum is an organism in existence rather than one coming into existence. But there seems to be no attempt here to take (iii) as the basic case of teleological causation in terms of which (iv) is to be understood.

Outside the *Physics*, Aristotle does on occasion claim that in the order of exposition (iv) should be stated before (iii), and notes that because man has certain properties, the process of his formation must be such and such and take place in a given manner (*PA* 640^b1–3). But Aristotle is, it seems, doing nothing more here than saying that you cannot understand the goal or nature of generation if you do not understand what the goal or pre-existing agent is—a man with his parts. So we need to understand what man is before we understand the processes of generation in which other men are produced. But this in no way commits Aristotle to thinking that he can or should define

A's *coming to be* for the sake of a goal (e.g. man)

in terms of

A's *being* for the sake of a goal.

Rather they seem to be treated as separate cases of teleological causation.

It appears that Aristotle accepts a plurality of cases of teleological causation without attempting to explain them in terms of one more basic case. If he had tried to do this, he could have unified his account of teleological causation by showing how all other cases could be explained on the basis of cases of (e.g.) the processes of formation of the organism which occur for the sake of the whole organism. This would have been a major task. It would have required him to show that parts of animals exist for the sake of an end solely because they are the focus of capacities which play a role in producing whole organisms. Also, it would have been incomplete unless he could also show that human actions are done for the sake of a goal only when they are steps towards the formation or maintenance of the relevant organism (or thought to be such). But there are no indications in the *Physics* or elsewhere that Aristotle actually engaged in attempting to carry through this type of project. Nor is there clear evidence that he was convinced that it could be carried through, or even that he thought it a project worth pursuing.[4]

This raises the question of whether there is one unified conception of teleological causation in Aristotle's account. If he did not attempt to take one of these four cases as basic, and explain the others as derived from this, he could have established the unity of teleological causation only by showing (as an alternative) that all these cases exemplified a common form. Indeed, this might appear to be the strategy he wishes to follow in the *Physics* when he treats agency and natural-organism cases as analogous with one another, with neither as more basic. Thus, he argues both 'as in art, so in nature' and 'as in nature, so in art'. (*Phys.* 199^a9–10, 15–16) However, superficially, cases (1) and (2) seem

[4] For a contrasting (but unsupported) claim see Gotthelf, 'Aristotle's Conception', n. 19. Gotthelf reiterates this claim in the 1986 Postscript to his earlier article in A. Gotthelf and J. Lennox (eds.), *Philosophical Issues in Aristotle's Biology* (Cambridge, 1987), 239 ff., and provides one consideration, based on one introductory text in *PA* I, in favour of it. It should be noted that even if this argument were successful, it would not establish Gotthelf's general claim, which concerns *all* cases of teleological causation, as the relevant text does not discuss examples of human agency. However, the consideration he adduces is at best inconclusive. In *PA* 640^a22 ff. Aristotle notes that Empedocles failed to recognize that in cases of animal generation the producer was a pre-existing agent ('as man produces man': 640^a25–6). He then continues (after a break) in 640^a34: 'Hence we should say that because man is such and such, therefore a man has these things'. But this sentence appears to emphasize *not* that man has certain parts because they come to be in a given way (as Gotthelf claims), but rather that generation occurs as it does *because* man has certain parts (cf. 640^b1–3). And this appears to favour, if anything, *some* form of priority of case (iv) over case (iii) (even if only in order of exposition). Gotthelf noted in 'Aristotle's Conception' that he needed to establish that Aristotle accepted the general explanatory priority of case (iii) over cases (i), (ii), and (iv). This claim, which is central to Gotthelf's important thesis, has yet to be supported by detailed analysis of either *Phys.* II. 8–9 or the *Parts of Animals* (with its wealth of specific examples). A full investigation of the latter cases would be a major, and worthwhile, undertaking.

different. In (1), the goal appears to be the object of an agent's desire or choice. And this encourages the thought that an action is done for the sake of a goal *G* if and only if the agent does that action with the further desire to produce *G*. Similarly, an instrument may exist for the sake of goal *G* if and only if that instrument is made by an agent with the desire to achieve *G*. And in these cases the agent is sensitive to the value of the goal. It is because the relevant state seems valuable that the agent pursues it (*Met.* 1072^a27 ff.). Were it not to seem valuable, she would not do so. But in (2), desire or choice are lacking. Aristotle rightly shows no sign of attributing desires or choices to all natural organisms. So in (2) the goal is not itself the object of an agent's desire or propositional attitude. And it appears that the organism need be not sensitive to the value of the end state produced. Were it to cease to be valuable, the organism would not immediately cease to pursue it.

In cases of agency, the agent is also sensitive to what is the best *means* to the goal. In a variety of cases, were the best means to achieve a goal to be different, the route the agent would follow would be correspondingly different. And this shows that the agent acted as she did *because*, for instance, the action she chose was the best means to the goal. It was not that the action chosen merely happened to be the best means to the goal. The agent's internal states (e.g. of perception or belief) were sensitive to the fact that this was the best route to her goal and she chose it for that reason.

The situation seems different in the cases of natural organisms, considered in (2). Parts of organisms (such as hearts) exist for the sake of the whole organism in that they serve a function in (e.g.) the maintenance of the life of that organism. Thus we say (e.g.) that the heart produces certain consequences which contribute to the survival or flourishing of that organism. This thought is sometimes strengthened by the further claim that organisms of this type have survived and reproduced in the past *because* they possessed hearts of this type. On both these views, a part has a function if and only if it tends (in one or these ways) to ensure the survival or other characteristic goals of the organism.[5] But the goal is obviously not the propositional object of the

[5] Similarly, a process would have a function if and only if it tends to produce a given state which is beneficial to the organism's survival or flourishing. And we might add to this the additional historical claim that in the past processes of this type played this role, and this contributed to the survival of the species. For claims of this general form, see L. Wright 'Functions', *Philosophical Review* 82 (1973), 161, and Ruth Millikan, *Language, Thought and other Biological Categories* (London, 1984), 28 ff.

relevant part's 'desire', nor is it selected by the part *because* it seemed or was good. Rather it is determined by an external characterization of what flourishing or survival for that organism consists in. The part in operation contributes to that end, but need not itself be sensitive either to the fact that the state produced contributes to the organism's flourishing, or to the goodness of the state produced. Further, it is not claimed that were the best means to that state to be different, the part would function in a correspondingly different way. Thus, if the best way for the heart to function to contribute to the organism's survival were to change (and it was required that it pump blood in the opposite direction), the heart's activities would not alter accordingly. The heart itself is not sensitive to the fact that the route it follows is the best way to achieve its goal. Even if it does take what is, in fact, the best route to its goal, it does not do so *because* it is the best one to achieve that goal.[6] At most, the heart has a tendency to produce a state which is a good for the organism to which it belongs by a route which reliably produces that goal. But in this case neither the route itself nor the goal is produced by an internal state of the organism which is sensitive to the fact that this is good or the best route to the relevant goal.

There are, it appears, major differences between the types of teleological causation at work in the cases of agency and natural organisms (in this account). In the case of agency, both the goal and means are selected because they are good for the agent and the agent is sensitive to this fact. In the examples of natural organisms, parts may indeed function in ways that are good for the organism. But neither the goals nor the means need be selected just because the organism is sensitive to their goodness. The goodness of the goals or route is not a necessary feature of the explanation of what occurs. The part (or organism as a whole) may have a tendency to promote a state, which is in fact good. But it need not itself be sensitive to the goodness of the end state.[7]

[6] A similar point could be made if the route was specified as the 'one required for the goal'.

[7] This is not, of course, to say that one can determine what things are goals without reference to the good of the organism. It is to say only that the parts themselves are not *sensitive* to the goodness (or goal-status) of the end-state produced. In 'The Place of the Good in Aristotle's Natural Teleology', in John J. Cleary and D. C. Shartin (eds.), *Proceedings of the Boston Area Colloquium in Ancient Philosophy* iv: 1988 (Lanham, Md. 1989), Gotthelf argues for the more radical thesis that Aristotle thought that one could determine what an organism's end is (e.g. its full actuality) without reference to what is good for it. This claim is difficult to sustain in *Phys.* II as Aristotle here introduces the notion of an end in a way which is at least partially explained by what is good (195^{a}23–6 ff., 194^{a}28–

This point can be brought out by the following example. In certain cases, a species (e.g. the mosquito) may develop over time new methods for overcoming its environmental challenges (e.g. DDT) and, eventually, find a route which allows it to escape this danger. This is, on one account, because DDT-resistant mosquitoes have higher successful reproduction rates than those that are not resistant, and eventually come to predominate in the mosquito population. This differs from the agency case in that it takes a period of time for the adjustment to occur. But more radically, the adjustment occurs because (e.g.) the non-DDT-resistant mosquitoes die out, and the ones which survive and reproduce are DDT-resistant. There is nothing in the mosquito (species) itself which is sensitive to the fact that the latter route is better for its survival chances. Rather the environment blindly destroys those mosquitoes which are not DDT-resistant, without the species itself selecting *as good* a means to overcome the challenge of the environment.[8] And this contrasts sharply with cases of purposeful agency. While in both cases one may use the locution '*a* occurs for the sake of goal *G*', it involves something completely different in examples of functional and purposive explanation.[9]

The difference in ontology between Aristotle's varying teleological causes and effects is startling enough. But the apparent differences within the genus of teleological relation between cases of agency and of natural organisms' function are even more threatening. And this naturally prompts the question: is there one unified account of teleo-

33). Elsewhere, Aristotle characterizes full actuality (partially) in terms of certain intrinsic goals (*Met.* 1048^{b}17 ff., 1050^{a}21–3) rather than explicating these goals in terms of the relevant actuality. But what are these goals if not what is good for the relevant organism? (I intend to argue for this claim in more detail elsewhere. It is, in my view, essential for a proper understanding of Aristotle's strategy in *Met.* Θ.)

[8] This is not to suggest that Aristotle himself countenanced or developed this type of causal story. The issue is whether his account leaves a gap which could be filled by a story of this type.

[9] There is a further possible difference between the cases of agency and natural function which should be noted. Animals' natural flourishing has some specific content in Aristotle's account which remains relatively unchanged under different conditions. Human agents, by contrast, deliberate towards the apparent good, and their actions can be caused teleologically even when they do not, in fact, aim at their real good. (cf. *Phys.* 195^{a}25–6). Since their real and apparent good can differ, their desires are sensitive to the goodness of their (proposed) goals in a way not paralleled in the case of non-deliberating animals. For it is this, and not (e.g.) their tendency to promote their full actuality, which is common to all their relevant desires. (Lindsay Judson alerted me to this.) However, this is a separate point from the one on which I am concentrating; for mine concerns the flexibility of means pursued for ends taken to be good rather than the variability of ends of a human agent.

logical causation at work in Aristotle's discussion of these topics? Did he believe that these two kinds of teleological causation were the same? Alternatively, did he think that the second, natural organism, model was the basis for understanding all cases of teleological causation? Or did he take the first, agency model, as his starting point, and represent all cases of function as exemplifying this stronger form of teleology? Or perhaps the differences between them did not matter for his concerns? It is only when we have an answer to these questions that we can begin to assess Aristotle's account of teleology, and grasp the distinctive kind of causation it involves.

Recognition of the unreduced plurality of types of teleological causation at work in the *Physics* serves to render untenable one otherwise attractive view of what its operation consists in. If one focuses exclusively on discussion of case (iii)—the process of formation of natural organism—and adopts the 'function-model' interpretation, it is natural to construe Aristotle's teleological causes as examples of efficient causation with a distinctive type of cause (e.g. the potentiality for or *nisus* towards the fully-formed organism) or a distinctive type of effect (fully functioning organism, or its constitutive activities). On this view, the basis for teleological causation would be a *nisus*, or potentiality, to bring into being a fully formed organism. And the connection between the *nisus* and its effect need only be that the former is the efficient cause of the latter.[10] One may doubt whether this is the correct model to understand cases of type (iii).[11] But even if it were, it

[10] This model for understanding teleological causation is influential. Gotthelf, 'Aristotle's Conception', adopts the view that what is distinctive of teleological causation is the presence of a certain type of efficient cause (e.g. an irreducible potential for form). Martha Nussbaum has regarded teleological causation in certain cases as marked out by its distinctive effects (*Aristotle's De Motu Animalium* (Princeton, 1978), 84).

[11] This view would account for cases of type (i) if one took (e.g.) desires to exemplify a *nisus* of this type, and gave up the idea that (e.g.) the *desire for* ϕ was itself essentially teleologically grounded in ϕ *being* or *seeming good*. But it would appear to require a major change in Aristotle's view of desire (cf. *Met.* 1072^{a}29–31). I argued for the alternative, good-based, interpretation of Aristotle's account of desire in *Aristotle's Philosophy of Action* (London 1984), 85–8, 229–33, 237–40. If it is correct, Aristotelian desires are essentially defined in terms of the goodness (real or apparent) of the state at which they aim and the good-based rationalizing connections with other psychological states in which they stand (e.g. in practical reasoning). Indeed, these two teleological features (sensitivity to the goodness of the goal and the means) serve to define desire in a way which (I claimed) is inconsistent with the standard functionalist account of the goal-directedness of desire (in terms merely of the state it tends to produce). For in the latter, the goodness of the end-state and the good-based rationalizing connections play no essential explanatory role. If my account of Aristotelian desire is correct, it seems better to argue by analogy that case (iii) is mishandled on any view which takes the existence of the relevant *nisus* as a basic datum (e.g. a potential for Form), and fails to explain its presence or nature in

could not account for cases (iv) or (ii). For here there is no possibility of the existence of parts being the *efficient cause* of the existence of the whole. There is no process of which the part is the starting-point or controller which culminates in the existence of the whole. If so, the general form of Aristotle's teleological causation cannot be given by construing it as an *efficient* cause with a distinctive type of antecedent or consequent.

III *Physics* II. 8: How Many Models at Work Here?

In discussing teleological causation in this chapter (and the next), Aristotle is addressing a specific type of opponent. This opponent, in effect, makes three separate claims.

(1) Everything that occurs of necessity, occurs as a causal consequence of the nature of simple bodies and their movements.[12]

(2) The necessity involved in the nature of simple bodies and their movements is simple, non-hypothetical, necessity.[13]

(3) Whatever occurs not as the result of the nature and movement of simple bodies, occurs by chance. Amongst the latter category are the development and continued propagation of a given species of animal.[14]

Given that Aristotle's opponent holds both (1) and (2), he will hold that everything that occurs necessarily is an example of simple necessity. If anything occurs which is not a case of simple necessity, it occurs 'by chance'—(3). So his world is divided into two exclusive and exhaustive sets of phenomena: those that occur by simple necessity,

terms of its being (or seeming to be) for the good of the organism. (For a contrasting view of these issues, see Susan Suavé's Review of D. Charles, *Aristotle's Philosophy of Action*, *Philosophical Review* 97 (1988), 413. Suavé regards as 'not foreign to Aristotle' a minimal view of teleology according to which a process is teleological if and only if the essence is to produce a given result, which need not itself be specified as good. In the account offered here, the goodness of the good is explanatorily essential in Aristotle's account of action, desire and function. See p. 122 below.)

[12] *Phys.* 198^{b}12–16. The opponent does not take seriously the possibility that there are other causes beyond the hot, the cold, and things of this sort. This latter claim need mean no more than this: whenever there is a causal relation whose relata are not (e.g.) the hot and the cold, but (e.g.) a man, this causal relation depends on one which involves (e.g.) the hot and the cold. The opponent is not saying that *everything* that exists occurs of necessity. He is making the weaker claim that everything that occurs of *necessity*, is caused in a given way. This is consistent with holding that some things occur by chance.

[13] *Phys.* 199^{b}35–200^{a}4.

[14] *Phys.* 198^{b}27–32.

and involve simple bodies and their movements, and those that occur by chance. *Tertium non datur.*

The opponent's three claims are importantly distinct. He could have held (1) and (2), but denied that whatever occurs not as a result of simple necessity happens 'by chance'. Thus, he could have allowed that apart from occurrences which result from simple necessity there are ones which are caused (e.g.) probabilistically but are not necessitated at all. Or he could have held that everything occurs as a result of simple necessity. To reject his third claim is not to reject either (1) or (2).

In arguing against this opponent, Aristotle begins by attacking the weakest point in his position: claim (3). Against this, Aristotle argues that chance is not sufficient to account for the regularity of animal generation, because chance only operates in unusual or exceptional cases (198^b35–199^a3). On this interpretation, Aristotle's first argument is deployed successfully against a specific opponent who accepts that natural animal generation occurs by chance, and also accepts Aristotle's view of chance as confined to exceptional or unusual cases.[15] So understood, Aristotle is not moving in 198^b35–199^a3 from the premiss that natural animal generation is not planned to the radical conclusion that (if so) it must be a series of exceptional co-incidences. Rather his opponent will himself have accepted this conclusion when he stated that everything that does not occur by simple necessity occurs by chance (provided that he also agreed with Aristotle's own account of chance). Since the opponent himself has suggested chance as the only alternative to purposive explanation (198^b21–3) in cases of natural generation, Aristotle can conclude as follows.

> If therefore these things are taken to be either by chance or for the sake of something, and they cannot be either by chance or automatically, they would be for the sake of something. All things of this kind occur by nature—as the opponents themselves would agree. Thus that for the sake of which is present in things that come to be and are by nature. (199^a3–7)

[15] As deployed in *Phys.* II. 5–6. The most cautious move for this opponent would be to accept Aristotle's account of chance, and then to modify his original claim that chance and teleological goals are the only two possible candidates for the explanation of the development of natural organisms. The opponent clearly would be foolish to maintain his original claim while at the same time accepting Aristotle's account of chance. The interesting question is why Aristotle envisaged his opponent to be so 'flat-footed' at this point as to refuse to modify his formulation in the light of the earlier discussion of chance.

For, so understood, this argument is directed *ad hominem* against someone who had offered chance and teleological goals as the only possible causes of, for example, natural animal generation. Given that chance is not a successful candidate for this role, teleological goals must be present (on the opponent's assumptions). If so, the scope of 'thus' in the final line is: 'given these assumptions, it follows that'. Aristotle need not himself be asserting that 'by chance or for the sake of something' are the only alternatives. All that he is doing is saying that, given an assumption which is acceptable to the other participants in the debate, it follows that teleological goals are present in natural changes. On this reading, Aristotle does not attempt to establish by this argument (198^b32–199^a8) that teleology was present in natural processes, because this conclusion rests on nothing more than his opponent's concession (coupled perhaps with his own inclination to believe it as true) of its basic premiss. Rather the argument is designed to show that chance cannot play the role postulated in his opponent's theory (so (3) must be rejected).[16] It further suggests that if chance is the one candidate offered in place of teleology in accounting for natural processes, then one can establish dialectically that teleological goals are present in natural processes.[17]

In the remainder of II. 8 and II. 9, Aristotle focuses on his opponent's claims (1) and (2). He does not challenge (2) directly. However, he

[16] This argument may be a case of Aristotle's dialectical argumentation in the *Physics*. It rests on what seems to be the case (199^a3), and on the opponent's acceptance of certain other claims (199^a5–6). If so, it may be legitimate for Aristotle to employ the opponent's premiss that either these occurrences result from chance or for the sake of something in his discussion, without himself attempting to establish it as true. Other parts of the chapter, however, appear not to be dialectical. This seems especially true of Aristotle's basic argument for the presence of teleology in 199^b15–18: that if there were not teleological direction towards a goal, there would be no natural processes or natures.

[17] John Cooper presented an alternative account of this argument in 'Aristotle on Natural Teleology', in Malcolm Schofield and Martha Nussbaum (eds.), *Language and Logos* (Cambridge, 1982), 207 ff. His suggestion rests in part on the claim that no set of coincidental occurrences can be eternal. However, it is unlikely that Empedocles would accept that species are eternal (cf. 198^b31–2). If so, Aristotle cannot refute his view merely by assuming that species are eternal. This point is well made by Charlotte Witt, *Substance and Essence in Aristotle* (Ithaca and London, 1989), 92–3, and by Lindsay Judson, pp. 87–8 of this volume. In Cooper's view, the argument also rests on the claim that unplanned results are coincidences which do not happen in more than 50% of a very large number of cases. But there is no reason to think that *all* materialists will accept that unplanned results are coincidences in this sense. So it seems preferable to take this argument as directed *ad hominem* against Empedocles who believed that unplanned results did happen 'by chance', and so were (given Aristotle's view of chance) coincidences in the required sense.

does dispute (1) by arguing that there is a mode of causation present in cases of agency and of natural change which is not the result of the movements of simple bodies, and that this mode of causation involves necessity. Further, he claims that this mode of necessity is not simple necessity, and hence denies that everything that occurs of necessity occurs of simple necessity (in II. 9, see below). It is the first stage in this argument that gives us the clearest insight into the nature of teleological causation. For this attempts to display both what teleological causation is and how it involves necessity.

In 199a8–b13, Aristotle deploys his analogy between artistic creation and natural processes. These sections have attracted considerable criticism on two counts. Some have objected that Aristotle construes natural processes on the inappropriate model of purposive agency. Others have thought that he argues for the presence of teleology in nature on the basis of a highly contentious analogy with human agency. However, the general pattern of Aristotle's considerations is far more cautious than either of these criticisms allows. He seems concerned to point to two specific features of teleological causation which are present in cases of nature and of agency. Nature is subject to teleological causation because it, like purposive agency, exhibits these features. This does not commit Aristotle to thinking of nature as *like* an agent in all respects (e.g. as possessing intentional states). Nor does it depend on his arguing for teleology in nature because nature is like a purposive agent. Rather, he is suggesting that two basic features of teleological causation are equally present in the two cases. To understand his conception of teleology requires us to grasp what these features are, and to see why they are required in Aristotle's scheme.

The first feature runs like this (199a8–33). In any case where there is a goal, the prior and subsequent stages in the relevant sequence occur for the sake of this goal. Thus, if G is the goal of a sequence $a_1 \ldots a_3$, a_1, a_2 and a_3 all occur for the sake of G (199a8–9). Aristotle's conclusion is that this structure is to be found in cases of skill and nature alike (199a18–20). In any such case, a_1, a_2 and a_3 all individually occur for the sake of G. It is not enough that a_1 occurs for the sake of G, and that this produces a_2 and a_3 which result in G: a_2 and a_3 have also to occur for the sake of G. There has to be an organized process each relevant part of which occurs for the sake of the goal, and is itself caused teleologically. Further, each part's occurring where it does in the process (e.g. as first, second, or third member) has itself to be explained teleologically (199b7–9). So, the occurrence and order of each part of what occurs has to be explained as for the sake of a goal, if a *goal* is present.[18]

Aristotle suggests that this feature is present in cases both of human agency and of natural process (199ª9–10). He argues that if results of agency were to be produced naturally, they would occur in precisely the same way. And conversely that if results of nature were produced by an agent, they too would occur in the same way. Art both imitates the teleology of nature and seeks to perfect it (199ª15–16). While he sometimes argues 'as in agency, so in nature', he also argues 'as in nature, so in agency' (199ª9–10). He does not give any special explanatory priority to the case of human agency, nor argue that nature itself is an 'artificer'. The force of the comparison is limited to the specific point being made, which (it is argued) applies equally to both cases. Indeed, Aristotle seems concerned to emphasize the fact that he is not considering nature to be like an agent. For he focuses first on cases of animals where no deliberation or search is present (199ª20–3), and then on examples of plants where intentional states are absent (in Aristotle's view: *An.* 414ᵇ1 ff.).

The considerations which Aristotle adduces are non-trivial. They succeed in imposing a general condition on teleological causation in all these cases: each relevant part of the process must itself be *for the sake of the goal*, and it must occur where it does in the process for the sake of this goal. If a process or set of actions is to be caused teleologically, there must be this type of organization in the *explanandum*. And they further suggest a generalization to the case of organisms. If they are subject to teleological causation, each part of the organism must itself be for the sake of the goal, and be situated where it is for the sake of the goal. Indeed, Aristotle appears to move in this direction when he speaks of things 'that come to be and are by nature' (199ª29–30). In these cases, the 'shape' of the organism provides the goal *for the sake of which* all the relevant parts are present where they are. In both natural processes and natural organisms, the relevant 'parts' are individually for the sake of a goal, and are 'positioned' where they are for the sake of that goal. Indeed this degree and kind of organization is precisely what makes them natural.[19]

[18] There could, of course, be one-stage processes where only a_1 occurs for *G*.

[19] This strategy is at work in Aristotle's discussion of processes in *Phys.* III. 1–3, esp. 202ª22–4. I have discussed this in *Aristotle's Philosophy of Action*, 23–5. What makes the type of process the one it is is the goal it is directed to achieve. This is, of course, consistent with tokens of that type being incomplete or inadvertent. (For a contrasting view,

However, while these considerations are non-trivial, and serve to specify the range of teleological effects, they do not spell out the basis for Aristotle's comparison between purposive agent and natural process or natural organism. Need the latter be sensitive to the goodness of the means or the goal adopted? Or is it sufficient that the organism functions in a way which is indeed beneficial without itself selecting this way because it is good for the organism? Nothing in this passage suggests grounds for preferring one of these interpretations. And this is because the notion of '*for the sake of*' seems to be taken as basic throughout and not further explained.

Aristotle's second characterization of teleological causation rests on the notion of *a mistake* (199^a33–$^b7$). A mistake occurs in the case of agency when a stage in an action-process is not completed successfully in the way required to reach the goal at which the agent aims (199^b1–3). A mistake occurs in the case of a natural process when a stage is not completed successfully—in the way required to reach the relevant goal. Such mistakes, which result in 'monsters' in the case of generation, occur because the internal source of change is destroyed in a way that accounts for its failure to reach its goal. This, presumably, prevents the internal source producing the goal towards which it is directed. In both cases, a stage occurs successfully for the sake of a goal if it leads to that goal, and in a mistaken way if it does not. The notion of mistake depends on the thought that, for example, a_1 occurs for the sake of G, but G itself does not in fact result because of the way in which a_1 occurs.

This further feature does not suggest that Aristotle thinks of nature as 'like' an intentional 'artificer'. It requires only that there be an internal source of change in an organism whose function is to produce G, and that this be in some way parallel to the desire or plan of an intentional agent. Thus, this case in no way deepens our understanding of which (if either) of the two models of teleology Aristotle is employing in this context. It appears that he is merely noting two general features of the *effects* of teleological causation: they are organized

see Sarah Waterlow, *Nature, Change and Agency in Aristotle's Physics* (Oxford, 1982), 127 ff.). Aristotle's basic claim in these chapters is that one cannot individuate natural processes (as types) without referring to the goals they are directed to achieve. Elsewhere, he argues that one cannot individuate natural organisms without reference to their goals (e.g. *Met.* 1044^b1–3, 1050^a21–3 etc.). This provides (as I intend to argue elsewhere) the basis for his pivotal analogy between processes and organisms (cf. *Met.* 1048^b8–11). See also n. 7 above.

structures, 'parts' of which can operate successfully or unsuccessfully. This may indeed be true whenever the notion of 'for the sake of' applies; but it says nothing about its proper basis. Indeed, it appears that Aristotle is taking this notion as a primitive or basic one in this passage. But if so, he fails to indicate which form of teleological causation is at work here. The most he does is to delimit the range of teleological effects.

The same lacuna is found in the final, and most basic, argument in II. 8 for the necessity of teleological causation. This begins with a critical remark directed against Empedocles, but proceeds to make a far more general point. It runs as follows.

> Moreover among the seeds things come to be by chance. But anyone who asserts this destroys what is 'by nature' and 'nature' itself. For whatever arrives at some goal by a continuous movement from an internal starting point happens 'by nature'. The result is not the same for each, but yet it is not a matter of chance; but what occurs always tends towards the same goal, if nothing intervenes. (199^{b}14–18)

The salient point is this: if one removes talk of goals, one thereby removes talk of natures and natural processes. For they are both essentially defined in terms of goals. Natural processes are essentially composed of stages which are for the sake of the goal, which will be achieved if nothing intervenes. Natures are essentially composed of parts which are for the sake of a goal, which will be achieved if nothing intervenes. This is what natures and natural processes are: they both essentially involve the goals of the stages or parts from which they are composed. This does not, of course, require that each process always results in the same goal, or that it always proceeds in the same way towards that goal. All that it requires is that it *always* reaches the relevant goal in favourable circumstances. In the case of natures and natural processes, achieving the relevant goal is not an isolated, one-off, occurrence. If there are to be natures or natural processes, they must always achieve the same goal in favourable conditions. At this point, one reaches bedrock in Aristotle's defence of teleological causation: it must be a genuine form of causation, because if it were not, the world would contain no natures and no natural processes. Teleological causation is required if the world is to be as Aristotle envisages it to be.

While this provides the basis for Aristotle's attachment to teleological causation, it does not serve to illuminate its *nature*. These claims would be compatible with Aristotle holding that natural organisms must be sensitive (like purposive agents) to the goodness of their goals and of the means to be achieved to reach them. But they would also be compatible with his adopting the 'functional model' of teleological causation, in which parts or processes have the function of producing consequences that contribute to the goal of the organism by a route which regularly achieves this (without any sensitivity to value being required). Nothing in this passage enables us to decide between these alternatives.

Similar remarks apply also to Aristotle's famous conclusion to this chapter: 'The best illustration (or for the sake of something) is the doctor doctoring himself. Nature is like that' (199^b30–2). For this does not determine whether nature is like the self-doctoring doctor in finding desirable means to a desirable goal (the agency model), or merely in acting for the sake of a goal. The latter would be neutral as between the Purposive Agency and Function Models of Teleology.

To summarize: in II. 8, Aristotle does not provide an analysis of the notion of 'for the sake of'. Nor does he come down on one side or the other of the Agency/Function construal of teleology. He does argue that teleology is basic for an understanding of natures and natural processes, and suggests that teleology is the key to understanding organized structures which can operate successfully or unsuccessfully. But he goes no further than this. Thus II. 8 by itself is consistent with Aristotle's taking teleological causation to be a basic independent notion (not itself to be further defined). Equally, for all this chapter says, he could have taken it as a basic notion interdefined with natures and natural processes. (*Natures* are what the subject to teleological explanation; *teleological explanation* is what enables us to understand natures.) It is also consistent with his preferring one or other of the two suggested understandings of teleological causation. Equally, for all II. 8 tells us, he may have remained agnostic on the issue of whether teleological causation could be defined further, and contented himself with noting several features of its operation.

This ambivalence, however, is dangerous.[20] For, as it stands, it leaves

[20] This ambivalence is not confined to *Phys.* II. 8, but seems to permeate Aristotle's discussion earlier in *Phys.* II. 3 ff. The remarks in 194^b35–195^a3, and 195^a20 ff. are too general to speak to the present issue. Aristotle seems to wish to ignore the differences between intentional and non-intentional cases (cf. 195^a25–6). At the end of II. 7, he writes of teleological causation as operating 'because it is better so (not without qualification, but with reference to the nature of each thing)' (195^b8–10) but this point is not further developed. It appears to favour the purposive agency model, but does not do so

Aristotle open to the criticism of conflating two quite distinct accounts of teleological causation, or of not having 'one' unified picture of what teleological causation is. And either of these would be damaging because they call into question whether Aristotle had a clear account of teleology on which to base his understanding of natural processes and actions. In Aristotle's own terms there might be more than four causes, if there are two distinct kinds of teleological causation. It remains to be seen how far *Phys.* II. 9 deepens our understanding of this issue. This should illuminate further Aristotle's strategy in the *Physics* on these issues.

IV *Physics* II. 9: Hypothetical Necessity and Teleology

The argument of II. 9 is directed to show that not everything that occurs necessarily occurs absolutely necessarily as the result of simple bodies and their movements. In addition to this mode of necessity, there is hypothetical necessity, which is present in cases of teleological causation. Thus Aristotle can conclude that not everything that occurs not as a result of simple necessity occurs by chance. For there will be in addition to simple necessity and chance a third mode: hypothetical necessity.

In this chapter, Aristotle speaks indifferently of the existence and nature of artefacts (houses, axes), of the occurrence of actions, and of the existence and nature of natural phenomena (200^a31–3). However, he gives greatest attention to the case of action, about which he makes the following claims.

(1) If the goal is to be, it is necessary that certain other things come to be or are present (200^a20–1, cf. $^a11$–12).

(2) If these other things are not present, the goal will not result (200^a11–12).

(3) The goal is not present *because* of these things, although they are its material cause (200^a27–9; cf. $^a13$–15).

decisively. For one might say that (e.g.) birds sit on their eggs because this is the best way to achieve incubation, without meaning more by this than that this method brings about a favoured functional end, and has done so in the past. It need not mean (e.g.) that if the way required to bring this goal about were to change, the bird's nature would adapt. Nor need the bird's nature itself be sensitive to the goodness of the route in follows. As it stands, this text appears to be underdetermined as between the two alternative lines of interpretation suggested here.

(1) may be understood initially thus:

if G is to be, necessarily $A_1 \ldots A_n$ occur.

(2) may be understood as follows:

if $A_1 \ldots A_n$ are not present, G will not be present.

In (3), Aristotle claims that the occurrence of the goal is not explained by the antecedent states $A_1 \ldots A_n$.

(1) and (2) appear to instantiate the general form of hypothetical necessity:

If G is to be, necessarily $A_1 \ldots A_n$ are present.[21]

In these cases, one cannot deduce that

$A_1 \ldots A_n$ are necessary

except on the hypothesis that G is to be. If so, $A_1 \ldots A_n$ are not simply necessary. They would only be simply necessary if G also was necessary, and one could then detach the necessity of $A_1 \ldots A_n$. But in these cases one cannot, since it is not *simply* necessary that G occurs. Hypothetical necessity involves (at least) a modal sufficient condition.

Assume that G is to be, necessarily then $A_1 \ldots A_n$ occur.

In its most direct application to teleological causation, this would be understood as follows:

Given that goal G is to be, necessarily then $A_1 \ldots A_n$ occur.

Aristotle develops an interesting parallel with mathematical reasoning to illuminate this point:

Given that the straight is thus and so, necessarily the triangle has angles whose sum is two right angles.

This analogy is revealing. In it, it is true that

Given that G is thus-and-so, necessarily A_1 occurs.

But it is also true that

Given that A_1 is so, necessarily G is so.

[21] I defend this view of hypothetical necessity in 'Aristotle on Hypothetical Necessity', 5-20. As I understand it, hypothetical necessity may characterize both efficient and teleological causation. For a somewhat contrasting view, which confines hypothetical necessity to teleological causation, see J. M. Cooper, 'Hypothetical Necessity' in A. Gotthelf (ed.), *Aristotle on Nature and Living Things* (Pittsburgh and Bristol, 1985). For reasons which will be clearer below (cf. n. 22), I agree with Cooper that hypothetical necessity is not present in *every* case when 'given P, necessarily Q' is true. The relevant relation is the stronger one: 'given P, necessarily because of this Q'. On this, see 'Aristotle on Hypothetical Necessity', n. 25.

However, while it is because G is so that A_1 is so, it is not because A_1 is so that G is so. If so, Aristotle's notion of 'because' here is stronger than that of a (modal) sufficient condition. One might thus represent the relevant notion as follows:

Given that G is so, necessarily because of this A_1 occurs.

Thus, in these cases, it may be possible that

Given that A_1 occurs, necessarily G is so.

is true, while it is not true that

Given that A_1 occurs, necessarily *because* of this G is so.

Thus, in teleological cases, even if it is true that

Given $A_1 \ldots A_n$, G necessarily occurs.

where $A_1 \ldots A_n$ are the matter for G, it will not be true that

Given $A_1 \ldots A_n$, necessarily *because* of this G occurs.

If this is correct, the matter does not 'hypothetically necessitate' the goal, even if it is modally sufficient for its occurrence. And this is because the matter does not *cause* in the appropriate way the presence of the goal, even though it may *necessitate* its occurrence.[22] The occurrence of the goal is not 'because' of the presence of the matter even in this case. And this explains Aristotle's attachment to claim (3).

In all teleological cases, Aristotle can offer the following account of the form of teleological causation: if a goal G is to be, then necessarily because of this $A_1 \ldots A_n$ will be present. But this form is neutral as between the two types of teleological causation discussed in the previous sections. In the example of purposive agency, $A_1 \ldots A_n$ will be present, given that G is the goal for S, because S will be sensitive to the goodness of $A_1 \ldots A_n$ as means to that goal. This rests on the presence in S of a desire for G, and beliefs or perceptions about the desirability of $A_1 \ldots A_n$ as means to G. However, for natural organisms, this form might be realized because $A_1 \ldots A_n$ function to bring about a benefit G for the organism, and G could not occur without $A_1 \ldots A_n$

[22] The realization that 'because' is thus stronger than 'necessitation' is, of course, consistent with Aristotle's believing that whenever there are true 'because' claims, the results are necessitated. Further, it is fully consistent with Aristotle's insistence that material causes necessitate in ways comparable to that in which premisses necessitate conclusions (198^{b}6–8). Material causes may necessitate the presence of (e.g.) goal-directed states, but do not cause their nature to be as it is (200^{a}26–8). In these ways, the goal is prior to the matter (cf. 200^{a}34–5), as the starting-points of mathematics are prior to their conclusions. This point supports the general account I offer of the role of teleological causation in 'Aristotle on Hypothetical Necessity', 37 ff.

occurring. And this might be true even if the organism was not itself sensitive either to the goodness of $A_1 \ldots A_n$ as means or to the benefit involved in G, provided that the organism itself was organized so as to achieve goal G by these means in given circumstances. And this would be so whether $A_1 \ldots A_n$ were parts of an organized whole, or stages in a process which led to a benefit for the organism. The form of teleological causation in II. 9 is blind to the major differences between these two quite separate models of teleology. It is true that in II. 9 Aristotle concentrates on the example of purposive agency. But this does not show that in this chapter he thought that all cases of teleological causation employed the first, agency, model. In particular, it does not require that the organism itself must be sensitive to the value of the ends or of the means being employed.

Aristotle's silence on this point may be excused to a certain extent. In the case of agency, the agent's desire for G is the desire it is *because* it is for a goal G (cf. *Met.* 1072ᵃ28 ff.). And in the case of an organism's capacity for processes, the capacity is the capacity it is *because* it is for goal G. Similarly, the parts of an organism are the parts they are because they contribute in given ways to the goals of the organism. In each of these cases, 'goal-directedness' is taken to define the relevant parts, capacities, or psychological states, which in favourable conditions produce the goal. Further in each of these cases, the goal sought is beneficial to the agent (or taken to be so). Indeed this is what makes it a goal. So in each case the relevant part, capacity, or psychological state is what it is *because* it is goodness- or benefit-directed. Finally, in the case of agency, given that G is a goal for S, he will do $A_1 \ldots A_n$ (in favourable conditions). And this is because what it is to be a rational agent is to be one who given that G is a goal, does $A_1 \ldots A_n$ (if they lead to G) in favourable conditions. Similarly in the case of organisms, given that G is a goal of the organism, it will do $A_1 \ldots A_n$ *or* possess $P_1 \ldots P_n$ (in favourable conditions). And this is because what it is to be an organism of a given kind is to be something which is organized to achieve its goals (in favourable conditions). Indeed, it is the organism it is because it is organized to achieve those goals in given conditions. The goals, and their implementation, make these organisms what they are.

These points of analogy are, no doubt, important. The goal is identified with the nature of the organism (199ᵃ30–1), and all else is said to be for the sake of it. Given that the goal is identified with its nature, its other capacities will be as they are because they contribute to this goal.

And it is because this is so, that we must begin our account of what (e.g.) the organisms are with their goal (200^a34–5). We could not begin to explain what these parts are without adverting to the goal of the organism. For they are essentially directed towards this latter goal. Similarly, the definition of the whole organism begins with its goal; for this is what makes it the organism it is. The definition may involve elements which are not themselves essentially defined by reference to the goal (e.g. being iron); but these will be derived from elements (e.g. having teeth of a given sort) which are themselves defined by reference to the goal of the organism.[23] For these reasons, Aristotle can claim with justification that to remove the teleological cause is to dispense with natures altogether (199^b15–18).

This said, the lacuna already exposed still persists, and infects the discussion of natural organisms at precisely this point. For if natures are conceived as on the agency model, Aristotle's natural organisms will be sensitive to the goodness of the means they follow to their goals. Thus if the best means to achieve their goals change, they will themselves change accordingly. Conversely, if natures are conceived as on the function model, one could not expect organisms to change their routes to achieve their goals if the best means to them were to alter. Dramatically put, in the first model, organisms would have the ability to change their routes to their goals or their internal parts if the environment were to alter so as to require them to do so if they are to survive. And this would be because they would be sensitive to the fact that the best means to their goals had changed. In the second model, similar changes in the environment would lead to the destruction of the organism if its failure to adapt to the new *best* means for its goals proved decisive in its struggle for survival. On the first view, natures have an internal dynamism for change to meet the new or varying challenges of their environment. On the second, they need not.[24]

[23] This suggests that in II. 9 the notion of matter is being employed to characterize phenomena which are not themselves defined in terms of their teleological role in the life of organisms or artefacts. Of such matter, Aristotle notes that it may be governed by (simple) necessity, and that it is not sufficient to explain the presence of goals (200^a27) except as the relevant matter. This is, of course, consistent with the occurrence of matter being sufficient for the presence of states with goals. But even in this case it would not *explain* the presence of the goal.

[24] These accounts will produce quite different grounds for the eternity of the species (in Aristotle's theory). On the first, species will be eternal because they possess the internal resources to overcome the relevant environmental challenges they face in the best way available to them. On the second, species will be eternal because the environment will not give rise to challenges which require the species to alter their favoured

Aristotle's failure to distinguish between these two models leaves his concept of nature, and of natural organism, importantly unclear in the *Physics*.

V How many causes? How are they individuated?

There are two related difficulties in the *Physics*' account of the four causes that stem from this source. If there are indeed two types of causation within the general framework of teleological causation, Aristotle's doctrine of the *four* causes is (to this extent) undermined. For within the fourth cause—the teleological—there will, in fact, be two quite different kinds of causal relation. And this might be taken to show that (*pace* Aristotle's explicit claim) there are more than four causes in his theory.

It is perhaps interesting that in *Phys.* II. 3 Aristotle is concerned to enumerate different types of causes, and not different causal relations. Thus, it might seem that, if in the two cases mentioned there are two different types of causal relation (in the Agency and Function model), but only one type of cause (namely, what is beneficial), Aristotle can still hold to his doctrine of the four *causes*. Differences between types of relevant causal relation are not sufficient to overthrow this claim.

However, if this is Aristotle's defence, it is superficial. For to understand the answer to the 'Why?' question in these cases (194^b18–19), it is not enough to know that something occurs for the sake of a good or a benefit. One also needs to know how it is that the benefit is connected with its effect. And if there are two distinct ways in which this can occur—either as in the Agency or Function model—Aristotle has not isolated the relevant type of causation merely by adverting to the goal involved. He also needs to show how this produces the relevant effects. Hence, if it is correct to think that he allows for the presence of both these models without distinction, (still worse if he employs them in a confused way), his famous doctrine of the four causes is either superficial or false.[25]

ways of achieving their goals. In both, the eternity of the species is a *consequence* of Aristotle's view of teleological causation (together with other of his beliefs) and not its basis. Nor could it be, since individual mortal agents are subject to teleological causation (even when acting in ways not relevant to the survival of the species). For a contrasting view, see Cooper, 'Aristotle on Natural Teleology'.

[25] Nor will it help to say that one is analogous to the other, since the mechanisms involved in the two cases are quite different. If there is an analogy, it is an extremely weak one.

There remains a further question. In the case of agents, Aristotle insists that there can be more than one cause of the same effect: (thus) both efficient and teleological causes may be at work. And the same is no doubt true of natural organisms with goal-directed capacities. For an organism with a capacity of this type will, on occasion, be the efficient cause of certain beneficial effects by means $A_1 \dots A_n$. And the occurrence of $A_1 \dots A_n$ will at the same time be teleologically caused by the good or benefit to be gained (namely, G). So why are there *two* causal stories in these examples? Why not say that there is just one—an efficient causal one involving capacities of a distinct kind, and treat teleological causation merely as an interesting way of describing the efficient causal process? In other words, why does Aristotle regard teleological causation as a separate objective mode of causation at all?

This objection could be stated as follows. Either Aristotle's teleological causation is a type of efficient causation or it is an illusion. If there is no efficient causal story to support it, teleological causation is not causal at all. One can, no doubt, state claims of the appropriate form in the case of an organism's possession of parts:

> Given G is to be, necessarily because of this $P_1 \dots P_n$ are present.

But these claims are ill-founded because one cannot support or legitimize the claim 'necessarily because of this'.[26] In cases where these claims are supported, it is *because* they are underwritten by an efficient causal story. Thus they are elliptical and should be expressed as follows:

> Given the presence of a capacity for G, $A_1 \dots A_n$ occur (in favourable conditions).

While the capacity may be defined in terms of the benefit it produces, and $A_1 \dots A_n$ may be means to that benefit, the relation between antecedent and consequent in this formula is *au fond* that of efficient causation. So far from there being more than four causes, it appears that there are fewer.

This objection misunderstands Aristotle's strategy in these chapters of the *Physics*. The basis for claims of the form

> Given G is to be, necessarily because of this $A_1 \dots A_n$ are present

[26] Or, if one can, it is because it is an a priori consequence of the animal being the type of animal it is. Thus, one might say that dogs, by definition, have goal G and parts $P_1 \dots P_n$. In this case, the two kinds of modal claim are to be sharply distinguished. In one they are those of efficient causation, in the second that of analytical (or a priori) necessity. Aristotle, it appears, confused these if he thought that there was only one type of necessitation involved.

lies in the *nature* of the animal (or rational agent). This is why, given the nature of S, $A_1 \ldots A_n$ necessarily are present. These claims do not rest on the presence of capacities of the organism which are the efficient causes of the occurrence of $A_1 \ldots A_n$. Rather, it is because G is to be that the capacities themselves are present. Further, while these capacities are causally active (in the efficient mode) to produce $A_1 \ldots A_n$, they are not essential to the explanation of why $A_1 \ldots A_n$ are present. For that rests on S's nature as the type of animal which fulfils G by an acceptable route. The (efficient) causal mechanisms involved, no doubt, are in operation when $A_1 \ldots A_n$ are achieved. But they are not essential to understanding the type of necessitation involved in teleological causation.

This mode of necessitation is, in Aristotle's view, no less objective than that which characterizes efficient causality. For the latter also depends on the nature of the organism in question (see 198ª24–7). Man is the efficient cause of the production of man. Why is this so? Because it is in the nature of man to produce after its own kind.[27] The source of efficient causal necessitation lies in the nature of the agent or organism in question. This view makes Aristotle's views on causal necessitation, at once, more foreign to us and more challenging. On his account, the basis for efficient causal claims of the form

$$\Box\,[\text{water is heated in } C \rightarrow \text{water boils}]$$

lies in the nature of water. It is because water is as it is that it must boil when heated in C. By contrast, the fashionable post-Humean view would take this necessity as basic, and treat the relevant modal claim as true in any world where water exists. By locating the source of the necessity in the nature of the animal or kind, Aristotle legitimizes efficient causal modalities in the same way as he does teleological ones. Both follow from the *nature* of the animal or kind in question. For this reason, he has reason to treat both as equally objective. Indeed, from his perspective, it is only if one (mistakenly) treats efficient causal modalities as basic (and perhaps unanalysable) that there is any temptation to analyse teleological necessitation in terms of them.

[27] This is why Aristotle insists that (1) man produces man (198ª26–7), and (2) the final and efficient causes are similar in type (εἶδος) in this case. Both are essentially involved in specifying the natures (198ᵇ2–7), which are the principal focus of *Phys.* II. These suggestions about Aristotle's strategy in *Phys.* II require further development elsewhere. In particular, more needs to be understood about his view of the nature of efficient causation in these chapters.

Since, in his theory, natures are themselves teleologically understood, teleological necessitation is more basic than efficient causal necessitation. The source of the latter lies in distinct natures which are themselves teleologically defined (198^{b}3–5).[28]

VI Conclusion

These remarks suggest that Aristotle in *Phys.* II. 8–9 did not articulate an account of teleological causation which distinguished properly between the Agency and Function model. He possessed a general strategy for vindicating claims concerning teleological causation; but he did not advance *one* determinate model of teleological causation which he aimed to vindicate. And for this reason, his notions of nature (the key concept in *Phys.* II), and of natural process (the key concept of *Phys.* III) are left, in certain important respects, ill-defined.

This lacuna in the *Physics* could be interpreted in more than one way. One might see it as the result either of confusion, or of a failure to distinguish between two alternative, and quite distinct, models of teleology. Or one might interpret Aristotle more charitably as proceeding in the manner of a cautious builder leaving his options open for future developments and refinements elsewhere (for example, in his biological works). It remains for further work elsewhere to see whether his discussion of these topics outside the *Physics* adds detail or clarity to the model we have considered in this essay.[29] But the lacuna

[28] For some further discussion of these issues, see my 'Aristotle on Hypothetical Necessity', 42–3.

[29] The *Parts of Animals*, in particular, requires study in this respect. Aristotle there speaks of nature as operating like a craftsman which (e.g.) makes the best of the possibilities available to it (687^{a}16–18), provides what is most necessary for survival (674^{b}24), compensates animals for deficiencies elsewhere (674^{b}31, 695^{a}10), and makes organisms suitable for tasks (694^{b}12–13). He also mentions that it is possible to explain phenomena by saying that 'it is well that they are so' (640^{a}37 ff.). These references at first sight favour the attribution of the Purposive Agency Model of teleological causation. But in order to establish that this interpretation is correct, one would need to show (1) that the references to nature acting 'like a craftsman' are non-metaphorical, and actually play a role in Aristotle's account of the *causation* of these phenomena, (2) that, nature selects these options *because* they are the best means to a given goal, (rather than doing something which is, in fact, the best means to the goal without selecting it for this reason), and (3) that if the best means to the goal were to change (e.g. through an environmental cataclysm), nature would for this reason adopt other means to the same goal. I am at present agnostic on these issues of interpretation. If detailed study shows that in the *Parts of Animals* Aristotle did utilize with the Purposive Agency Model of teleological causation in specific cases, this will establish that he was more committed there than in his explicit

in the *Physics*-model reveals a problem in Aristotle's account of several major concepts, such as nature and natural process, which are there defined in terms of it.[30]

and general discussion in the *Physics*. David Balme has argued forcefully that Aristotle did not use the Purposive Agency Model in the biological works; see 'Aristotle's Use of Teleological Explanation', Inaugural Lecture, Queen Mary College, University of London (London, 1965).

[30] An earlier version of this essay was read at a conference on Aristotle's *Physics* held at the University of Southern California in December 1989. I am grateful to those who organized and participated in this event for their comments and criticisms. Earlier discussions with Jonathan Barnes, Robert Bolton, John Cooper, David Depew, Michael Frede, Cynthia Freeland, Allan Gotthelf, Frank Lewis, and Michael Morris were very helpful in developing some of the ideas in this essay. Recently, I have gained considerably from Rowland Stout's account of teleological explanation, and from his insistence on the need to distinguish between functional and purposive explanation. I am further indebted to Lindsay Judson's useful and constructive comments on the penultimate draft. None of those mentioned agree with all of this paper, and some agree with almost none of it.

6

Aristotle's Potential Infinites

WILLIAM CHARLTON

'It is not true', says Descartes, 'that the infinite is understood through denial of boundary or limitation, since on the contrary all limitation contains a denial of the infinite'.[1] As a claim about the finite and the infinite this is notoriously dubious; but something of the sort can, I think, be maintained about finitism: we understand a thinker's finitism by seeing what kind of infinity he rejects and how. Aristotle rejects at least four kinds of infinite. In the first place he holds that there cannot be an infinitely large body or quantity of material. His argument for this conclusion (which occupies *Phys.* III. 5 and *Cael.* I. 5–7) depends mainly on physical considerations about motion and force. He also has metaphysical objections to infinitudes of parts in undivided wholes, infinitudes of points and infinite numbers of actual things generally; but in connection with each of these classes of entity he allows the possibility of what commentators call a 'potential infinite'. Many philosophers profess to find the notion of a potential infinite obscure or question its utility.[2] Although Aristotle's exposition is indeed marred by obscurity, indecision, and even error I shall argue that the concept is fundamentally sound and does apply to parts, points, and actual things. First, however, I shall say something about his views on the impossibility of an infinitely large quantity of material, since these affect his development of his more metaphysical contentions.

There is a Greatest Magnitude

Aristotle holds not only that there cannot be an infinitely large quantity of material but that there is some finite amount that cannot

[1] *Œuvres*, ed. C. Adam and P. Tannery (Paris, 1964), vii. 365.

[2] For example, D. Bostock, 'Aristotle, Zeno and the Potential Infinite', *Proceedings of the Aristotelian Society* 73 (1972–3), 38–9; J. Lear, 'Aristotelian Infinity', *Proceedings of the Aristotelian Society* 80 (1979–80), 188.

be exceeded. He says, 'There cannot be an infinite in respect of addition such as to exceed every magnitude' (*Phys.* III. 207^{a}33–5); 'It will not exceed every magnitude' (206^{b}18–19, cf. ^{b}20–1). To put it formally, he holds:

> (1) It is not possible that there should be an x such that x is a physical reality, and for all y, if y is a finite magnitude, x is greater than y.

But he also holds something stronger, namely:

> (2) There is a y such that y is a finite magnitude, and for all x, if x is a physical reality, it is not possible for x to be greater than y.

At 206^{b}7–12 he says,

> If in a finite magnitude one takes always a limited amount in the same proportion (but not the same fraction of the whole) and adds that, one will not go through the finite amount; but if one always increases the proportion so as to embrace always the same quantity one will go through, since every finite amount is exhausted by [repeated subtractions of] any limited amount [however small].

If we follow the first procedure, if, say, we add first 2/3, then 2/9, then 2/27, etc., we never exceed the original quantity. If, on the other hand, we keep adding the same amount, even if it is minute like 10^{-100}, there is no finite amount we shall not eventually exceed. Aristotle thinks the first procedure can be carried on as long as we please but not the second. Of course there are series of diminishing additions, for example, 2/3, 4/9, 8/27, which will eventually exceed any given magnitude. But Aristotle says at 207^{b}2–5:

> In number there is a limit in the direction of the least but in the direction of more it is always possible to exceed every plurality; but with magnitudes it is the opposite: in the direction of less it is possible to exceed every magnitude but in the direction of more there is not an infinite magnitude.

The claim that 'with magnitudes it is the opposite' implies (2). His going on to say 'there is not an infinite magnitude' rather than 'there is a magnitude that cannot be exceeded' might suggest that he does not distinguish (2) from (1), but I think it is more probably an infelicity due, perhaps, to an unconscious fear of being asked what the magnitude is which cannot be exceeded. On the whole his grasp of the difference between (1) and (2) seems firm enough.

But though Aristotle holds (2) as well as (1) he offers no arguments for (2) as distinct from (1). Why did he accept it? If (1) is true the universe must be finite, but might it not expand indefinitely? Aristotle might have at least two reasons for thinking not. First, so far as he would be able to see, there could be no causal mechanism to account for an expansion. Expansion through growth would require some additional material for the universe to take in, and expansion through rarefaction would surely require some heating agent. Secondly Aristotle conceives the universe as a kind of organic whole (*Cael.* I. 268^{b}8–10); and he says there is a limit to the size and growth of anything of that kind (*An.* II. 416^{a}16–17). With the latter point we may agree. It is not too clear how tall a man there could be, but surely there could not be a man more than fifty feet high. Since, however, Aristotle did not know how big the universe is, he was in no position to name any size it could not exceed.

Aristotle Versus Leibniz on the Reality of Parts

The infinite first appears, we are told (*Phys.* III. 200^{b}17–18) in the division of the continuum. Aristotle's treatment of this forms an instructive contrast with that of Leibniz. Both hold that what is continuous is infinitely divisible, that is, 'divisible into parts that are always further divisible' (*Phys.* VI. 232^{b}24–5, cf. I. 185^{b}10, III. 200^{b}17–20). Leibniz here speaks of 'the labyrinth of the continuum' (G ii. 379[3]). He reasons that a mere aggregate or set has no reality except what it derives from the aggregated items. A pair of diamonds, for example, is real or exists because each of the diamonds exists (G ii. 96). I take it that so far both Aristotle and we would agree with him. But he further thinks that 'a thing which can be divided into several (already actually existing) is an aggregate of several . . . and has no reality except what it borrows from its components' (G ii. 267). A yard of string, he would say, exists because the three feet into which it could be divided exist. Since material objects are conceived as continua and continua are divisible into parts that are always further divisible, it appears to follow that a material thing can never have any reality which is unborrowed, and hence that the notion of a real material thing is incoherent: material objects as ordinarily conceived do not really exist but are only the way in which unextended, mindlike entities appear to us or are imagined.

[3] 'G' abbreviates *Die Philosophischen Schriften*, ed. C. I. Gerhardt (Berlin, 1875–90; reprinted Hildesheim 1965).

There are various ways in which a modern thinker might avoid this disquieting conclusion. We could say that physical objects are no longer conceived as continua: quarks are not corpuscles. Since many philosophers are prepared to believe that every event in the history of the universe was caused by an earlier event, some might accept Leibniz' regress of borrowings as harmless. Rom Harré declares we have no reason to think that physical objects 'are ultimately anything but infinitely complex in structure'.[4] This solution would be closed to Aristotle by his finitism. But Leibniz' problem is anticipated, in essence, in Zeno's Dichotomy paradox, and Aristotle's preferred solution to this constitutes a reply to Leibniz too. The Zenonian problem is that since distance is infinitely divisible to cover a finite distance it is necessary to 'pass through infinitely many halves, and this is impossible' (*Phys.* VIII. 263^{a}4–6). Zeno treats the traversing of a complete distance as an aggregate of the traversings of the distances into which it could be divided. In *Phys.* VI. 233^{a}13–28 Aristotle points out that the time is infinitely divisible too, but in 263^{a}15–23 he says that to provide a really adequate solution we should invoke the doctrine that the halves of a continuum 'do not exist ἐντελεχείᾳ but only δυνάμει' (^{a}28–9). This doctrine appears in many parts of the Aristotelian corpus.

In *An.* III. 430^{b}10 he says that if a length is not in fact divided the parts do not exist except δυνάμει. The same doctrine is stated in the *Metaphysics*. He says that a thing can exist, or be a thing that is, either δυνάμει or ἐντελεχείᾳ, and as examples of things that exist in the former way he gives a statue of Hermes in the uncarved stone and the half of a line (*Δ*. 1017^{a}35–^{b}8). Picking up the point at *Θ*. 1048^{a}30–5 he says:

> ἐνέργεια is a thing's being present not in the way in which it is when we say it is δυνάμει. We say that the statue of Hermes in the wood, for example, or the half in the whole, is [i.e. exists] δυνάμει in that it could be got out; and that the man who is not contemplating is a knower δυνάμει if he is capable of contemplating.

What exactly does he mean? When he says that the parts could be got out he cannot mean that they could be got out in the way in which rabbits in a warren could be got out by ferrets: if that were the case they would exist actually. Commentators usually see no problem here. They take the expression 'exists δυνάμει' as parallel to 'is hot δυνάμει'

[4] Rom Harré, *The Principles of Scientific Thinking* (London, 1970).

in 'the cold is δυνάμει hot' (*Phys.* VIII. 255^b6); and they ascribe to Aristotle the view that the parts exist *potentially* or are *potential* existents.[5] If Aristotle thought that he was seriously confused. To say that *A* is potentially *f*, or a potential *f*, is to say that *A* could become *f*. What is potentially a source of strife could become a source of strife. But what could become something must already be something else and already exist. A potential Pope is an actual human being, to say the least. Nothing, therefore, can exist potentially or be a potential existent. Existence is not, like a room, something into which a thing can have a potentiality of coming. But was Aristotle guilty of this confusion?

The ground for thinking so is his use of the example of the knower (ἐπιστήμων). The knower δυνάμει who is capable of contemplating (or perhaps of contemplating the truth of a particular theorem) but is not actually doing so is surely a potential knower; so the existent δυνάμει should be a potential existent. The non-contemplating geometer, however, is certainly not a potential knower in the way he was before he learnt geometry (*An.* II. 417^a30–$^b2$, III. 429^b8–9) and we have to consider how close the analogy is supposed to be and what point Aristotle is trying to bring out. Aristotle is prepared to call 'analogous' relationships as different as those of power to exercise and matter to 'substance' or form (*Θ*. 1048^b6–9). In the present passage he is trying to bring out the meaning of ἐνέργεια, not of δύναμις, and the point he wishes to make is surely that just as a knower is not an actual knower until something activates the knowledge (in Aristotle's opinion this would be something's *appearing* to the knower) so a part of an undivided whole is not an actual existent until the division is carried out.

How does this bear on the Leibnizian problem? If, with Leibniz, we say 'A divisible whole depends for its reality on the reality of the parts into which it can be divided' we imply that the parts can be referred to. Suppose I have a thick wooden beam and I could saw it into five large cubes. Then 'The beam depends for its reality on the reality of the five cubes into which it can be sawn' must be taken to mean 'There are five cubes, and the beam can be sawn into them, and it depends for its reality on them.' To deny that the parts of an undivided whole exist ἐνεργείᾳ is to deny that 'The beam could be sawn into five cubes' is more perspicuously rendered by 'There are five cubes and the beam

[5] So W. D. Ross, *Aristotle's Physics* (Oxford, 1936), 53; Lear, 'Aristotelian Infinity', 189.

could be sawn into them'. If I say 'I might have driven my car into five large cubes of wood' I do indeed claim that there were (right in the middle of your drive) five large cubes of wood, and that I might have driven my car into them. But dividing a beam into cubes is not like driving a car into cubes. The cubes do not exist in the beam in the way in which obstacles can exist in drives. That may be more evident if we reflect that I might have sawn the beam into twenty-five squares or three planks. If the cubes and the squares and the planks all exist, the beam is three times the size we supposed.

The word to express the status of the parts of an undivided whole is not 'potentiality' but 'possibility'. Dividing into five cubes is a possible way of dividing the beam, so five cubes are a possibility or exist as possibilities. They are like a whole into which a number of separate objects might be assembled. If I say 'Those planks could be built into a hen-house' I do not say 'There is a hen-house and those planks might be built into it'; at most I say that a hen-house is a possibility.

This seems to me an attractively economical way of solving Leibniz' problem, and I think Leibniz himself missed it only through an avoidable mistake. He distinguishes between the actual and the possible or, as he sometimes says, the actual and the ideal, and observes, 'It is the confusion of the ideal and the actual which has embroiled [*embrouillé*] everything and produced the labyrinth concerning the composition of the continuum' (G iv. 491). He says correctly that when parts can be taken in any way we please they are not actual (G ii. 379). But he continues, 'In actuals, simples are prior to aggregates, in ideals the whole is prior to the part' (ibid.), and by this he clearly means that if a whole is actual its parts must be actual and if the parts are merely possible or ideal the whole must be possible or ideal. That is precisely the embroilment. If the parts out of which a whole might be constructed are actual, the whole is merely possible. If I have five actual lengths of wood the box I could construct out of them is a mere possibility. If I have an actual box the pieces of wood into which I could resolve it are possibilities. No doubt when an artefact is merely fitted, nailed, or glued together we are free to treat it, depending on our purposes, either as a true unity or as an aggregate. But in general the principle stands: if the whole is there to be referred to the parts are not, and conversely.

If the parts of an undivided whole exist as possibilities and continua are infinitely divisible, does it follow that for any continuum or material object infinitely many parts exist as possibilities? Although 263^a28 and

$^b5$–6, taken by themselves, could suggest this, and Ross says, 'He accepts too easily the view that the infinitely numerous parts of a spatial whole are only potentially existent in it',[6] we shall see as we proceed that he does not think a division into infinitely numerous parts is a possibility at all.

'Division Everywhere'

In *GC* I. 2 Aristotle considers a problem which arises if we suppose that a body is divisible everywhere (*πάντῃ*), i.e. at every point (316^a14–16, $^b21$–3). If the division at every point actually occurred the body would be resolved into extensionless points, which seems impossible. Hence we must suppose that bodies are composed of extended but indivisible parts. Aristotle replies to this reasoning in a passage C. J. F. Williams[7] translates as follows:

> Since no point is contiguous to another point, there is one sense in which divisibility at every point belongs to things possessed of size and another in which it does not. When this is asserted it is thought that there is a point both anywhere and everywhere, so that the magnitude has necessarily to be divided up into nothing; for because there is a point everywhere it is formed either out of contacts or out of points. In one sense there is a point everywhere because there is one anywhere ÷ and all are like each one; but there is no more than one since they are not consecutive, so it is not the case that there is a point everywhere. For if it is divisible at the middle it will also be divisible at a contiguous point. For position is not contiguous to position or point to point, and this is division or composition.÷ (317^a2–12)

The error in the reasoning to indivisible parts can in fact be brought out with the aid of modal operators and quantifiers. It is correct to say:

(3) For all x, if x is a point on A, it is possible that A will be divided at x

but wrong to say:

(4) It is possible that for all x, if x is a point on A, A will be divided at x.

Aristotle might be accusing his opponents of confusing (3) and (4) at $^a4$–5 when he says they think there is a point 'both anywhere and

[6] Ross, *Aristotle's Physics*, 53.

[7] C. J. F. Williams, *Aristotle's De Generatione et Corruptione* (Oxford, 1983).

everywhere'. 'There is a point anywhere' would then correspond to (3) ('You can divide anywhere') and 'There is a point everywhere' to (4) ('You can divide everywhere'). But as Williams emphasizes in a long and perceptive note, it is extremely difficult to read the following lines as a development of this solution. Aristotle seems to be saying that it is not possible that there should be more points on *A* than one at which it is in fact divided, and supporting this doubtful doctrine by a bad argument. If, he seems to say, *A* could be divided at more points than one it could be divided at two adjacent points; but no two points are adjacent. The natural objection is that if no two points are adjacent, then even if *A* is divided at every point it will not be divided at adjacent points. And why, anyhow, should it follow, if *A* can be divided at two points, that it can be divided at two adjacent points?

Williams may be right that the passage is impenetrably obscure and straining one's eyes in the darkness might simply cause hallucinations. The following interpretation, however, gives Aristotle a line of thought which is not crass. 'In one way there is a point everywhere, in that there is a point anywhere' [i.e. (3)]; 'but there are not more points than one anywhere' [i.e. it is not possible for *A* to be divided at two, we must understand, *adjacent*[8] points anywhere; it is not the case that for some *x* and *y*, *x* and *y* are adjacent, and it is possible for *A* to be divided at both]. 'So there is not a point everywhere' [so (4) is false]. 'For [suppose that (4) is true:] then *A* could be divided at its midpoint and at the next point. But there are no two [a body cannot be divided at two] consecutive points'.

On this interpretation Aristotle is saying that there are adjacent points on an undivided body, and the body can be divided at either but not both. I have to concede that if this is so he has expressed himself very badly. That will be true on any interpretation. But is it not totally un-Aristotelian to suppose that there can be adjacent points on an undivided body? Not totally. Aristotle usually says that when you divide a length in half, the mid-point becomes the end of the first half and the beginning of the second; hence the first half consists of the mid-point and all points before it. But in *Phys.* VIII. 263^{b}9–15 he describes a different way of dividing a time-stretch: you can count the mid-moment only with the second half. The analogous procedure in dividing a one-foot length would be to include in the first half every

[8] This interpretation would be helped by David Sedley's proposed reading at ^{a}9, *οὐκ εἰσὶν ⟨ἐφεξῆς⟩: ἐφεξῆς γὰρ*; but it does not require that reading.

point that is less than half a foot from the start, but not the mid-point itself. You would not then, strictly speaking, be dividing the length at a different *point* from the mid-point. If you adopt the *Physics* VIII procedure the first half-foot will have no last point. But these two ways of dividing the line are such that there is no possible intermediate way of dividing it. Aristotle is surely right that whereas in thought at least we can divide a length in either way, it cannot actually be divided in both ways. Any two points at which it is in fact divided must be such that between them there is a further point at which it could have been divided. When he talks about the division of movements we shall find him distinguishing actual and possible points, and the distinction might helpfully have been drawn here. Possible points would be points at which a body might be divided and actual points, points at which it is divided. He could then say that possible points are next to one another and correspond to infinite series of rational fractions, whereas actual points are not next to one another and correspond to the rational fractions themselves. Not that between any two actual divisions there is a third actual division; but there always could be.

Possible and Actual Points

The distinction between possible and actual points is drawn in *Phys.* VIII. 8. Speaking of movement in a straight line Aristotle distinguishes intermediate points from terminal points. It is not absolutely clear whether he wants at 262ᵃ22–5 to say that an intermediate point does not actually exist, or only that it is not actually intermediate, 'unless the movement is divided at it and the moving body comes to rest and later starts to move again'. I think he means both; and certainly at 262ᵇ31–2 he declares that the intermediate points exist δυνάμει and only a point at which something actually stops exists ἐνεργείᾳ. David Bostock, however, maintains that the distinction is faulty and that its unsoundness invalidates Aristotle's treatment of the Zenonian problem.[9]

Holding that continua are infinitely divisible Aristotle must hold that no finite number is the greatest number of parts into which a mile, for example, can be divided. (For our present purposes it does not matter whether we take the mile to be a mile of, say, ground across which a moving body moves, or simply a one-mile distance it moves,

[9] Bostock, 'Aristotle, Zeno'.

though I think it more elegant to take it the latter way.) Bostock rightly, I think, understands Aristotle to deny that it is possible for a mile to be divided into an infinite number of parts, and rightly represents him as holding that a body divides a mile and actualizes parts of it if it actualizes a point on it. His criticism is that Aristotle has no good grounds for denying that a moving body could actualize infinitely many points. A body actualizes a point by 'doing something' at it (pp. 41–2). But a body can do something at an infinite number of points in a finite distance in a finite time. In the first place, it can *reach* or *arrive at* infinitely many points. Aristotle does indeed 'argue that this is not after all true' but the argument is 'fallacious' (p. 42). Secondly Aristotle admits or rather tries to prove that a moving body divides a movement and actualizes a point when it reverses direction, and a body might reverse direction at infinitely many points. Imagine a bouncing ball the mid-point of which rises after each fall to half the height from which it fell, and which completes each bounce in half the time of the preceding bounce. The mid-point of the ball will change direction infinitely often, actualizing infinitely many points, in a finite time. So 'Aristotle does not, after all, have the right answer to Zeno's problem' (pp. 43–6).

This critique is unsatisfactory. First, it is unfair to attribute to Aristotle the vague idea that a body actualizes a point by 'doing something' at it. He has the quite precise idea (262^a22–5) that a body actualizes a point by coming to rest at it.

Next, it is not clear that Aristotle argues that a moving body does not arrive at points it passes through. What he says is that if we suppose 'a body *A* to have reached a point *B*' which it passes through 'and to have left it, it will always come to rest. For it is impossible for it simultaneously to have arrived and to have left'. This is not so much fallacious as obscure: why not say that at any moment at which *A*'s reaching *B* is past, *A*'s leaving *B* is past? And in any case what is excluded is less arriving than having arrived.

Although Aristotle does not say so in so many words, I suspect he may have felt obliged to exclude arriving by some reasoning such as this. (i) If *A* passes through *B* without stopping, then if *A* arrives at *B* and leaves *B* there can be no time interval between arrival and departure. (ii) So arriving at *B*, leaving *B*, and being at *B* must be simultaneous. (iii) So they must be a single event under three different descriptions. (iv) But that is impossible. Although (iv) looks plausible at first, Aristotle's own conception of continuity should have led him to reject it. If *A* passes through *B* without stopping, the move to *B* is continuous with the movement from *B*, and 'things are continuous if their ends are identical' (*Phys.* VI. 231^{a}22). To say, however, that reaching *B*, being at *B*, and leaving *B* are a single event under three descriptions is not to say that there is an actual occurrence which is so describable. *A*'s passing through *B* without stopping is not an occurrence of a non-stop, but a non-occurrence of a stop. It is no more an actual event than the point at which stopping does not occur is an actual point.

Discussing a little later a case like Bostock's ball, Aristotle says:

> If a body *E* moves to a point *D* and reverses direction and moves down again, it uses the terminus *D* as end and beginning, it uses the one point as two, and therefore must come to rest. And it has not simultaneously reached *D* and left *D*, for if it had it would both be there and not be there at the same instant (262^{b}23–8).

Aristotle's reasoning here seems to be (i) if *E* does not come to rest, its having finished the movement to *D* must be simultaneous with its having started the movement from *D*. (ii) At any moment at which *E* has completed the movement to *D*, *E* is at *D*, and at any moment at which it has started the movement from *D* it is no longer at *D*. (iii) So if *E* does not come to rest, then there is an instant at which *E* is both at *D* and no longer at *D*. This reasoning is certainly unsound, and the flaw seems to be in (i). What Aristotle should have said is: 'If *E* does not come to rest, its *finishing* the movement to *D* must be simultaneous with its *starting* the movement from *D*'. But the argument cannot then proceed. For *E* starts the movement from *D* at *D*, not when it is no longer at *D*.

Aristotle, then, fails to show that *E* must come to rest at *D*, and in fact I agree with Bostock that it does not. But that deprives Bostock's criticism of all but *ad hominem* validity. If the bouncing ball does not really stop at the top of each bounce, and if, as Aristotle maintains, it would have to stop to actualize a point and divide its movement, then its movement is not, after all, divided into infinitely many parts but continuous. Hence Aristotle does not have to admit that a movement could be divided into infinitely many parts and his solution to Zeno's problem stands. It is merely debarred to himself so long as he thinks that in order to reverse direction a body must stop.

It might be asked whether there are any good grounds for holding that a point can be actualized only by stopping. I do not feel this is the right place for a proper answer, but I shall make a couple of remarks. First, we may speak both of points at which a body, and of points at which a movement, can be divided. It is only the latter which would be actualized by stopping, and application of the δύναμις–ἐντελέχεια distinction to them can hardly be discussed independently of its application to movements. Second, movement, or at least translation, is change in distance from something, and stopping is ceasing to change in distance. If stopping divides a movement so, presumably, would starting or ceasing to change speed or direction; though simply passing through a speed or a point of the compass would not divide the movement.

The 'Potential Infinite'

Aristotle could maintain that an undivided body does not contain an infinite number of actual parts, and that a body like Zeno's runner or Bostock's ball does not make an infinite number of actual movements, without denying the possibility of dividing a continuum into infinitely many parts. Nevertheless it is commonly and, I think, correctly, supposed that he does deny this. The denial is a consequence of his finitism concerning numbers of actual things. But what exactly does this finitism amount to? The passage to which we first look for enlightenment (*Phys.* III. 206ᵃ14–33) is compressed and bewildering. Aristotle seems to say that the infinite exists in connection with time, the division of magnitudes, and number, and then draws at least half a dozen distinctions. We might paraphrase, numbering the distinctions, as follows:

(i) There are two ways of being, δυνάμει and ἐντελεχείᾳ (ii) and two kinds of infinite, by addition and by division. We have already said that there is a finite size things cannot exceed [so that rules out (iii) one kind of infinite by addition completely]. There can be infinites by division. But the infinite exists only δυνάμει. (iv) But it does not exist δυνάμει in the way in which a statue exists δυνάμει if something could be made into a statue. There are a lot of different ways of being. (v) The being of a day or an athletic contest is different from that of a man or house. (vi) And the distinction between being δυνάμει and being ἐνεργείᾳ applies to things like athletic contests too. (vii) The infinite is different in time and for men and for the division of magnitudes. But in general it is that outside of which there is always more to take.

Some of these distinctions are unproblematic. The impossibility of infinitely many additions of the same quantity, as distinct from diminishing quantities—(iii)—has already been noted (p. 130). Distinction (ii) also is straightforward. Speakers of English might express (v) with the aid of 'exist' and 'occur' or 'take place'. Men and houses exist whereas days, i.e. revolutions of the fixed stars, and athletic contests take place. Existence and occurrence are the ways of being actual which attach to objects and processes respectively. Hence we may explain (vi) by saying that a process exists δυνάμει if it might occur, and ἐντελεχείᾳ if it is occurring. As to (vii), the infinite in time is presumably a series of processes, namely sidereal revolutions, and the infinite for men a series of objects, not all existing simultaneously. The infinite in division would be separated parts, object-like entities, existing simultaneously.

It remains, then, to clarify (i) and (iv). Aristotle apparently holds that the infinite exists δυνάμει in connection with sidereal revolutions, members of biological species, and parts of magnitudes; and he later adds that it also exists ἐντελεχείᾳ in the same way as an athletic contest. By this he presumably means that the series of sidereal revolutions, the series of men, the cutting up of pieces of material, all go on. So the question arises: in what way does the infinite exist δυνάμει and not ἐντελεχείᾳ? I shall first sketch the answer I favour, and consider some alternative answers in the course of defending it.

Let S be a set or series of sidereal revolutions, men, or other entities in connection with which the infinite exists; and let R be a relation which is transitive and irreflexive like being larger than. Then the claim that the infinite exists δυνάμει in connection with S (which English-speaking commentators formulate as 'S is a potential infinite') is the claim:

(5) For all x, if x is a member of S, there could be a y such that y is a member of S, and y is R to x.

The claim that it exists ἐντελεχείᾳ ('S is an actual infinite') is:

(6) For all x, if x is a member of S, there *is* a y such that y is a member of S, and y is R to x.

It is obvious that (5) does not entail that S *could* be an actual infinite. It does not entail:

(7) It is possible that for all x, if x is a member of S, there is a y such that y is a member of S, and y is R to x.

And Aristotle's claim that the infinite does not exist ἐντελεχείᾳ is the claim that (7) is false. When he says (iv) that the infinite does not exist δυνάμει in the sense that it could exist ἐντελεχείᾳ he is rejecting (7) and showing that the thesis that it exists δυνάμει should be understood as in (5).

If this interpretation is correct Aristotle's position concerning numbers of actual things is a clear and controversial kind of finitism which some people today wish to attack and others to defend on its own merits; a survey of some recent controversy is provided by Richard Sorabji.[10] But simple and attractive as it seems to me to be, this interpretation has to face a number of objections.

Objections and Replies

First of all I am suggesting that 'the infinite exists δυνάμει in connection with the division of a magnitude' should be understood as equivalent to: 'For all *x*, if *x* is a part cut off from a magnitude, it is possible that there should be a *y* such that *y* is a part cut off from that magnitude, and *y* is less than *x*'. It might be complained that this is a deviant use of δυνάμει, the normal usage being that in 'The cold is δυνάμει hot'. Bostock (p. 38) makes some such observation and adds 'but in fact it is not the sense of "potentially" that he [Aristotle] mentions as unusual but the relevant sense of "is": the relevant sense of "is" is the sense in which a process such as a game may be said to "be".' This seems to me an incorrect reading of the passage. Bostock is explaining 'We must not take "is δυνάμει"' at 206^{a}19 by distinction (v); he is running together (iv) and (v). But they are clearly different; (v) is a fresh point; and at (iv) Aristotle is issuing a warning about his use of δυνάμει, a warning reiterated at *Θ*. 1048^{b}9–15.

Bostock rightly interprets Aristotle as denying (7) though he thinks that the denial is not properly expressed and also that it is arbitrary. Jaakko Hintikka suggests that for Aristotle the infinite exists δυνάμει only in that it has the kind of being which attaches to processes, i.e. it rather occurs than exists, or occurs successively instead of existing all at once.[11] This is an implausible interpretation and as Jonathan Lear (pp. 190–2) points out, combined with other Hintikkan doctrines it

[10] R. Sorabji, *Time, Creation and the Continuum* (London, 1983), ch. 14.

[11] J. Hintikka, *Time and Necessity: Studies in Aristotle's Theory of Modality* (Oxford, 1973), 116.

would commit Aristotle to saying that (at least in infinite time) for every part cut off there actually is a smaller part cut off.

English-speaking commentators make things harder for themselves by using the phrase 'potential infinite'. The Greek for that would be τὸ δυνάμει ἄπειρον. In fact, however, Aristotle uses δυνάμει not with ἄπειρον but with ὄν, something I have tried to recapture by 'the infinite exists δυνάμει in connection with . . .'. As I said earlier, things cannot exist potentially. On my interpretation Aristotle uses δυνάμει in three different ways when he says that the cold is δυνάμει hot, that the Hermes exists δυνάμει in the wood, and that the infinite exists δυνάμει. Such a variety of uses does not seem to me at all uncharacteristic.

A second objection is this. On my interpretation Aristotle rejects (7) and *a fortiori* (6). But it appears that there are some sets of series for which he accepts (6). At 272ᵃ1–2 he says, 'We say that number is infinite in that there is no greatest number'. To put it formally:

(8) For all x, if x is a finite number, there is a y such that y is a finite number, and y is greater than x.

I reply by distinguishing between numbers, the finite cardinal integers, and numbers of objects—that is, sets with one or more members. Aristotle's finitism concerns sets or series of things like men, sidereal revolutions, and parts which have actually been cut off. These are what P. T. Geach calls 'things that actually exist' in contrast with entities such as numbers which exist, indeed, but not actually.[12] I am not aware that Aristotle uses δύναμις and ἐντελέχεια (unless perhaps at *Met. M.* 1078ᵃ29–30) to mark this distinction, but he is surely alive to the distinction itself. He would presumably hold that numbers do not exist in the same way as men or even sets of men, on the ground that for a number to exist is for a set with that number of members to be *possible*.

But does not Aristotle also accept (6) for entities which actually exist by Geachian standards, including men and sidereal revolutions themselves? Does he not think that for every sidereal revolution there has been an earlier, and for every human being a parent? There are passages where this is said or implied. But it is difficult to reconcile them with other passages where Aristotle says or implies that the infinite does not exist otherwise than δυνάμει.

[12] P. T. Geach, 'What Actually Exists', *Proceedings of the Aristotelian Society* Suppl. vol. 42 (1968).

The traditional view (appearing in Philoponus, Aquinas, etc.,) is that Aristotle allows there can be infinitely many successive things but not infinitely many simultaneous things. This interpretation raises several awkward questions. First, why does he say that the infinite exists only δυνάμει? It will not do to reply 'Because its existence is of the kind that attaches to athletic contests'. Not only does Aristotle say that both kinds of being, δυνάμει and ἐντελεχείᾳ, are open to processes, there is no ground for saying that a set of actual things which exist one after another is less of an actual aggregate than a set of things which all exist at once. Next, why cannot there be infinitely many simultaneous men? If the reason is that there cannot be an aggregate of infinitely many actual things at all, then (for the reason just given) this objection applies equally to aggregates of infinitely many successive objects. An alternative reason would be that there cannot be an infinite quantity of material: if there were infinitely many simultaneous men there would be an infinite quantity of human flesh. In the first place, however, this does not rule out an infinite number of simultaneous objects diminishing in size, like the parts which would be produced by repeated dichotomous division; yet the traditional view (here surely correct) is that Aristotle would reject such an infinite. In the second place, Aristotle maintains, as we saw, not only that there cannot be an infinite quantity of material, but that there is a finite quantity which cannot be exceeded. That being so I think it is easiest to interpret *Phys.* III. 6–7, at least, as allowing only what I call potential infinites even of successive things. When Aristotle says at 206^{a}33–^{b}3 that the infinite exists 'in the case of magnitude, with what is taken remaining, but in the case of time and men, with what is taken passing away', he is saying that the set of simultaneous parts produced by division is a potential infinite but the set of simultaneous men is not: for any number of men, there might have been more successively, but for some number there could not have been more simultaneously.

Edward Hussey offers the following reconstruction of Aristotle's finitism, using '*F*' as a place-holder for predicates like 'man':[13]

(i) The number of *F*s existing at any one moment is finite.

(ii) The number of *F*s existing between any two moments (a finite time, presumably, apart) is finite.

(iii) Before a given time there could have existed an infinite number of *F*s.

[13] E. Hussey, *Aristotle's Physics, Books III and IV* (Oxford, 1983), pp. xx–xxiii.

(iv) Aristotle's 'central notion' of the infinite is that *F*s are an infinite if for any future moment *t* there will be, or it is possible that there will be, an *F* which has not yet existed by *t*. For example sidereal revolutions are an infinite because there will never be a time after which there will not be another.

(v) Such an infinite is potential because it has unfulfilled potentialities: it always will or could have members it does not yet have.

The second thesis is not explicit in Aristotle but I have no doubt Aristotle would accept it. It would be entailed by the first together with the plausible proposition that there is some finite time for less than which no *F* can exist. Hussey's interpretation, however, seems to me open to at least two objections. First, the fifth thesis does not fit Aristotle's text. Aristotle does not say that the infinite has unfulfilled potentialities; he says it exists δυνάμει. The difference between existing δυνάμει and existing ἐντελεχείᾳ cannot consist in having or not having unfulfilled potentialities; otherwise we should all exist δυνάμει.

Second and more serious, the fourth thesis does not fit with the first and third. According to the first, the set of men now living should have only a finite number of members. But since there (probably) will be born additional men, the 'central notion' of the infinite applies to it. Hussey has, indeed, a reply to this. If we take *F*s as men, *F*s are indeed an infinite, but if we take *F*s as men living now, they are not; for there will not be born an additional *F*. Additional men will be born, but not additional men living now. This reply depends on there being dated properties, the property not just of being a man but of being a man-in-1983. Hussey is not the first philosopher to invoke such properties. McTaggart postulates the property of being red-on-Monday in order to prove that pokers cannot change, and the only possible change is an event's changing its *A*-determination.[14] But Aristotle thinks pokers do change, and I do not believe he would accept dated properties. Hussey requires us to understand his first thesis as 'For any time *t*, the set of *F*s-at-*t* is [?timelessly] finite.' The thesis Aristotle accepts is surely 'For any time *t*, the set of *F*s is-at-*t* finite.'

Suppose, however, we accept dated properties: then even if the sidereal revolutions which have already occurred are infinitely many the 'central notion' will not apply to them. For though there will be

[14] J. McT. E. McTaggart, *The Nature of Existence*, ii (Cambridge, 1927), ch. 33.

further revolutions, there can be no further revolutions-prior-to-1983. Hussey's account, then, has the consequence that Aristotle either counts as infinite aggregates we could as finite, or as not infinite aggregates which we count as infinite.

In places outside pages xx–xxiii Hussey suggests doubts about whether Aristotle really thinks every revolution might have had a predecessor: that would 'involve various undesirable infinities' (p. xlvii) and make him 'dubious whether this infinity is more than potential' (p. 171; Hussey here seems to be going along with more traditional ideas of what Aristotle means by an infinite that exists δυνάμει). I agree. I also think Aristotle would have been less ready than Aquinas (*Summa Theologiae* I. 46 2 ad 6) to believe that the whole series of historical events could have taken an infinite length of time, given (*Phys.* IV. 222^a24–9) that no event in the series occurred more than a finite time ago. On the other hand he cannot see how there could have been a first event (*Phys.* VIII. 251^a9–28). I am inclined to think that in the face of this Kantian antinomy Aristotle cannot make up his mind. Jonathan Lear (pp. 206–8) suggests an escape route. Aristotle could say that the successive states of the universe are a 'vaguely determined totality'. That would allow him to say that there is no first event of any specified kind, e.g. no first human birth, and yet insist that there is some number that is not exceeded—there have not been 10^{100} human births. Whether this solution would appeal to Aristotle I am unsure. He is attached to his revolutions of the heavens, and they hardly seem to form a 'vaguely determined totality'. But the official teaching of *Phys.* III. 6–7 is:

(9) For all *x*, if *x* is a revolution of the heavens, it is possible that there was an earlier revolution.

It is reluctantly, if at all, that Aristotle accepts

(10) It is possible that for all *x*, if *x* is a revolution, there was an earlier one.

The third and last objection I wish to consider is easily formulated: Why should Aristotle hesitate to accept (10)? Granted that (7) does not follow from (5), why should not both be true? Hussey makes three suggestions. Aristotle thought that the infinite cannot be traversed and if there were an infinite number it would be (pp. xxiii–xxiv); he thought the infinite formless and unknowable (p. xxiv); and he thought past time was somehow mind-dependent (p. xlviii). Although all these suggestions have a basis in the *Physics* (204^b5–10, 207^a21–32 and

223^a21–9 respectively) they do not constitute very good reasons for rejecting (7) or (10).

An infinite set is one that can be put into one-to-one correlation with a proper part of itself; the number of its members is not affected by the addition or subtraction of one. A possible motivation for rejecting (7) and (10) would be a feeling that if S is a set or series of things which are actual in Geach's sense, the number of its members must be affected by the addition or subtraction of one. The number would be unaffected only if the members of the set were mere possibilities. The number of times a moving body might but does not stop is unaffected by the number of times it does stop; but these non-stoppings, as I said earlier, are non-events. I should like to add some further considerations.

What is the difference between (9) and (10)? How, indeed, should (10) be understood? The possibility in (9) is logical: there is nothing in the definition of a sidereal revolution to put a limit on how many there can have been of them. Presumably the possibility in (10) is logical too. It would be feeble to construe it as epistemic and take 'It is possible that' to mean 'For all we know to the contrary it is the case that'. The difference between (9) and (10) ought to be analogous to the difference between:

(11) For all x, if x is entered for this race, it is possible that x will win.

and:

(12) It is possible that for all x, if x is entered for this race, x will win.

(11) says that there is no horse that is hopeless, (12) that the race might be a dead heat. The difference is fundamentally that between:

(13) $\Diamond p \,\&\, \Diamond q$

and:

(14) $\Diamond (p \,\&\, q)$

Suppose, then, that we number the revolutions, going back from the birth of Christ, BC1, BC2 etc.; (10) will say:

(15) It is possible that (BC1 had a predecessor & BC2 had a predecessor & BC3 . . .)

where no revolution that occurred is not said to have had a predecessor.

It might be objected that (15) says it could be that every revolution which has occurred in the actual sequence of revolutions had a predecessor, whereas what is in question is whether it could be that there is some sequence every member of which had a predecessor. (10), it might be claimed, is analogous not to 'It could be that every horse in this race will win' but 'There could be a race in which every horse wins.' It seems to me, however, that we want to know precisely whether it could be that every revolution in this sequence has a predecessor. Besides, to suppose there could be a sequence in which every member *has* a predecessor is to suppose there could be an S such that (M_1 has a predecessor & M_2 has a predecessor & M_3 . . .) where M_1, M_2, M_3 etc. are all members of S, and I do not see how we can make sense of this supposition if we cannot make sense of the supposition there could be an S such that *it is possible that* (M_1 has a predecessor & M_2 had a predecessor & M_3 . . .).

But (15) could be true even if there has been only a finite number of revolutions. Suppose there have been just three. We can still say 'It is possible that all of the following should be true: BC1 had a predecessor & BC2 had a predecessor & BC3 had a predecessor.' BC3 did not in fact have a predecessor, but it could have had. Hence (15) turns out to be asserting not that the number of revolutions might have been infinite but that it might have been greater than it was.

This last proposition is, indeed, ambiguous. It could mean:

(16) For some x, the number was equal to x and might have been greater.

But it could also mean:

(17) For some x, it might have been the case both that the number was equal to x and that it was greater than x.

If we define being more numerous than v as having a proper part which is equinumerous with v then (17) is not self-contradictory; rather it asserts precisely that the number of revolutions might have been infinite. If there were $\aleph_0$ revolutions, the number of revolutions will be equal to the number of finite integers, and so will the number of revolutions not counting BC1. But (17) is hardly a possible interpretation of (10). Rather it is a fancy way of saying:

(18) There is an x such that x is a transfinite number, and the number of revolutions might have been x.

Now there are plenty of transfinite numbers, but it does not follow that the number of actual things in any set or series might be one of them.

There are vulgar fractions, but while the number of bottles of wine I drink might be 7/3, the number of children I have could not be. There are negative numbers, but even the peers in *Iolanthe* saw that the number of years old a mother is when she gives birth to a son could not be minus 8. The probabilities, I think, are rather against than in favour of the supposition that Aristotle should have thought the number of revolutions there have been might be $\aleph_0$. Contrary to first appearances, (10) goes no way towards making us see how this could be a genuine possibility. I adhere, then, to my claim that so far as actual things are concerned, Aristotle wants to accept (5) and reject (7); and this is a finitist position which I myself would be unwilling to dismiss without better arguments than I have yet seen.[15]

[15] Much of the material in this paper was presented to the Boston Area Colloquium in Ancient Philosophy in two seminars at Brown University in September 1988. I am grateful for the perspicacious remarks made then, and also for many helpful comments by the editor of this volume.

7

Aristotle on the Reality of Time

MICHAEL INWOOD

At *Phys.* IV. 217^b32 ff. Aristotle asks whether time is 'among the things that are or among the things that are not' (217^b31).[1] He presents three 'exoteric' arguments which suggest that time either 'is not at all or scarcely and dimly' (217^b32–3):

Ia. Some of it has been and is not, while some of it is to come [μέλλει] and is not yet. Both infinite time and any given time [i.e. period or stretch of time] consists of these. But what consists of things that are not cannot, it would seem, participate in being [μετέχειν οὐσίας]. (217^b33–218^a3)

Ib. Of any entity consisting of parts it is necessary that, if it is, either all or some of its parts should be when it is. But whereas time consists of parts, some of them have been and some are to come, but none of them is. The now is not a part. For a part measures [the whole] and the whole must consist of parts. But time seems not to consist of the nows. (218^a3–8)

II. It is not easy to see whether the now, which seems to bound the past and the future, always remains one and the same or is successively different. Suppose (i) that the now is always different. Then, assuming that no two distinct parts of time can be simultaneous (unless one *includes* the other, as the greater time includes the lesser), and assuming that the previous now which no longer is must have perished at some time, then the nows too will not be simultaneous with each other, and the previous now must always have perished. (*a*) The now cannot have perished in itself, since that is when it is; and (*b*) it cannot have perished in any other now, for, granted that nows, like points, cannot be next to one another, the now would have perished not in the next now but in another [later] now and would thus be simultaneous with the infinitely many intervening nows. But this is impossible. But (ii) the now cannot remain always the same. For (*a*) nothing that is divisible and bounded has only one limit, whether it is continuous in one dimension or several. But the now is a limit, and it is possible to take a limited time. And (*b*) if to be simultaneous in time and neither before nor after is to be in one and the same now, then if both

[1] I am grateful to Dionisios Anapolitanos, Lindsay Judson, and Vasilis Karasmanis for their valuable comments on earlier drafts of this essay.

earlier and later [events] are in this [persistently identical] now, events of ten thousand years ago would be simultaneous with today's events and nothing would be before or after anything else. (218^{a}8–30)

Ib supplements Ia: it answers the objection that at least the present 'is'. I shall refer to these two arguments together as 'Problem I'. II might be seen as a reply to Problem I, casting doubt on the existence of the now on which Problem I depends, but Aristotle presents it as a distinct argument against the reality of time, casting doubt on the existence of the now on which time depends. I shall therefore refer to it as 'Problem II'. Aristotle provides no explicit solution to either problem. I shall ask whether he provides implicit solutions to them. I initially consider the two problems separately, but in the final part of the essay I argue that Aristotle's attempt(s) to solve Problem II are also relevant to the solution of Problem I.

I Problem I

1. Problem I may be rehearsed at any time: according to it, it is never true to say that time or any part of time is (now). Thus one premiss of the argument, namely that 'Some of [time] has been and is no longer, while some of [time] is to come and is not yet', is arguably false. For on a plausible, and possibly Aristotelian, account of past- and future-tense statements, the truth of a past-tense statement implies that there was a time when the corresponding present-tense statement was truly assertible, and the truth of a future-tense statement implies that there will be a time when the corresponding present-tense statement is truly assertible.[2] But since, by Problem I, there never was nor ever will be a time when it can be truly asserted that some of time is, we cannot truly say, now or at any other time, that some of time either was or will be. But this hardly subverts Problem I: the premiss might be rewritten so as to exclude the past- and future-tense claims (or to sever their links with present-tense counterparts) and to confine itself to the main point, namely that no part of time is now. These considerations do cast doubt on the claim that time consists of stretches of time rather than of nows (since no stretch of time ever is, only a now); but this claim too

[2] Cf. *GC* 337^{b}4 ff.: 'If it is true to say that something will be [ἔσται], it must at some time be true to say that it is. But if it is true to say that something is about to be [μέλλει], there is nothing to prevent its not happening.' (Aristotle uses μέλλει in stating Problem I, but it is unlikely that he has in mind a contrast with ἔσται.)

might be eliminated from Problem I. Alternatively, we might say that the force of Problem I is not diminished by its implicit appeal to the negation of its conclusion, since to argue 'If *p*, then not-*p*. So not-*p*' (or more elaborately, 'If *p* and *q* and *r*, then not-*p*. But *q* and *r*. So not-*p*') is a valid way of arguing for the truth of not-*p*.[3] If anything the case for the non-being of time is strengthened rather than weakened by the additional claims that time was not and will not be, as well as is not.

2. Aristotle asks whether time is τῶν ὄντων or τῶν μὴ ὄντων (217ᵇ31). This implies that some things are, even if time is not among them. But what things? Problem I can be naturally extended so as to conclude that no temporal entity, no event or object, is: what exists now is not a κίνησις, but an instantaneous cross-section of a κίνησις (a διαίρεσις, *Phys.* 220ᵃ19, or a κίνημα, *Phys.* 241ᵃ4, where Aristotle says that a κίνησις does not consist of κινήματα); what exists now is not an object, but an instantaneous cross-section of an object. It may be that τὰ ὄντα include only things that are not in time, such as the incommensurability of the diagonal of a square with its side (222ᵃ5–6). But Aristotle gives no hint that he is using ὄντα so restrictedly. Moreover he shows no sign of wishing to make these natural extensions of Problem I. He takes it for granted that objects are ὄντα; if Homer is not (although he once was, 221ᵇ32–3, and although he is a poet, *Int.* 21ᵃ25–6), this is because he is long dead, not because he presents only a punctual cross-segment at any given moment. At *Phys.* 206ᵃ21 ff. he acknowledges that for an event like the Olympic games to occur is not for it all to happen at once, but he does not suggest that all that exists now is a κινήμα.

There may be several reasons why Aristotle does not extend Problem I to objects and events: (i) He is not very interested in scepticism, unless it serves to develop his positive views. (ii) He perhaps feels that since, on his account, time is a derivative phenomenon, a πάθος of κίνησις (223ᵃ19), the ontological instability of time cannot automatically affect the ontological stability of objects and events, apart from such uncontroversial claims as that Homer is not. (iii) In the case of objects it is less easy to argue that none of their parts is at any given time, since the parts of an object are more naturally taken to be its contemporaneous physical parts (e.g. arms and legs) than the segments of

[3] A similar reply can be made to the objection that Problem I implicitly presupposes the reality of time by its use of tensed verbs or by engaging in the temporal process of arguing.

its temporal career (e.g. infant, child, adolescent, etc.). By contrast, the parts of a motion or event are more naturally taken to be successive stages than contemporaneous aspects or elements of the motion or event: the parts of a train journey from Oxford to London are the journeys from, say, Oxford to Reading and from Reading to London, rather than the journeys of the engine and of each carriage from Oxford to London. Motions are immediately vulnerable to Problem I in a way that objects are not. At all events Aristotle's reluctance to extend Problem I to objects and events plays a crucial role in his answer to Problem I (and II): the ontological status of time is secured by the fact that it rides on the back of existent and persistent objects.

3. At *Int.* 9 Aristotle argues that the future differs from the past (and present) in that if there was a sea battle yesterday, it is true *now* that there was a sea battle yesterday, whereas if there will be a sea battle tomorrow, it does not follow that it is true *now* that there will be a sea-battle tomorrow: future events and the truth-value of future-tense statements are indeterminate in a way that past events and the truth-value of past-tense statements are not. There is thus a sense in which the past is among τὰ ὄντα while the future is not. Hussey implies that this doctrine plays an important part in Aristotle's solution to Problem I (and perhaps II): 'The solution that passage [219^b9–33] seems to favour is the conceding of some kind of not too tenuous reality to the past'.[4] There are however several reasons for thinking that the *De Interpretatione* doctrine is irrelevant to Problem I: (i) Problem I concerns the past and future as such, not past and future events. Neither in the *Physics* nor in *De Interpretatione* does Aristotle doubt that there will be future time. It is true that the existence of time, and therefore of the future, depends on the occurrence of changes; but Aristotle tends to argue not that the continuance of time is therefore precarious, but that, since time cannot cease (because the now is essentially the beginning of the future, 222^b6–7), the appropriate changes must occur (*Phys.* 223^b12 ff., 250^b11 ff.). (ii) Consequently not all future events are undetermined now: the heavens will revolve in the future as they have in the past (223^b12 ff., 250^b11 ff.). (iii) The *Physics* contains no hint that past and future differ in reality or determinacy, nor is it obviously entailed by the doctrine of the persisting now (219^b9 ff. See Part II,

[4] E. Hussey, *Aristotle's Physics, Books III and IV: Translated with Notes* (Oxford, 1983), 140. Cf. pp. xxii f., xlvii, 138–40 and esp. 156, which refers to 'the problem about the "now" (218^a8–30) which . . . is a way of raising doubts about the reality of the past'.

section 3 below). In $221^{b}31$ ff., for example, he draws no relevant distinction between entities which are not but have been, e.g. Homer, and entities which are not but will be. Thus even if the *De Interpretatione* doctrine were to play a part in Aristotle's considered answer to Problem I, it is not the answer he gives in the *Physics*.

4. Problem I assumes that to be is to be now, that if *x* is then *x* is, in part at least, now. In Greek, as in English, the tense of a sentence is indicated by a modification of the verb, so that 'is' contrasts with 'was' and 'will be', and in Greek τὰ ὄντα contrasts with τὰ γεγονότα and with τὰ ἐσόμενα or τὰ μέλλοντα. Does the assumption that the temporally remote (unlike the spatially remote) is less than fully real rest on anything more than the contingent fact that Greek and English indicate tense in this way? It probably does, for at least three reasons.

(i) Temporally remote entities are perceptually inaccessible in a way that entities fairly remote in space are not. The temporally remote may be accessible to memory, expectation, etc. but these depend for the most part on some past or present perception in a way in which the perception of something at a distance does not depend on the perception close to either of it or of anything else.

(ii) The spatially remote is physically accessible in a way that the temporally remote is not: I can go over there and then return here; the spatially remote can play an intelligible part in the plans I lay here and now. Future (unlike past) entities can figure in my present projects, but what future entities there are depends in part on what I or we do now, whereas the spatially remote is regarded as fully determinate independently of my/our projected interventions in it: the present occurrence of a battle over there cannot depend on what I now do here.

(iii) Consequently we are tempted to regard the future, at least, as indeterminate, and future entities as not existing at all rather than simply not existing now. The spatially remote presents no similar temptation.

Although these considerations help to explain and justify our association of being with the present, they do not exclude a tenseless use of εἶναι, 'to be' or 'to exist'. In the *Timaeus* Plato marks out a tenseless use of εἶναι for application to atemporal entities such as God. He argues that application of εἶναι to temporal entities is improper (as is the application of 'was' and 'will be' to the ἀίδιος οὐσία) and should be replaced by γίγνεσθαι (*Tim.* 37 e 2 ff.). But Proclus, in his commentary on this passage, acknowledges two legitimate senses of εἶναι, one tenseless and applicable only to genuine, timeless entities, the other tensed and applicable only to temporal entities.[5] Modern philosophers have tended to extend the tenseless 'to be' or 'to exist' to temporal entities. Until recently at least they were prepared to say that Socrates (in contrast to Pickwick) *exists*, though they were less ready to say that dodos (in contrast to unicorns) exist.[6] Perhaps time is among τὰ ὄντα in this tenseless sense, even if it is not *now*.[7] But this is an unsatisfactory solution to Problem I on several counts:

(i) Since Problem I can be extended to any temporal entity (see section 2 above), if we follow Problem I in accepting that time does not exist now, we shall have to accept that nothing exists now, and probably that no present tense verb (except perhaps an achievement verb such as 'win') is ever applicable to anything. (This objection makes the solution unsuitable for us, if not for Aristotle—who, as we have seen, does not acknowledge this extension of Problem I.)

(ii) It is hard to decide whether time itself is temporal or atemporal. On Aristotle's account of being 'in time' (220^b32 ff., 221^b23 ff., 222^b30 ff.), time itself is not in time: there is no time when time is not, and time is not measured or embraced by time. Again, despite his official view that time is a number *counted* (219^b6 ff., 220^b5 ff.), Aristotle often assimilates time to the 'numbers by which we count' (220^b14 ff., 221^a9 ff., 223^b5 ff., 224^a2 ff.), the abstract numbers, 1, 2, 3 etc., which are presumably timeless. On the other hand, time in some sense consists of finite periods of time (e.g. 218^a7, $^b1$–2) and these (whether conceived as the time of a particular motion or more abstractly as, say, July 1989) are in time. It is no doubt artificial to impose on time a distinction between temporality and atemporality which derives its sense from time itself. But the same may be true of the distinction between the tensed and tenseless uses of εἶναι.

(iii) With regard to temporal entities Aristotle rejects the tenseless εἶναι: Homer is not (*Phys.* 221^b32–3, *Int.* 21^a25–6).

(iv) In the case of temporal entities the tenseless use of 'exist' seems

[5] Proclus, *in Tim.* 42, 17–43, 4, esp. 42, 31–3.

[6] Cf. S. R. L. Clark, *Aristotle's Man: Speculations upon Aristotelian Anthropology* (Oxford, 1975), 121 f.

[7] Cf. R. Sorabji, *Time, Creation and the Continuum: Theories in Antiquity and the Early Middle Ages* (London, 1983), 13: 'it is only in the irrelevant sense of being present that the past and future do not exist. In the sense that matters, there is a past and there is a future, and so there is time.'

to presuppose the tensed use of 'exist': Socrates exists (tenselessly) only if it is at some time true of Socrates that he exists (tensedly) now. But, by Problem I, it is *never* true to say that time exists now (see section 1 above). So, if time is a temporal entity, time cannot exist tenselessly. If it is objected that time is an atemporal entity, the reply is this:

A. Since time consists of periods of time, if time exists tenselessly, then periods of time exist tenselessly.
B. Periods of time are temporal entities.
C. By Problem I, no period of time exists tensedly now.
D. Therefore, no period of time exists tenselessly.
E. Therefore, time does not exist tenselessly.

These considerations suggest that Problem I cannot be solved in an unproblematic (or an Aristotelian) fashion by recourse to a tenseless εἶναι.

5. Problem I depends on the assumption that the now that divides past from future is an instant without duration. If the now had some duration, it would be a part of time and 'measure' time (218^a6 ff.)—in so far as any finite period can measure infinite time. There are at least four ways in which the doctrine of the durationless now might be challenged:

(i) We might argue that, independently of our perception of time, there are time atoms, finite periods of time which cannot be further divided.[8]

(ii) At *Phys.* 222^b14–15 Aristotle says that the sudden is what happens in a time too small to be perceived (τὸ ἐν ἀναισθήτῳ χρόνῳ διὰ μικρότητα ἐκστάν). This suggests that there are *periods* of time too short for us to perceive anything happening in them. The extended now might then be a period just large enough to be perceptible.

(iii) We think of even longer periods, lasting for a second or more, as strictly present and as perceived rather than remembered or expected. How else could we perceive change or motion (which does not occur at an instant, *Phys.* 234^a31), as opposed to inferring its occurrence from the change in our perceptions?[9]

[8] Sorabji examines several theories of this type which were proposed in late antiquity (ibid., pt V).

[9] Cf. Thomas Reid, *Essays on the Intellectual Powers of Man* (Glasgow, 1785), III. v: 'though in common language we speak with perfect propriety and truth, when we say,

(iv) Indefinitely long periods may be referred to as 'now', 'the present', etc. if a contrast is drawn with an appropriately remote past or future time: now as opposed to yesterday, last year, the nineteenth century, antiquity, etc. At 222^{a}20 ff. Aristotle acknowledges a similar use of 'now' to refer to events which have happened or will happen today, but he argues that this use of 'now' depends on the proximity of the event to the strict, durationless now.

There are, however, several considerations in favour of a durationless present. Some were presented by Aristotle and others may have influenced him implicitly:

(*a*) The expression τὸ νῦν means both 'instant' and 'the present'. An instant is instantaneous, the temporal analogue of a point. But this does not guarantee that the present is instantaneous. For first, position (i) above denies, and position (ii) might deny, the existence of durationless instants. And second, even if we grant the existence of instants, we might still argue that the present is a period of time, containing an infinity of instants none of which is more strictly present than the others. We would then be less inclined to call an instant a 'now'.

(*b*) Pressure to restrict the scope of 'now' derives from the fact that relevant changes may occur within what is thought of as the present. I cannot say, for example, 'Earlier (from Oxford to Reading) the train was doing 80 m.p.h., but now (from Reading to London) it is doing 100 m.p.h.', unless the train is moving at approximately 100 m.p.h. at the time of my utterance—even if its *average* speed between Reading and its present location is 100 m.p.h. Similarly I cannot claim that entities exist now solely for the reason that they existed at an earlier phase of what I currently regard as the present: I can say, for example, 'Elephants did not exist in 10 million BC but they do now', but not 'Dodos did not exist in 10 million BC but they do now'. If the relevant state of affairs is constantly changing, precision requires a shorter rather than a longer now, and the limit of such successive reductions is the durationless now. Position (iv) above will allow these successive

that we see a body move, and that motion is an object of sense, yet when as philosophers we distinguish accurately the province of sense from that of memory, we can no more see what is past, though but a moment ago, than we can remember what is present: so that speaking philosophically, it is only by the aid of memory that we discern motion, or any succession whatsoever. We see the present place of the body; we remember the successive advance it made to that place. The first can then only give us a conception of motion, when joined to the last.'

reductions, conceding that some part of what is currently treated as the present is more strictly present than the rest of it. But positions (i), (ii), and perhaps (iii) will insist that there is a minimum period beyond which the present cannot be further reduced in length. (Relevant variations within a spatial area may similarly lead us to restrict the scope of 'here'—the spatial counterpart of 'now'—but, owing perhaps to the size of our bodies and the spatial range of our senses, we are not tempted to reduce the scope of 'here' to a point; when I do refer to some spot as 'here' it may be at some distance from my body.)

(*c*) The belief that what is merely past or merely future (in contrast to what is *there* rather than *here*) does not strictly exist exerts pressure both to narrow the scope of the present and to sharpen its boundaries, on the assumption that the boundary between existence and non-existence cannot be hazy, shifting, or arbitrary. Positions (i) and (ii) may resist this pressure to reduce the present to an instant, but (iii) and (iv) can supply no single, obviously correct answer to the question 'If the present is not an instant, how long is it?' The indeterminacy of the future might be thought to require an especially sharp and non-arbitrary boundary between the present and the future. But this is not so, for in general the plausibility of regarding future-tense statements as indeterminate increases with the remoteness of the future to which they refer: 'That vase will break tomorrow' may be indeterminate in truth-value, but 'That vase will hit the floor in two seconds' is not, if the vase is now falling and no one is near enough to catch it. Nevertheless the doctrine of the instantaneous present is not wholly independent of the association of *being* with the present.

(*d*) Aristotle attacks position (ii) at *Sens.* 7, 448^{a}19 ff.,[10] arguing that if there is a time too small to be perceived, two unacceptable consequences follow. (1) A person will be unaware, during such a time, both of his own existence, and of the fact that he is perceiving (448^{a}26–30). (2) It will never be strictly true that a person perceives, say, a line for the whole of a period *t*1–*t*10 (or indeed the whole of the line, since a similar argument will apply to the perception of a line as to the perception of time). For if, e.g., *t*9–*t*10 is an imperceptible time, the line will not be perceived during *t*9–*t*10, so that the line is perceived only during *t*1–*t*9. But the same argument will apply to the claim that I perceive the line through the whole period *t*1–*t*9: since *ex hypothesi* *t*8–*t*9 is imperceptible, the line is perceived not through the whole of

[10] This apparent contradiction of *Phys.* 222^{b}14–15 puzzled Alexander of Aphrodisias (Simplicius, *in Phys.* 753. 9 ff. *ad* iv. 13, 222^{b}14).

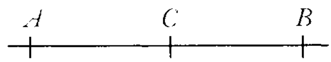
*t*1–*t*9, but only from *t*1–*t*8. We are meant to infer, though Aristotle does not explicitly conclude, that by successively removing each imperceptible segment of the period *t*1–*t*10 we can show that the line is not perceived at all, since on the hypothesis under consideration any finite time will consist of a finite number of imperceptible times (448ᵃ30 ff.). His argument for consequence (1) requires the assumption that time in fact, if not as perceived, is 'continuous' (*ἐν συνεχεῖ χρόνῳ*, 448ᵃ27; *ἐν τῷ συνεχεῖ*, 448ᵃ28), that is, that any time is divisible into shorter times. For if time itself were divisible only in the way that, on the hypothesis under attack, our perception of time is, Aristotle could not infer from it that there is a time in which we are unaware of our own existence. Both arguments seem to ignore the fact that two or more imperceptible times might together constitute a perceptible time. If I perceive myself or anything else for a period *t*1–*t*10, consisting of imperceptible times *t*1–*t*2, etc. we cannot say of any one these imperceptible times, as opposed to the others, that nothing is perceived in it, since each one, when added to its immediate neighbour, will constitute a perceptible time; nor can we successively remove imperceptible times without effect, any more than we can successively remove the hairs on a head without eventually producing baldness. (Conceivably, what underlies Aristotle's arguments here is the feeling that if I am aware of a movement, lasting e.g. two seconds, as present, my awareness of the movement cannot itself take two seconds or unfold sequentially as the movement does. How, in that case, could I at any time be aware of the movement as a whole? Such considerations may have suggested to him a now or present of my awareness, which, if not instantaneous, is narrower than the present of which I am aware.)

(*e*) Aristotle's main argument for the view that the now is primarily an indivisible instant, and that it is by reference to this now that extended periods are said (*καθ' ἕτερον*) to be now occurs at *Phys.* VI. 3, 233ᵇ33 ff. There is a boundary (*ἔσχατόν τι*) of the past and a boundary of the future. Within the boundary of the past there is none of the future, and within the boundary of the future there is none of the past. The boundary of the past and the boundary of the future must be one and the same now or present. For if they were distinct nows, the past and the future would not be continuous with each other. But if the now is an extended and therefore divisible period, three impossible consequences result:

(1) If the now is divisible, some of the past will be in the future and

some of the future will be in the past. For that at which the now is divided will demarcate past and future time. (2) The now will not be καθ' αὑτό but καθ' ἕτερον, since the division is not καθ' αὑτό. (3) A part of the now will be past and a part will be future, and it will not always be the same part that is past or future. Thus the now will not be the same. For the time can be divided in many ways ($234^{a}11$–19).

Suppose that the past ends at instant *A*, while the future begins at instant *B*. Then *ex hypothesi*, *AB*, the now, is an extended period, divisible at any instant *C*:



AC and *CB* are again extended periods. Aristotle's arguments can then be interpreted as follows: (1) *AC* is earlier than *CB* and is past with respect to *CB*. Conversely *CB* is later than *AC* and is future with respect to *AC*. But, on the other hand, *AC* is later than *A*, and is therefore future with respect to *A* and what precedes *A* (the past) and, conversely, *CB* is earlier than *B* and therefore past with respect to *B* and what succeeds *B* (the future). So *AC* is both past (with respect to *CB*) and future (with respect to *A*), while *CB* is both future (with respect to *AC*) and past (with respect to *B*). (2) The extended now *AB* will not be a now καθ' αὑτό, but only καθ' ἕτερον. For there is nothing intrinsic to *A* that makes it, rather than some earlier or later instant, the appropriate boundary of the present and the past, and nothing intrinsic to *B* that makes it the obvious boundary of the present and the future. What gives them this status is the fact that the period *AB* includes an instant *C* which is the now καθ' αὑτό and does not owe this status to anything else. (3) *AB* need not be divided only at *C*. Since it is continuous it may be divided at any of an infinity of points, C^1, C^2, C^3, etc. Suppose that we divide it at C^1. Then AC^1 will be past and C^1B will be future. If we divide it again at some later point C^5, then AC^5 will be past and C^5B will be future. But the period C^1C^5 will be future on the first division and past on the second. Similarly all of the infinitely many different divisions that we could make will produce a different past and a different future within *AB*. Depending on how it is divided then the now will be different, unlike the punctual now which, since it cannot be divided, is always the same.

If these interpretations are correct, none of the arguments is compelling: Argument (1) confuses the notion of past and future *simpliciter* with the notion of earlier and later (than) or of past and future with respect to something. The battle of Waterloo is/was later than the battle of Hastings and is/was future with respect to the battle of Hastings, but it is past *simpliciter*. Similarly, *AC* is earlier than *CB* and is past with respect to it; it is later than *A* and future with respect to *A*. But it does not follow either that *AC* is past or that it is future *simpliciter*. On the hypothesis under consideration *AC* is simply present. Conversely, *CB*, though future with respect to *AC* and past with respect to *B*, need be neither past nor future *simpliciter*, but present. (We might also question Aristotle's assumption that the boundaries between the past and the present (*A*) and between the present and the future (*B*) are punctual instants rather than hazy boundaries, but this assumption is not essential to his argument here.)

Argument (2) seems to beg the question by assuming that since the extended now can be divided at any of infinitely many points, there is some point *C* which is strictly and primarily the present and that it is in virtue of including *C* that *AB* is regarded as secondarily and loosely the present. Aristotle may be troubled by two questions about the extended present: Why is the present *AB* located now (e.g. from 6.05.15 to 6.05.18 p.m.) rather than earlier or later (e.g. from 6.05.19 to 6.05.22.)? Why is the extended now three seconds long rather five or two seconds? But it seems unjustified to assume that the first question awaits some such answer as, 'Because it includes the punctual now, 6.05.17 p.m. precisely', while the question, 'Why do you locate the punctual now at 6.05.17 p.m. rather than earlier or later?' requires no answer. The second question is more difficult: the extended now differs from the punctual now, for there is no analogous problem 'Why does the punctual now last for n, rather than $n + m$ or $n - m$ seconds?' The question cannot be answered by invoking the haziness of *A* and *B*, the boundaries of the extended present, since hazily bounded periods may nevertheless be longer or shorter. Nor can it be resolved by referring to the extended now's inclusion of the punctual present, since a longer or a shorter extended present will include the punctual present (if we allow, with Aristotle, that there is one). Aristotle might still claim that the boundaries of the extended present are located as they are not in virtue of their intrinsic properties, but καθ' ἕτερον, in virtue perhaps of our perceptual and retentive capacities or in virtue of our current interest in, for example, contrasting the weather now with the weather last week (or last year). But at all events the location of the boundaries of the hypothesized extended present seems not to require a punctual present included within them.

Argument (3) also fails: an extended present, of a given length and location, will not be homogeneous in the way that a punctual present must be. But that need no more matter than the fact that the past and the future are not homogeneous in that way. That different segments of the extended now may be alternatively regarded as past and future with respect to each other will not entail that these segments are past and future *simpliciter* unless, contrary to the hypothesis under consideration, the different points of division are each covertly regarded as the genuine (punctual) present. (Conceivably, Aristotle here exploits the point that successive extended presents will be overlapping, rather than consecutive like railway carriages. But it is hard to see how this will improve his argument.)

Whatever the merits of the extended present, Aristotle clearly rejected it and we must look elsewhere for his solution to Problem I.

II Problem II

1. Problem II asks when, if at all, the now perishes. Corresponding to the double sense of *τὸ νῦν* this question can be formulated in two ways: When, if ever, does it cease to be a given instant, for example, 6 o'clock precisely? When, if ever, does it cease to be now? These are not necessarily the same question: it is more tempting to suppose that it never ceases to be now than that it never ceases to be 6 o'clock (see section 2 below). But it is initially plausible that the core of Problem II is adequately represented by the first question.

Problem II rejects two possible answers: (i) it stops being 6 o'clock at 6 o'clock; and (ii) it stops being 6 o'clock at some time after 6 o'clock. Sorabji argues for a third alternative: (iii) there is no time at which it ceases to be 6 o'clock, but at any time after 6 o'clock it *has ceased* to be 6 o'clock.[11] He compares the claim that someone has aged, when there is no time at which he or she aged. The comparison is not compelling, since aging is usually a gradual process, while its ceasing to be 6 o'clock is not.[12] It might be argued that solution (iii) does not differ significantly from solution (i): to say that at any time after 6 o'clock, however brief the interval, it has ceased to be 6 o'clock suggests that 6 o'clock is the limit of a series of progressively shorter intervals, and thus looks like an elaborate way of saying that it ceased to be 6 o'clock

[11] Sorabji, *Time, Creation*, 10.
[12] Ibid., 12.

at 6 o'clock. But there is, in any case, more to be said for (i)—that, despite the fact that at 6 o'clock it *is* 6 o'clock, it stops being 6 o'clock at 6 o'clock (and by parity of reasoning also starts being 6 o'clock)—than Aristotle and Sorabji allow. It is true that we cannot say, 'It is/was ceasing to be 6 o'clock (at 6 o'clock)',[13] since this would suggest that it takes time to stop being 6 o'clock; but the non-continuous present and past tenses ('It ceases, ceased to be 6 o'clock at 6 o'clock') are not similarly objectionable. Normally events and times are not said to start, occur and stop at the same time (except loosely, as in 'No sooner had he started than he stopped again'). But this is because they do not normally start, occur and cease simultaneously rather than because their doing so would be intrinsically contradictory: 'It stopped being 6 o'clock at 6 o'clock' does not obviously entail 'It was not 6 o'clock at 6 o'clock'. Nor is it obviously odder to suppose that it stops being 6 o'clock at 6 o'clock than that it has stopped being 6 o'clock even though there is no time at or during which it did so.

2. Aristotle adopts neither (i), (ii) nor (iii) as his answer to Problem II. His answer is given, if anywhere, at 219^b9 ff.:

And as the motion is always other and other, so too is the time. But all simultaneous time is the same [ὁ δ' ἅμα πᾶς χρόνος ὁ αὐτός]. For the now is the same whatever-it-may-be [ὅ ποτ' ἦν], but its being [τὸ δ' εἶναι] is different. The now measures[14] time *qua* before and after. The now is in one way the same, and in another way not the same. For in so far as it is in successively different states [ἐν ἄλλῳ καὶ ἄλλῳ], it is different—that is what it is to be now—but in respect of whatever-it-may-be [ὅ δέ ποτε ὂν ἐστι τὸ νῦν] it is the same. For, as was said earlier, motion follows magnitude, and, we claim, time follows motion. Similarly, then, the moving object, by which we recognize the motion and the before and the after in it, follows the point. The moving object is the same whatever-it-may-be [ὅ μέν ποτε ὄν] (for the point is either a stone or something like that,[15]) but different in account [τῷ λόγῳ], just as the sophists take being Coriscus in the Lyceum to be different from being Coriscus in the market-place. The moving body, then, is different by being in different places. But the now follows the body, as time follows motion. For it is by the moving object that we recognize the before and after in motion, but the now is the before and after *qua* countable. So that in these too [viz. in the before and

[13] Ibid., 10.

[14] Retaining the MSS reading μετρεῖ.

[15] I follow Owen in reading ἡ στιγμή at 219^b19 and in taking Aristotle to say that the point on his diagram represents, say, a stone. See G. E. L. Owen, 'Aristotle on Time', in P. K. Machamer and R. G. Turnbull (eds.), *Motion and Time, Space and Matter: Interrelations in the History of Philosophy and Science* (Columbus, Ohio, 1976) 22 and n. 42.

after[16]] in respect of whatever-may-be-now [*ὅ μέν ποτε ὂν νῦν ἐστι*], it is the same (for it is the before and after in motion), but its being is different (for the now is the before and after *qua* countable). This is also especially knowable. For motion is known through the moving object and locomotion through the locally moving object. For the locally moving object is a 'this' [*τόδε τι*], while motion is not. In one way then the now is the same always, but in another way it is not the same. For the same is true of the moving body.

The now remains the same over time in the way that Coriscus does, but differs in so far as it coincides with different stages of the movement of a body; the now is the same in so far as it is always simply a now, but differs in so far as it may be specified as, e.g. 6 o'clock. The persistence of the now in this sense will not entail that there cannot be periods of time bounded at either end or that remote events are simultaneous with each other. For a period of time will be bounded by the now under different descriptions (or differing in being, 219^{b}11, 14, 17) and non-simultaneous events occur at, or are bounded by, the now differently described.[17] This solution to Problem II does not however answer the question, 'When does it cease to be 6 o'clock?'.

3. An initial difficulty concerning this doctrine of the persisting now is that the claims that time 'follows' motion, and that all motion is the motion of a persisting object, do not entail that the now persists in the way that the object does. If we press the analogy between the now(s) and the object, we could reasonably say that the nows are the counterparts of successive instantaneous states of the object, while the now regarded as the present corresponds to the present state of the object. But why need we suppose that there is some persistent temporal entity which underlies or consists of the successive nows in the way that the persisting object underlies or consists of its successive states? And if there must be such an entity, why might it not simply be time itself? Aristotle insists that the counterpart of time is not the moving object, but the motion or change; what corresponds to the object is the now. But this seems arbitrary. Considered simply as an object, an object is quite different from the now and from time itself, since it may remain

[16] W. D. Ross, *Aristotle's Physics: A Revised Text with Introduction and Commentary* (Oxford, 1936), 601 *ad* 219^{b}26–8 takes *ἐν τούτοις* to refer to the nows. But a reference to nows (in the plural) would be quite out of place here. I take 'these' to refer to the before and after, and I interpret Aristotle as distinguishing the identity-in-difference of the now over time from its identity-in-difference over space at any given time—the question raised earlier at 219^{b}10 ff. See Pt II, section 4.

[17] So Philoponus *in Phys.* 725,16–728,14 *ad* IV. 11, 219^{b}12.

the same object without changing or moving, at least in the relevant way; while the object *qua* moving (if we grant Aristotle that any change or movement must be that of a single object) differs only verbally from the movement itself. Aristotle's neglect of the distinctions between the object as such, the object *qua* undergoing this particular movement, instantaneous states of the moving object, and the present state of the moving object, is also apparent in the further parallels which he draws between the object and the now:

(i) Aristotle draws an epistemological parallel between the moving object and the now: We recognize the before and after in motion by the moving object (219^b23–5). The object is most knowable: for we know the motion by the moving object, because the moving object is a τόδε τι, but the motion is not. Since the now is the before and after *qua* countable,[18] corresponding claims are implied for the now: we recognize the before and after in time (which does not differ significantly from the before and after in motion) by the now (cf. 219^b25). The now is more knowable than time, and time is known by the now, since the now, though not a τόδε τι, is a persistent entity in a way in which time as such is not. This epistemological parallel depends, however, not on an analogy between the now and the object, but between nows and instantaneous states of the moving object: the before and after in motion are presumably picked out or recognized by earlier and later states of the object; and if the whole motion is recognized as the motion of such-and-such an object or as the motion proceeding from a certain initial state of this object to a certain final state, the stretch of time it takes is recognized not simply by reference to a single undifferentiated now, but as the time which lasts 'from *now* to *now*', or 'from 6 o'clock to five past six'. (Cf. 219^a27 ff., where the passage of time is marked 'when . . . the soul says that the nows are two, one earlier and the other later'.)

(ii) At 219^b33 ff. he says that just as there can be no motion (φορά) without the moving object (τὸ φερόμενον) and no moving body without the motion, so there can be no now without time and no time without the now—the now is the number of the moving body and time the number of the motion.

(iii) At 220^a4 ff. Aristotle says, 'Time is both continuous by the now and divided at the now. For this too follows the motion and the

[18] 219^b25; and 220^a3 f.: time is the number of movement, but the now corresponds to the moving object, like the unit of number.

moving object. The change and the motion is one in virtue of the moving object, because it [the object] is one, not [only] whatever-it-may-be—for there could be a gap—but in definition.' It is unclear whether, in contrasting ὅ ποτε ὄν with τῷ λόγῳ he is contrasting the object as such with the object as moving or, rather, the object as moving with some particular state of the object. Probably he has both points in mind: If object *x* moves from *A* to *B*, and just as *x* reaches *B*, a distinct object *y* (whether or not impelled by *x*) moves from *B* to *C*, the movements from *A* to *B* and from *B* to *C* do not together constitute one movement, but two. If one object *x* moves from *A* to *B* and then from *B* to *C*, the movements from *A* to *B* and from *B* to *C* constitute one movement only if *x* remained in motion and did not stop at *B* and then restart, if, that is, its state of arrival at *B* is the same as its state of departure from *B*. Since Aristotle again stresses that the now follows the motion and the moving object, he implies that to the question 'What is required for two stretches of time to constitute a single stretch of time (rather than two temporally separated or even temporally unrelated stretches of time)?' the answer is that there is a single persisting now analogous to the persisting object. However he goes on to discuss an apparently quite different doctrine (see section 5 below).

In suggesting, though, that the now is analogous to an instantaneous state of the moving object rather than to the object as a whole, we are not only offending against Aristotle's reluctance to carve up objects along temporal lines (see Part I, section 2), but also against the spirit of the doctrine of the persisting now. For the doctrine seems intended to account for those features of our experience—such as our ability to perceive movement and the smoothly continuous rather than jerky character of our temporal experience—which have inclined some philosophers to postulate an extended present (see Part I, section 5). It perhaps represents a compromise between Aristotle's austerely mathematical conception of time and the phenomenology of our temporal experience, to which he often appeals (e.g. 218^{b}21 ff., 219^{a}22 ff., 223^{a}16 ff.). But the more precise significance of the doctrine is unclear. Is Aristotle drawing, as Hussey believes,[19] a distinction between passing *instants* and the persisting *present*? Or is he simply claiming that the now, the undifferentiated instant/present, may be

[19] Hussey, *Physics III and IV*, p. xliv: 'while we must distinguish between the fixed, unrepeatable instant and anything that persists in time, still there *is* something that persists in time, and, in some sense, *is* each instant successively: something that is persistently 'in the present' and divides past from future'.

regarded (like Coriscus) either as successively different or as persistent, depending on how much accompanying detail we embody in our reference to it? And in either case, is he affirming a potently metaphysical doctrine of an eternal now (such as we find in, for example, Tatian[20])? Or is he only making some such verbal point as that we may always refer to the present simply as 'now', with the additional suggestion perhaps that the now has insufficient ontological depth for the criteria of its individuation to differ from those of the object which it follows? To answer these questions we need to consider two other doctrines with which the doctrine of the persistent now is deeply entangled in Aristotle's text.

4. The first of these is the postulate of simultaneity, which Aristotle variously expresses as, 'Time is equally both everywhere [πανταχοῦ] and with everything' (218^b13); 'Every simultaneous time is the same' (219^b10); 'The same time is everywhere simultaneously' (220^b5–6); and 'Time is everywhere the same' (223^b10–11). Aristotle's definition of time as 'the number of motion with respect to the before and after' (219^b2), together with his beliefs that 'the change and motion of anything is in the very thing that changes only, or wherever the moving and changing object happens to be' (218^b10 ff.) and that time is the number of motion in the sense of a number that is counted, not the number by which we count (219^b5 ff.), and is therefore *κινήσεώς τι πάθος ἢ ἕξις* (223^a18–19), make it hard to see how distinct changes can be simultaneous, and more generally how all changes can be located in a single time, in which they are before, after, overlapping, or simultaneous with each other rather than each having its own time. The comparison of time to the number by which we count (esp. 223^b1 ff. and possibly 220^a21 ff.)[21] does not help much here. It may well enable us to say whether or not two different motions are of equal duration, as

[20] Cf. Tatian, *Or. Graec.* 27, 22–8 in S. Sambursky and S. Pines, *The Concept of Time in Neoplatonism: Texts with Translation, Introduction and Notes* (Jerusalem, 1971) 103: 'Why do you divide time, saying that a portion of it is the past, a portion the present, and a portion the future? For how can the future pass by, if the present is there? Just as passengers believe, out of ignorance, as the ship is carried on, that the mountains are going [past], so are you unaware that you are going past, whereas eternity is [and will be] static for as long as its Maker wills it to exist.'

[21] The comparison is made elsewhere for different reasons: at 220^b4–5 time, like the numbers by which we count, cannot be fast or slow; at 221^a13 ff. finite things and events are in time in the way that finite collections are 'in number'; at 220^b14 ff. the measurement of time by motion is like measuring 'the number by what is counted, e.g. the number of the horses by the single horse' (220^b19–20); at 221^b10 ff. time measures

it enables us to say whether two collections of objects are equal in number. But the fact that one and the same time can be used to measure the duration of different motions need not enable us to say whether the motions are simultaneous, overlapping, or one wholly before the other.[22] For that, we would need not only to say how many temporal units each motion occupied, but to assign each motion to some definite segment of the number series, one of them to, say, 1832–9, the other to 1842–9—which we do not need to do when we simply count objects. Nor does the introduction of a single, dominant heavenly motion (223^b12 ff.) remove the difficulty: we still need to know how to regard a certain terrestrial motion as simultaneous with some portion of the heavenly motion.

Phys. 219^b9 ff. starts with a discussion of the postulate of simultaneity ($^b10$: 'all simultaneous time is the same')[23] and concludes with a discussion of the doctrine of the persisting now, giving no clear indication of where the one discussion ends and the other begins.[24] It is, however, hard to see a close relationship between the two doctrines. At 220^b5 ff. Aristotle reaffirms that time is the same everywhere, but insists that, before and after, time, and the nows, are different, with no mention of the persisting now. One connection between the two doctrines is that both can be expressed by saying that the now as whatever-it-may-be is the same, while its being is different: if Coriscus leaves Athens at the same time as Diares leaves Corinth, the same now as whatever-it-may-be is the now both of Coriscus' departure and of Diares' (cf. 219^b10–11);[25] and the now of Coriscus' departure is the

periods of rest as well as motion, since 'it is possible for what is at rest too to be in the number of motion' (221^b11–12); and 224^a2 ff. discusses the matter more generally.

[22] 223^b2 ff. conflates the question whether the durations of two motions can be compared and measured on the same time-scale with the question whether they are *ἅμα*, without explaining how they can be *ἅμα*. Possible sources of confusion are (i) Aristotle's attempt to make the comparison of time to abstract numbers answer both questions; (ii) his assumption at 223^b2–3 that if there *were* two distinct times they would be *ἅμα* (ambiguously 'together' and 'simultaneous') rather than temporally disconnected; and (iii) the fact that the measurement and comparison of two motions presupposes that each motion is (judged to be) simultaneous, if not with the other, at least with the motion of a 'clock'.

[23] Hussey's translation—'though the whole time in sum is the same' (*Physics III and IV*, 44)—obscures the presence of the postulate of simultaneity.

[24] But see n. 16 above.

[25] The note to 219^b10 f. in R. P. Hardie and R. K. Gaye, *Physica*, in W. D. Ross (ed.), *The Works of Aristotle Translated into English*, ii (Oxford, 1930) is apt: 'e.g. if you come in when I go out, the time of your coming in is in fact the time of my going out, though for it to be the one and to be the other are different things.'

same whatever-it-may-be as the now of his arrival at Corinth (cf. 219^{b}26 ff.). But this does not entail that either doctrine implies the other. One connection, which Aristotle may have in mind, is this:

(i) If *x* goes from Athens to Corinth via Megara, then (by the doctrine of the persisting now) the now of his departure from Athens is the same as the now of his arrival at Corinth, and the same as the now of every other point on his journey.

(ii) Therefore it is the same now in Athens, Corinth, Megara, and, by a natural extension of the argument, anywhere else that *x* does or could visit.

(iii) Therefore time is the same in Athens, Corinth, and indeed 'everywhere' (218^{b}13, 220^{b}5–6, 223^{b}10–11).

(iv) Therefore different motions are related to each other as simultaneous, earlier, or later. If for example *y* goes from Megara to Athens, and *z* goes from Megara to Corinth, we can say that they reached their destinations simultaneously or at the same now, in virtue of the identity of time and the now secured by *x*'s actual (or potential) journey through all three places.

This argument has at least two flaws: (*a*) The doctrine of the persisting now leaves the now attached to a particular motion. It does not licence an inference from (i), that the now of *x*'s departure from Athens is the same as the now of his arrival at Corinth, to (ii), that the now is the same in Athens and Corinth *simpliciter*. Of itself, therefore, it does not entitle us to relate temporally two or more distinct motions, arrivals and departures as simultaneous, earlier, or later. (*b*) The argument blurs rather than clarifies the distinction between the sense in which the now is the same over time, throughout e.g. the whole of *x*'s journey (as in (i)), and the sense in which the now of e.g. *y*'s arrival at Athens is the same as the now of *z*'s (simultaneous) arrival at Corinth (as in (iv)).

This argument fails to establish a connection between the persisting now and the simultaneity postulate, but it is sufficiently plausible to have figured among Aristotle's implicit motives for advancing the doctrine of the persisting now. It tells us little about the significance of this doctrine, however. It does not settle whether what persists is the present or simply the undifferentiated now. For the simultaneity postulate applies with respect to both 'senses' of 'the now': the present moment is the same in Athens and Corinth, or events in Corinth, simultaneous with what is happening now in Athens, are also happening now; but events in Athens may be simultaneous with events in Corinth, regardless of whether they are past, present, or future.

5. The second doctrine with which the persisting now seems to be intertwined is that of continuity, the doctrine that the now both makes time continuous and (potentially) divides it. The doctrine of continuity is introduced at 220^{a}4 ff.: 'Furthermore time is both continuous by the now and divided at the now. For this too follows motion and the moving object. For the change and the motion are one by the moving object, since it is one.' The doctrine is thus associated with the moving object, the counterpart of the persisting now. The association of the two doctrines is more explicit at 222^{a}10 ff.:

> The now is the link [συνέχεια] of time, as we said. For it links past and future time, and is a limit [πέρας] of time. For it is the beginning of one and the end of another. But this is not as clear as in the case of the stationary point. It divides potentially. And as such, the now is always different, but in so far as it binds together, it is always the same, as in the case of mathematical lines. For it is not always the same point for the intellect—when we divide, it is other and other. But in so far as it [more probably the line than the point] is one, it [more probably the point than the line] is the same everywhere [πάντῃ]. So too the now is both a division of time in potentiality, and a limit and unity of both. The division and the unification are the same thing and in respect of the same thing, but their being is not the same.

This passage invites several questions:

(i) Does it in fact contain the doctrine of the persisting now? The claim that 'in so far as [the now] binds together, it is always the same [ἀεὶ τὸ αὐτό, 222^{a}15]' might be taken to mean either (*a*) that the now is always the same over time—for example, the now at 6 o'clock is the same as the now at 7 o'clock—or (*b*) that any given now regarded as the end of a period is the same as itself regarded as the beginning of another period, for example, the now of 6 o'clock regarded as the end of the preceding period is the same as the now of 6 o'clock regarded as the start of the following period. (*a*) is the doctrine of the persisting now. It is not obvious that (*a*) is entailed by (*b*), but (*a*) may entail (*b*), since the continuity of time is presumably a necessary, if not sufficient, condition of the persistence of the now. Both (*a*) and (*b*) may be expressed by saying that the now is the same whatever-it-may-be, but different in being. (Cf. 222^{a}33–^{b}4 where, without a hint of the *persisting* now, the now as the end of the past and the beginning of the future is compared to the circle which is both convex and concave.) It is not certain, then, that 222^{a}10 ff. involves the doctrine of the persistent now, at least in the undiluted form in which it is distinct from the doctrine of continuity.

(ii) How far, on Aristotle's view, does the parallel extend between time and magnitude or the line, and between the now and the point? Time follows motion, which in turn follows magnitude (219^{a}10 ff., 219^{b}15 ff.), and the now follows the moving object, which in turn follows the point (219^{b}16 ff., 220^{a}6 ff.) Prima facie, however, there are at least two, but probably connected, differences between magnitude (primarily a line) and time: first, time (and a change or motion) have a direction, but a line as such does not. The line *AB*:

A *C* *D* *B*

does not essentially run from *A* to *B* or from *B* to *A*. The points *A*, *C*, *D*, and *B* are ordered (*C* is between *A* and *D*, etc.), but no one is before or after any other, unless we import direction into *AB* by seeing it as the route of a movement from *A* to *B* or from *B* to *A*.[26] Second, although points on the line may be seen as the counterparts of instants, the line as such provides no analogue of the *present* instant, or of the past and future, unless again we import them by regarding the line as the route of some actual or potential motion.[27] Aristotle seems to deny the first difference at 219^{a}14–15; τὸ δὴ πρότερον καὶ ὕστερον ἐν τόπῳ πρῶτόν ἐστιν. ἐνταῦθα μὲν δὴ τῇ θέσει. But this claim might be interpreted in at least three ways: (1) The positions on a line have a

[26] For the difficulties felt in antiquity about Aristotle's apparent denial of this at 219^{a}14 f., see Galen in Themistius, *in Phys.* 149,4 ff. (321, 16 ff.) and Simplicius, *in Phys.* 718, 13–719, 21; Plotinus III. 7, 9, 55 ff.; Simplicius, *in Phys.* 712, 7–719, 21; Philoponus, *in Phys.* 719, 10–720, 18 and 727, 3–7.

[27] J. McT. E. McTaggart, 'The Unreality of Time' in *Mind* 17 (1908), 457–74, argues that a 'series which is not temporal has no direction of its own, though it has an order' (p. 462). For a series of events which has order but no direction (a 'C series') to become a temporal series with a direction from earlier to later (a 'B series'), it is necessary (and sufficient) that the events should be successively past, present, and future, i.e. constitute an 'A series' (pp. 463 ff.). Hussey suggests that the 'Aristotelian theory of change ... produces a natural "time direction" (since Aristotelian changes cannot be read backwards, and since closed systems always degenerate)' (*Physics III and IV*, 169 *ad* 221^{a}26 ff.). But firstly, the theory of change and degeneration plays only an incidental role in *Phys.* IV. 10–14 (at 221^{a}30 ff. and 222^{b}16 ff.), and secondly, McTaggart would reply that though it may be natural for us to read the series of changes in one direction rather than the other (just as it is natural to count forward rather than backward, 'The Unreality of Time', 462), there is nothing in the series itself that compels us to do so, unless the series constitutes an *A* series.

certain order owing to their arrangement [τῇ θέσει], though none is *strictly* before or after the others. (2) Spatial magnitudes have a certain direction, as say, in the case of race-tracks;[28] but this direction is assigned to them by convention [τῇ θέσει].[29] (3) Hussey argues for an interpretation on which the before and after is intrinsic to magnitude as such:

['The before and after in place'] ought to be something that is present at every point along a path of change, independently of any change's actually occurring. The best solution is that what is meant is the 'ubiquitous point', representing a potential division, that is present anywhere and everywhere on an undivided line.[30]

Hussey's interpretation is unsatisfactory on several counts. First, it exploits, though it does not strictly require, an improbable reading of τὸ πρότερον καὶ ὕστερον. This, Hussey argues, 'ought to mean "that which is both before and after" ... i.e. it ought to pick out something (or *the* thing) that is, successively or simultaneously, present at all stages of the ordered series' (p. 147)—the ubiquitous point in place, the 'leading edge' of change (p. 149), and the 'persisting present' (pp. 152 ff.). This reading is implausible both because 'the before and after' is sometimes treated as plural[31] and because it seems unlikely that Aristotle would introduce a controversial and novel doctrine with this innocent-looking expression. More probably, the expression does not differ significantly from 'the before and *the* after' and means something like 'priority and posteriority', its point being to indicate an

[28] Plotinus III. 7. 9. 63 and Simplicius, *in Phys.* 712. 11–13 both refer to the stadium or race-track.

[29] Hussey translates τῇ θέσει as 'by convention' (*Physics III and IV*, 43 and 147). But it is hard to see how the claim that the before and after are in place *by convention* is to be reconciled with the claim that they are in it *primarily*.

[30] Ibid. 149. Cf. 153 *ad* 219^{b}16–20: 'a point of indeterminate location ... the underlying *x* of which all particular (located) points are only different aspects ... Thus a particular point at location *L* is the ubiquitous point as realised at *L*.' And p. 171 *ad* 222^{a}10 ff.: 'Here, and at *Gen. et Corr.* I. 2, 317^{a}2–17, Aristotle is a little more explicit [than at 219^{a}14 ff. and 219^{b}9 ff.]. The "ubiquitous point" is an unlocalised potentiality for division.'

[31] 219^{a}23–6: 'And we say that time has occurred when we have a perception of the before and after in motion. We determine them by taking them (αὐτά) to be other and other'. Ross takes αὐτά to refer to the nows (Ross, *Aristotle's Physics*, 386), but the nows are not mentioned in this passage until 219^{a}28, and the only possible referent is the before and after. Hussey says that '"them" denotes "the before" and "the after" considered as two' (Hussey, *Physics III and IV*, 150). This argues against the view that the force of the expression is to indicate one thing that is both before and after. Cf. 219^{b}26 and n. 16 above, and also 223^{b}5 ff. and n. 36 below.

obvious feature of time, change, and, supposedly, magnitude, rather than to commit us to the ubiquitous point or the persisting now. Second, none of the passages cited by Hussey unequivocally imply the doctrine of the ubiquitous point. For example, 222ª17: 'in so far as it is one, it is the same πάντῃ'—may mean that if a line *AB* is divided at any point *C*, *C* regarded as the end of *AC* is the same as *C* regarded as the beginning of *CB*. Third, the doctrine of the ubiquitous point does not secure the required parallel with the persisting now, since it does not entail that a line has an intrinsic direction or, therefore, that the point is *before* and *after*. For this, we require in addition some such doctrine as that a line is essentially the product of the movement or 'flow' of a point: the line *AB* is, or results from, the flow of a point from *A* to *B*, and not from *B* to *A*, so that *A* is intrinsically before *C*, etc. This doctrine is to be distinguished from such claims as: (1) A line *AB* may be *seen as* the product of the flow of a point from *A* to *B* (or alternatively from *B* to *A*);[32] (2) If a point moves from *A* to *B*, it generates a line *AB*;[33] (3) If we draw a line *AB*, we do so by moving the tip of our charcoal from *A* to *B*.[34] None of these claims entail that a line *AB* as such, independently of our view of it, the process of producing it, or the motion of an object along it, is directed from *A* to *B* rather than from *B* to *A*. Aristotle seems to disown the doctrine of the moving point, in any sense which could entail the intrinsic directedness of a line, at 220ª12–14, 17–18 and, more explicitly, 222ª13 (ἐπὶ τῆς στιγμῆς μενούσης). My last objection is that still more is required if we are to find in a line or magnitude an analogue of the temporal present. Not only would the line need to be generated by the flow of a point, but it should be possible to ask, of an already existing line, 'Where on the line is the generating point *now*?' Plainly this is not a sensible question, unless it is asked with an eye on the drawing of the line or the movement of an *object* along it.

(iii) What does the doctrine of continuity tell us about the doctrine of the persisting now? Despite Aristotle's language at 222ª10 ff. it is not

[32] Cf. Hussey, *Physics III and IV*, 154 *ad* 219ᵇ9 ff.: 'we cannot see a particular direction along a line without seeing the line as swept out by a moving point, and we cannot specify a particular direction along a line without reference to particular points ("from A to B").'

[33] D. Bostock, 'Aristotle's Account of Time', *Phronesis* 25 (1980) 148–69, refers to *An.* 1. 4. 409ª4–5 for the view that 'the spatial point . . . generates a continuous line by "flowing" along it' (p. 162 and n. 12). But the *An.* passage attributes to others (presumably Xenocrates and other Platonists) the view that a moving point generates a line, not the view that any line is intrinsically the product of such a unidirectional movement.

[34] Cf. n. 15 above.

certain that the persisting now is involved in the passage (cf. (i) above). But even if it is so involved, it is tempting to argue that Aristotle cannot be maintaining that the now *qua present* persists and unites time, while the now *qua instant* is always different and divides time.[35] For the line, on which Aristotle bases his account of time, provides no counterpart to the present. But Aristotle's thought on this matter is subject to conflicting pressures. For an earlier account of the continuity doctrine (220^{a}4 ff.) indicates a disanalogy between time and a line:

> When one takes it thus, treating the one point [viz. the boundary between two segments of a line] as two, it is necessary to halt [ἵστασθαι], if the same point is to be both a beginning and an end. But the now is always different, owing to the movement of the object. So time is a number not like that of the same point, in that it is beginning and end, but like, rather, the extremities of a line—and not like the parts, both for the afore-mentioned reason (for one will treat the middle point as two, so that a pause will occur [ἠρεμεῖν]. (220^{a}12–18)

The subject of ἵστασθαι and ἠρεμεῖν may be either the line or the person, giving two possible interpretations: (*a*) If one regards point *C* on a line *AB* as the end of *AC* and as the beginning of *CB*, this presupposes that *AB* can come to a halt at *C*, breaking into two lines, *AC* and *CB*, which can be separated and rejoined again. But time (and a motion) cannot come to a halt in this way. Aristotle perhaps refers to this at 222^{a}13–14, when he says that the fact that the now is the end of one period and the beginning of another is not so clear as in the case of the stationary point, [since] the now divides potentially (while a point can presumably divide a line actually). (*b*) If a person regards a point *C* on a line *AB* as the end of *AC* and the beginning of *CB*, he will take time to do so, stopping at *C qua* end of *AC* and after a pause proceeding along the line from *C qua* beginning of *CB*. This temporal gap does not disturb the continuity of the line, since it is a spatial, not a temporal (or a kinematic) continuum. But if one attempts to treat time in this way, regarding a now as the end of the past and the beginning of the future, the procedure again takes time, only in this case it disrupts the continuity of time: At 6 o'clock one takes the now as the end of the past and then as the beginning of the future. But by the time one performs this second operation, it is no longer 6 o'clock, but 6.00.01. So either one treats 6.00.01 as the beginning of the future (with the period

[35] As Hussey believes: 'The persisting present is the now considered as linking past and future. The unrepeatable instant is the "now" considered as a potential divider of time' (*Physics III and IV*, 170 *ad* 222^{a}10 ff.).

from 6 o'clock to 6.00.01 unaccounted for) or one treats 6 o'clock as the beginning of it (though 6 o'clock is not the time *now*).

If, as seems likely, Aristotle had both these interpretations in mind, his view is perhaps this: the now, whether as instant or as present, divides time only δυνάμει, not ἐνεργείᾳ. But while I can think of a given *instant*, specified as, say, '6 o'clock', as the end of one period and the beginning of another, that is, as dividing time (if only potentially), I cannot think of the *present* as dividing time even in this way, since it changes in the course of my thinking. A similar consideration suggests that it is the now *qua* instant which changes over time, while the now *qua* present persists. For, unlike an instant, the present cannot be specified or marked out by anything else than the fact that it is present or *now*. If a clock strikes 6 o'clock now, the stroke marks not the present as such, but the instant, 6 o'clock, which, as it recedes into the past and ceases to be present, is still associated with the stroke of the clock. The present is essentially continuous, and our efforts to mark off different presents from each other result only in our distinguishing the different instants with which the present successively coincides. An obscure and much-emended remark at 220^{a}21–4 may also indicate a connection between the continuity doctrine and the postulate of simultaneity (see Part II, section 4 above):

> So in so far as the now is a limit [πέρας], it is not time, but an accident [of time]: but in so far as it counts, it is a number. For the limits belong only to that of which they are limits, but the number of these horses, the ten, is also elsewhere [καὶ ἄλλοθι].

The suggestion may be that in so far as the now divides time (as a πέρας), it is marked by the beginning or end of some particular event and is thus especially associated with that event, while in so far as it makes time continuous it is detached from any particular event and can thus be the same everywhere, from one event to another. If this is so, Aristotle is trying to break the attachment of time and the now to a particular magnitude, motion and object with which he began. More generally, if he is to claim that it is the present as such, rather than an undifferentiated now, that persists, while the instant alters over time, then he needs to qualify the analogy between time and the line. Since there are only occasional and inconclusive signs of this, the precise significance of the doctrine of the persisting now, and of its relationship to continuity and simultaneity, remains unclear. The notions of the instant and of the present may be too closely fused in his thought for us to separate them out from his concept of the now.[36] Nevertheless enough has emerged from the examination of these doctrines to return to Problem I and sketch Aristotle's solution to it.

III Conclusion

Schematically Problem I is this:

A. Time has two parts, the past and the future.
B. Neither of these parts is now.
C. The now itself is not a part of time.
D. If an entity is (now), then one or more of its parts is (now).
E. Therefore, time is not.

Aristotle attacks this argument in two connected ways: First, it is incorrect to regard time as an entity composed of parts, the past and the future, as A. assumes. Second, although the now is not a part of time (in the way in which the nuts and bolts holding together the parts of an object are themselves parts of that object), it is far more than the simple meeting-point of the 'parts' of time (like the meeting-point of two separate but contiguous sticks), which C. implies it to be, and thus its occurrence *now*, contrary to D., is sufficient to secure the existence of time. The relevant considerations are these:

(i) The parts of an object can in principle be separated so as to leave a gap between them, a gap which is not a part of the object. They can be divided ἐνεργείᾳ. But the 'parts' of time cannot be separated in this way. The now divides only δυνάμει (222^a14, 18. Cf. Part II, section 5).

(ii) The parts of an object are, or may be, marked off from each other by a seam or by some qualitative difference between them. But the present is not like a seam: it is marked out only by the fact that it is present. The past and the future do not differ qualitatively from each other, but only in that one is past and the other future (cf. Part II, section 5).

[36] The fusion of these notions in the now goes together with a similar 'ambiguity' in Aristotle's expressions for the past and future. A conflation of the past and future *simpliciter* with the past and future in relation to any given instant, seems probable at 222^a10–11, where the now is said to link time past and future (τὸν χρόνον τὸν παρεληλυθότα καὶ ἐσόμενον), at *Phys.* VI. 3, 233^b33 ff. (cf. Part I, section 5 above); and at 223^a5 ff., where 'before'; and 'after' are defined κατὰ τὴν πρὸς τὸ νῦν ἀπόστασιν, τὸ δὲ νῦν ὅρος τοῦ παρήκοντος καὶ τοῦ μέλλοντος. (The seemingly irrelevant introduction of the now into the definition of 'before' and 'after' might derive from this conflation or from the epistemological priority implicitly assigned to the now at 219^b24 ff. But cf. n. 27 above. At all events 'the before and after' of 223^a7 cannot *be* the now.)

(iii) An object, however, may be all of a piece, with no seams, joints, or qualitative variations. It may be divided at an arbitrary point, either physically or in thought, but we do not need so to divide it in order to recognize its presence. By contrast, a non-arbitrarily located now or present is essential both to time (e.g. 220^a1) and to our perception of it (e.g. 219^a27 ff.): time could not be unless the now successively marked its passage and we could not notice or measure time's passage if we could not discern distinct, successive nows.

(iv) Conversely, if an object has distinct parts which meet at a point or a surface, the point or area at which they meet will still exist as a point or area, even if the parts are separated or the object annihilated. The now, by contrast, could not exist if time did not exist; the now is essentially—in a sense—the end of the past and the beginning of the future (219^b33, 222^a33 ff.).

(v) Nevertheless, while we can pick out a point or area on an object at which its parts meet or join, we cannot so easily pick out the now as the end of the past and the beginning of the future: if we try to do so, we either pick out not a single present, but two or more presents, or we pick out a single *instant* which is not the present as such and does not bound past and future (220^a12 ff. Cf. Part II, section 5).

(vi) Aristotle also held that the now persists. This bears on Problem I in two respects. First, it casts doubt on the idea that time is composed of past and future as *parts* which meet at the now: the now, in so far as it persists, has its counterpart not in the point(s) at which the parts of an object meet, but in the object as a whole: and if what persists is the *present* as such, an object provides no counterpart to this.[37] Second, it enhances the status of the now: although instantaneous and not a part of time, the now is not fleeting and evanescent as, say, the instant *qua* 6 o'clock is. Its being is secondary, or rather tertiary, to the being of objects, but its temporal persistence is secured by the persistence of objects. The now in turn is related to time as a moving object is to its motion (219^b22 ff.), and hence the being of the now is surely sufficient to secure the being of time—not simply tenseless being, but its tensed being *now*.

Problems I and II, presented initially as distinct problems, receive essentially the same answer in Aristotle's text. The answer is involved and often obscure and ambiguous, but it is neither negligible nor obviously incorrect.

[37] Cf. Tatian, quoted in n. 20.

8

Aristotle on Continuity in *Physics* VI

DAVID BOSTOCK

Aristotle's favourite model of the continuum is the same as ours, namely a geometrical line, or line-segment. He does not also have, as we do, the numerical model of the real numbers, but that is scarcely a handicap. After all, one of our main ways of understanding the structure of the real numbers (namely via the notion of a 'Dedekind cut') is very naturally viewed as drawing upon our prior grasp of the structure of a geometrical line, with the real numbers understood as corresponding to the points on that line. There are some features of this structure which Aristotle grasps very clearly, notably that no two points on a line are ever *next* to one another, for between any two points there is always a whole line-segment, which in turn will always contain further points within it. Undoubtedly, this is a major achievement on his part, but its importance should not be over-exaggerated, as it appears to be (for example) in Ross's classic edition of the *Physics*.[1] (Looking at the issue from our contemporary point of view, we may observe that the fact that between any two points there are always others does not yet distinguish the structure of the real numbers from that of the rational numbers.) Ross's general evaluation of the discussion in the first two chapters of *Phys.* VI is that 'it seems to me to indicate that Aristotle had a more mathematical turn of mind than he is usually credited with' (p. 70), and his commentary on these chapters contains no important criticism.[2] As for the remainder of Book VI, this he says 'develops with unwearied diligence and (so far as I can judge) with unfailing accuracy the implications of the infinite divisibility of time and of extension' (p. 71). We shall see, however, that Aristotle's discussion in fact leaves

[1] W. D. Ross, *Aristotle's Physics* (Oxford, 1936).

[2] Ross does note, on 233^{a}18–21, that it is a little surprising to find Aristotle asserting that if time is infinite in extent then so is distance, since his own position is that time is infinite but space is not. The difficulty is removed by recognizing that throughout this chapter Aristotle is speaking of the time taken, and distance covered, by a body in uniform rectilinear motion.

a number of important gaps, is at times seriously muddled, and contains some plain mistakes. I aim to make these inadequacies clear, and then to add some brief speculations on why they are present.

There is one significant feature of Aristotle's whole discussion in Book VI, which I would not regard as an 'inadequacy', but which deserves a mention here at the outset. That is the tension between Aristotle's reasoning in this book and his discussion of infinity in Book III. To put it very roughly, his position in Book III is that there is no 'actual' or 'completed' infinite, and all infinity is merely 'potential'. But in Book VI he shows no such tendency to be suspicious of the infinite, and apparently accepts without qualms that a line does ('actually') contain infinitely many points, that a stretch of time does ('actually') contain infinitely many instants, and so on. Thus his answer in this book to Zeno's most famous paradox on motion apparently accepts that one who traverses a finite distance has thereby completed a series of infinitely many distinct tasks, in traversing the infinitely many 'half-distances' contained within the original distance. This is possible, he here says, because the finite time available may equally be divided into a corresponding and infinite series of 'half-times' (233^a13–31, 239^b9–29). Later in Book VIII he will reconsider this paradox, and with the doctrine of Book III in mind he will deny that a moving body ever does pass over infinitely many 'actual' points (263^a4–$^b9$). But here in Book VI we find no such complications. A simple hypothesis evidently suggests itself: Book VI was written before the doctrine of Book III was worked out.[3] (In my concluding section I shall note some other pointers to the early date of Book VI.)

I The Definition of a Continuum

In chapter 3 of Book V Aristotle defines what it is for one thing to be continuous with another. He begins by defining 'succession' (τὸ ἐφεξῆς): one thing succeeds another when it comes after it, in position

[3] Books VI and III do share similar mistakes about infinity. In VI. 7, Aristotle argues that even an accelerating or decelerating object cannot be moving in such a way that it covers a finite distance in an infinite time (237^b34–238^a19), or an infinite distance in a finite time (238^a20–31). Basically his mistake is that he assumes that an infinite time, or distance, cannot be divided into one finite and one infinite part (238^a9–11). This may be compared with his claim in III. 5, that an infinite body cannot have several infinite parts (204^a25–6; cf. 204^b19–22). It may be that one of these errors is influencing the other, though there is no strong reason to suppose so. (Note that III. 5 raises no *logical* objection to the division of an infinite whole into one finite and *one* infinite part, 204^b13–19.)

or in some other way, and there is nothing between them that is of the same kind (226^b34–227^a6). A special case of succession is 'being next to' (τὸ ἐχόμενον): one thing is next to another when it succeeds it and is also in contact with it (227^a6), contact being defined as occurring when (some part of) the 'limits'[4] of the two things are in exactly the same place[5] (226^b21–3). A special case of being next to is 'continuity' (τὸ συνεχές): one thing is continuous with another when they are next to one another and in addition the limits at which they touch are the same thing, or 'have become one', being indeed 'held together' as the word συν-εχὲς implies. Aristotle's thought evidently is that things which are merely next to one another need not hold together—one may move one of them leaving the other where it is—whereas if they are continuous then they move together (227^a10–17).

At the opening of Book VI Aristotle recalls this definition,[6] and appears to suppose that it has also explained to us what it is for a single thing to be a continuous thing—or, as I shall say, a continuum—though clearly it has not explained this at all (231^a21–6). But if we follow through the argument that he at once plunges into, we can I think reconstruct the definition that he has failed to state but must have had in mind. For he claims that it follows from the definition that a continuum (such as a line) cannot be made up from indivisible things (such as points). The first argument is simply that an indivisible thing has no limits, for a limit (ἔσχατον) must be different from what it limits, and thus what has limits must have parts, and must therefore be divisible (231^a25–9). As a second argument Aristotle adds that indivisible things such as points cannot form a continuum by being in contact with one another, even if we waive the point about their lack of limits. For, he says, if a point could be in contact with a point then they would have to be in contact 'as wholes', and this (I think) he takes

[4] ἄκρα, 226^b23 and 227^a22–5; πέρατα, 227^a12; ἔσχατα, 227^a13.

[5] ἅμα, as defined at 226^b21–2. I take it that Aristotle means what he says, and is (rightly?) supposing that when the face of one (perfect) cube touches the face of another the two faces are different things but are in *exactly* the same place. (By contrast, *Met. B.* 5, 1002^a34–$^b3$ appears to suppose that when two such faces touch they automatically do become one, and are no longer two.) Admittedly, this perfectly proper use of 'same place' is somewhat difficult to reconcile with Aristotle's definition of 'place' in Book IV, which seems not to allow such things as limits to have places.

[6] As Ross notes (*Aristotle's Physics*, 626–7) being next to is defined in Book V as a *special* case of contact, namely where the items in contact are also successive. (Compare Plato, *Parmenides* 148e4–7, 149a4–6).) But usually Aristotle draws no distinction between contact and being next to, and simply equates the two, as he does here at the beginning of Book VI.

to imply that they would have to be in exactly the same place. At any rate he objects that contact of this kind will not produce anything continuous,[7] because a continuous thing will be divisible into parts that occupy *different* places[8] (231^{a}29–^{b}6). (The thought here seems to be this. When one (perfect) sphere touches another there will be a point on the surface of the one that touches a point on the surface of the other, and these are different points, just as the face of one cube in contact with the face of another are different faces (see n. 5 above). Hence one point may be said to be in contact with another, but such contact yields spatial coincidence, so no matter how many points we may put together that are in contact with one another in this way we shall still not form anything bigger than a single point. But this subtlety is generally ignored in what follows, where it is simply claimed that indivisible things cannot be in contact (e.g. 231^{b}17).) Finally he adds as a third argument that points cannot even be successive, since between any two points there will be a line (and hence, he presumably means to add, a further point) (231^{b}6–10). Summing up this reasoning,[9] Aristotle concludes that what a thing is made up from it may also be divided into, and so we have shown that a continuum cannot be divided into indivisibles. Anything whatever must be either divisible or indivisible, and if divisible then either divisible into indivisibles or divisible only into what may always be divided further; this last is a continuum. And conversely every continuum is divisible into what may always be divided further, for if it could be divided into indivisibles then one indivisible could touch another, 'for the limits of (two) continuous things touch, and are one' (231^{b}10–18).

Now, if this argument is to be intelligible, what must we take the definition of a continuum to be? First, it must evidently be given as a premiss that a continuum has at least two parts which do not coincide (231^{b}4–6). For Aristotle clearly does not count a point as *itself* a continuum, and this seems to be the minimal premiss needed to rule out that suggestion. To avoid some complications I should like to

[7] *ὅλον δ' ὅλου ἁπτόμενον οὐκ ἔσται συνεχές* (231^{b}4). I take *ἔσται* to be existential, and *ὅλον . . . ἁπτόμενον* to be what Ross calls a 'loose accusative absolute'.

[8] *τόπῳ κεχωρισμένα*. Earlier in Book v *χωρίς* was defined simply as meaning 'in a different place' (226^{b}22–3). I imagine that what Aristotle has in mind here is places that do not overlap but do touch.

[9] Ross sees the reasoning from ^{b}10 to ^{b}15 as still concerned with the suggestion that points might be successive. This is mainly because, though he brackets the words *τῶν στιγμῶν καὶ τῶν νῦν οὐθέν* at ^{b}13, he takes them to be a *correct* gloss. But my summary presumes that they give quite the wrong sense.

strengthen this premiss a little to the following: a continuum may always be *divided* (without remainder) into two parts that do not coincide. It is true that, given a 'logic of parts' that is nowadays orthodox, the supposedly stronger version would follow at once from the original. For the now orthodox logic supposes that, given any one proper part of a thing, there will always be some *one* further part of the thing, which comprises all the rest of it excluding the given part. Thus the whole may be said to be divided into these two parts in the sense that the two do not overlap one another—i.e. they have no common part—and together they exhaust the original whole, in so far as every part of it must overlap at least one of these two. But it seems to me doubtful that Aristotle would accept the assumption on which this reasoning is based. For example, it implies that if we begin with a line three inches long, and subtract a part one inch long from the middle of it, then the two separated line-segments remaining may be counted as together forming *one* part of the original. But this is not a very natural use of the notion of one part, and equally it is not very natural to say that the operation just envisaged would divide our three-inch line into just two parts. I shall come back to this problem shortly, but for the moment let us simply side-step it, by adopting the stronger premiss: any continuum may always be divided into two.

Then it is perfectly clear from the whole run of Aristotle's discussion that we must add a further premiss: any division of a continuum into two (non-coincident) parts must divide it into parts that are continuous with one another in the sense defined in Book v, i.e. the two parts must touch, and the limits where they touch must 'be one'. These two premisses are, I imagine, the premisses that Aristotle regards his argument in this passage as depending upon, and as given by the (unstated) definition of what a continuum is. We may therefore reconstruct the definition by designing it to yield just these premisses: a continuum is anything which (i) can be divided into two parts, and (ii) is such that any two parts into which it is divided must share a limit. But if this is indeed the definition Aristotle has in mind, then either it is incorrect or the conclusion that he attempts to deduce from it will not follow. That is, it will not follow that the parts into which a continuum may be divided must themselves be further divisible.

Suppose we take a finite line, and suggest 'dividing' it into these two parts: one part is to be an end-point of the line, and the other part is to be all the rest of the line, excluding this end-point. (Thus the second part is what we call a 'half-open' interval, containing all the points of the line except this one end-point.[10]) Now we may surely assume that a finite line is a continuum, if anything is, so if Aristotle's definition is correct then the two parts into which we have divided it must share a limit. According to one of Aristotle's lines of thought, they do not, since he claims that a point has no limit. But in that case we must simply reject the definition as incorrect. However, if my interpretation is right, then he also has another line of thought which allows points to be in contact with one another ('as wholes') despite their alleged lack of limits. This is most easily harmonized with his general position by allowing that a point may be said to be *its own* limit. If we do allow this, then the suggested definition of a continuum may be retained, even in the face of this example. For the end-point of the line, and the rest of the line, now do share a limit, namely the end-point itself. It is its own limit, and it must also be reckoned to be one of the limits of the rest of the line. For we cannot say that the rest of the line has no limit in this direction without once more subverting the suggested definition, and if it is to have a limit then clearly there is no other candidate than this same end-point. Thus, if the definition is to be retained, while admitting this example as a 'division', then we must conclude that it is possible to divide a line into two parts, where *one* of these parts is a mere point, and hence not further divisible. (And we may add, incidentally, that *any* point of the line may thus be exhibited as one of the parts into which it may be divided, by considering a division into three parts, one of which is the given point, one the rest of the line (if any) that is to the left of it, and one the rest of the line (if any) that is to the right of it.)

There are two possible avenues one might explore, in seeking to rescue Aristotle from this objection. One would be to note that I have merely conjectured his definition of a continuum (since he fails to state any definition himself), and to seek for an alternative conjecture. Here one may remark that he quite often appears to take, as *definitive* of a continuum, the thesis that in this passage he is attempting to prove. That is, a continuum is defined as what is divisible (only?) into parts that are further divisible. (This appears to be cited as a definition a little later in this very book of the *Physics*, at 232^b24–5; it is also cited as a definition elsewhere, e.g. *Cael.* 268^a6–7.) But there are two obvious

[10] One might claim that Aristotle himself may be said to recognize the existence of such things as half-open intervals at *Phys.* VIII. 8, 263^b9–264^a6, where he perceptively remarks that an interval without its end-point is no shorter than the interval with its end-point (264^a4–6).

objections to any such suggestion. One, based on the passage we have just been considering, is that such a definition simply omits a feature that Aristotle evidently regards as crucial: it does not include the condition that the parts into which a continuum may be divided must share limits. The other, based on the objection that we have just been considering, is that any definition designed to rule out that objection would run the risk of having as a consequence that a geometrical line is not after all a continuum. But this consequence is evidently intolerable. The alternative escape route is to put some restrictions on what is to count as a 'division' so that the example we have considered does not qualify. It is not easy to see quite what restriction would be appropriate, but whatever is suggested there will always be this reply: if we are only allowed to count as a division something which divides a continuum into extended parts, each further divisible, then it will no longer be plausible to claim (as at 231^{b}10–11) that whatever a continuum is 'made up of' it will also be 'divisible into'. That is, a line may perhaps be constructed out of points, and nothing else, even though it is not 'divisible' into points, according to the restricted understanding of a 'division'. I conclude that neither of these avenues of escape will in the end prove satisfactory.

But here we should notice a way of modifying Aristotle's argument, which still yields the conclusion he was aiming for, and derives this conclusion perfectly correctly. It requires, however, a much stronger premiss on divisibility. We have said so far that any way of dividing the line into just two parts must yield parts that share a limit (and we have observed that one of those parts may be taken to be a mere point without disturbing this premiss). By repeating such divisions, we can evidently infer that any way of dividing the line into a finite number of parts must yield parts that share limits (and several of those may be taken to be mere points, so long as others are line-segments joining them). But the generalization that Aristotle's argument requires is that any way of dividing the line into any number of parts—finite or infinite—must yield parts that share limits. However, no point shares a limit with any other point, and hence the line cannot be divided into *nothing but* points. But what a thing is made up of it may also be divided into, and hence a line cannot be made up of nothing but points. This, evidently, is an argument of some power, and it is not at all surprising that Aristotle found it compelling, but of course we are free to reject the strong premiss on divisibility that it is based on.

But still, though we may reject Aristotle's premiss, we can yet feel some sympathy with his position. How *can* we conceive of a line as made up of nothing but points, which cannot touch one another? Imagine that you have an inexhaustible supply of points to hand; how could you 'put them together' (as it were, 'one at a time') so as to make a line out of them? Well, we can say nowadays that the answer is this. First set down a denumerable infinity of points that is dense in the line, for example by putting down the two end-points of the desired line, and then by setting down a further point in the middle of any stretch between two points already set down. (Note that we do *not* try to put the points down 'in succession' from left to right.) When you have finished this (!), then consider all the Dedekind cuts in the points already set down, in other words all ways of separating those points into two groups, with all of one group to the left of all of the other. For each such cut, put in a further point to be the point at which the separation is made (if you do not have one already—i.e. if there is not either a rightmost point of the left group, or a leftmost point of the right group). We may note, incidentally, that to complete this stage you will in fact need more than a denumerable infinity of points (so it may take you a little longer than the first stage), but when you have done it you are through. The result is now a continuous line. But although we have a proof of this fact, must we not still admit that it is absolutely amazing? No point that has been put down touches any of the others, and yet the result is a line with no gaps in it anywhere! No wonder that Aristotle could not see how such a construction of the line could succeed. Indeed, I think I would admit that I cannot exactly 'see' it myself.

So much, then, for Aristotle's opening argument in Book vi, as seen from a modern perspective. I now add two observations (from the same perspective) about the definition on which it is based. The first follows naturally from the mention of Dedekind cuts in the last paragraph. Elsewhere in *Phys.* vi Aristotle is generally content to characterize a continuum as 'divisible into divisibles' (and hence 'infinitely divisible'), but he pays surprisingly little attention—either here or elsewhere—to the question of how much divisibility a continuum has. For example, when he is arguing that time is a continuum, he seems content with a proof that any time, however short, can be divided to yield parts that are yet shorter (232^b20–233^a12). He is prepared to add that any time can be divided in a specified ratio (233^a3–5), or again that any time can be divided both into two equal parts and into three equal parts (233^b19–32), but these points seem merely incidental to his main claim, which is just that time is infinitely divisible. One might have expected him to stress that a continuum can always be divided in *any* desired ratio, not only those ratios that can be specified by the natural numbers, but also those irrational ratios that geometry forces upon us, e.g. the ratio of $\sqrt{2}$ to 1. (These, of course, correspond to the Dedekind cuts mentioned above.) But in fact there is no emphasis on this point, and indeed no mention of irrational ratios in the whole of Book VI. I can only comment that I find this surprising. As a matter of fact it can be proved, from what I at least regard as plausible assumptions, that if there is no smallest time then it does *follow* that any time may be divided in any desired ratio, rational or irrational.[11] But one can hardly suppose that Aristotle knew of any such proof. As for the characteristic of a continuum that Aristotle stresses here, concerning the sharing of limits, this does indeed have a role to play in the modern conception, as elaborated in topology. I end this section with a few brief and sketchy remarks on this.

In modern topology one takes as primitive and undefined the notion of an open set of points. This, one might say, corresponds to the notion of something that is extended, that fills space. We may remark here that ordinary extended things will be extended in one, two, or three dimensions, and the dimensions will introduce an order, allowing us to speak of a series of extended things, coming one after another in the order corresponding to that dimension (e.g. from left to right, from top to bottom, from earlier to later). This explains the relevance of Aristotle's notion of succession, which receives little attention in his own account. But indeed it is not of fundamental importance. What is much more important is that an extended thing will (normally) have a limit, which of course corresponds to the points in the closure of the corresponding open set, but not in the set itself. For most purposes, it is natural to think of an extended thing as including its limits, but we have now learnt how important it is to be clear about this, one way or the other. Now any extended thing may be divided into two others (of the same dimensions), and this corresponds to the partition of an open set into two open subsets, plus the points, if any, that are in the limit of both. But in the general case there need not be any points in the limit of both, for an extended thing may consist of several scattered parts, not in any way joined up with another. That is to say, an open set may

[11] See e.g. my *Logic and Arithmetic*, ii (Oxford, 1979), 173–7, 237–9, 248–53. But my 'plausible assumptions' include a principle that provides for the existence of Dedekind cuts (see in particular pp. 270–4).

be a union of open sets that are not in any way *connected* with one another. This is where Aristotle's definition comes in: it introduces the requirement of connectedness. So a continuum is a special case of an extended thing, namely one where any two parts into which it may be divided must share a limit.

Now the notion of connectedness proves, on investigation, to be very much more complicated than one might at first have supposed. (For example, we need to distinguish between being connected 'pathwise' and being connected in other ways.) That is why Aristotle's definition is not the one used in modern topology, though I cannot elaborate here on the issues involved. But from our perspective it can easily be seen as an attempt to introduce just this notion; its role is to distinguish extended things which are 'all joined up' from those that consist of scattered parts. But then one must observe, with some irony, that it seems improbable that Aristotle himself was aware of this role, since it seems improbable that he recognized the existence of any extended things that were not continua.

II The Continuity of Space and Time

For the remainder of chapter 1 of Book VI (from 231^{b}18), and for all of chapter 2, Aristotle's aim is to show that time and space are each infinitely divisible, and hence continua.[12] His argument depends upon the premiss that motion is also continuous, which he explains in the same way as meaning that any motion is always divisible into further parts that are motions (231^{b}18–20), but this idea needs a little elaboration.[13] Throughout the discussion in these chapters, Aristotle is considering only uniform rectilinear motions, covering equal distances in equal times, and he counts such a motion as continous if the moving body traces out a path through space that is a continuous line. That is, each point on the line is occupied by the body at some instant during the time of its movement. The alternative that he considers is that motion proceeds, as we might say, 'cinematographically' (or, as Aristotle says, ἐκ κινημάτων rather than ἐκ κινήσεων, 232^{a}8–9), which

[12] Ross's title for ch. 2, viz. 'Further proof that every continuum is infinitely divisible', is thus misleading.

[13] From our point of view, the continuity of motion is a matter of the continuity of a *function*, namely the function which assigns to each instant during the motion the position occupied at that instant.

is to be envisaged thus: the moving body remains at one position for a stretch of time, and then 'leaps' instantaneously to another and different position, without ever occupying the positions in between. It then rests at the new position for a further stretch of time, then 'leaps' instantaneously to yet another position, and so on. Aristotle argues against this alternative hypothesis at 231^b28–232^a18.

Suppose we have a spatial stretch *A* and our allegedly moving body 'moves' across it by staying for a while at one end of it and then instantaneously leaping to the other. Then Aristotle in effect claims that there is no time when the body *is moving* across *A*; there is a time when it is not yet moving across *A*, and there is a time when it has moved across, but no time when it is crossing. But, he says, one cannot be in a position in which one *has* walked to Thebes without at some time being in the process of walking to it. And it is no use, he suggests, to suppose that a longish journey, such as that from Athens to Thebes, is actually made up of a whole series of such small instantaneous leaps, so that I can say that I *was* walking to Thebes during all the time when that whole series was going on. For that supposed movement would be made up of things that clearly were not movements, but states of rest (before an instantaneous transition) and states of *having* moved (after it). But a movement, he claims, can only be made up of movements, just as a line can only be made up of lines and not of points; no other alternative is possible.

To this one can only reply that the alternative that Aristotle himself quite clearly describes in this passage is evidently not impossible at all. No doubt we do not believe that movement takes place through such a series of instantaneous leaps, separated by rests, and for that reason our usual vocabulary for talking of movement is not well adapted to this possibility. But that is surely of no significance. It is simply an empirical question whether movement is in this way discontinuous, and one that is hardly settled by the observation that it does not look that way to the naked eye. But if the continuity of movement is best regarded as a straightforwardly empirical question, should we say the same of the continuity of space and of time?

What is undoubtedly true is that all of Aristotle's arguments for the infinite divisibility of space and time do depend upon the premiss that motion is continuous, and will collapse if that premiss is denied. I illustrate this claim by considering just his final argument at 233^b15–31, for (unlike the others) that argument appears at first sight to be independent of the continuity of motion. We may reasonably suppose that two bodies may move (uniformly) in such a way that the ratio between their velocities is as three to two. We may also suppose that if there are indivisible time-atoms and space-atoms[14] then the faster body is moving at such a rate that it traverses exactly one space-atom in exactly one time-atom. Thus after three time-atoms it has crossed three space-atoms, and in the same time the slower body will have crossed two space-atoms. Then, Aristotle argues, the slower body will cross one space-atom in half of three time-atoms, which is only possible if the supposed time-atoms are after all divisible, and hence not atoms at all.

Now, if we assume that the motions in question are continuous, this is no doubt a fair conclusion. But if we suppose instead that the motions are cinematographic then the conclusion is easily avoided. The faster body, we may assume, proceeds by remaining at rest for one time-atom and then instantaneously leaping a distance of one space-atom. The slower body must proceed in a manner which has the result that it has crossed two space-atoms in three time-atoms. Evidently the simplest hypothesis is this: it proceeds by remaining at rest for three whole time-atoms and then instantaneously leaping a distance of two whole space-atoms. We do not have to suppose that in cinematographic motion a 'moving' body never rests for more than one time-atom, nor that it never leaps more than one space-atom. We can make the first supposition if we wish to, and preserve the atomicity of time, while allowing the infinite divisibility of space. Equally, we can make the second supposition, and preserve the atomicity of space, while admitting the infinite divisibility of time. But we do not have to make either supposition. So we can if we wish maintain the atomicity of both space and time, so long as we also claim that the ratio of one 'uniform' velocity to another can always be measured by a pair of whole numbers. And clearly no straightforward empirical observation could provide a direct refutation of that claim.

It will be useful to add a word here on how one should conceive this hypothesis of the 'atomicity' of space and time. The basic idea is that although time, say, *may* be viewed as forming a continuum, such a view turns out to contain superfluous detail, in that it posits the existence of many temporal instants which need never be invoked in the explanation of physical phenomena. For the only instants at which any change

[14] I shall try to elucidate this notion in a moment. Meanwhile observe that in my usage an atom is extended, and may even be said to have parts, though there is a sense in which it is not divisible.

ever occurs are the instants that separate each time-atom from the next. Thus the time-atoms are conceived as having duration, and may be said to be 'divisible in thought', so that we may speak of the state of affairs that obtains 'half way through' a time-atom. (Indeed, the claim will be that exactly the same state of affairs persists all through the time-atom.) But this 'divisibility in thought' corresponds to no reality, and is merely a product of our way of conceiving of time, namely as a continuum measurable by the real numbers. Perhaps, if we had reason to suppose that time really was atomic in the way suggested, we might (for scientific purposes) adopt a different way of looking at it, taking it as a matter of definition that each time-atom was equal to every other, and thus measuring time by the natural numbers (used simply to *count* the time-atoms), rather than by the real numbers. (One might compare the way that we adopt, for scientific purposes, an 'absolute' scale of temperature, which eliminates those degrees of temperature, below -273 °C., to which no reality corresponds.)

Certainly, we could not adopt such a standpoint without abandoning many deeply engrained ways of thinking. Quite a nice illustration of this is to be found in Aristotle's argument in chapter 3. He claims there that what he calls 'a now, in the primary sense' must be conceived as an indivisible instant and not a stretch of time, offering this argument. There will be an instant which is the limit of time past, and an instant which is the limit of time to come, and we cannot suppose that 'the present time', i.e. 'now', is some stretch of time between these instants. If we try to think of it in this way, then we must admit that this stretch is divisible, and then a (suitably chosen) division would itself have to be 'the now, in the primary sense', dividing the past from the future (233^b33–234^a24). Putting his argument the other way about, we may say that his point is that if we try to conceive of an 'atomic present', which lasts for some time, but which is 'atomic' in the sense that nothing changes during that time, then we will find an insuperable obstacle. For at least *the time itself* must be conceived as changing. That is, the limit of time past must be envisaged as moving continuously through the supposed atomic present, and cannot be thought to rest at one end of it and then leap instantaneously to the other. But the reply is simply that although it may indeed be difficult to conceive the 'flow' of time in this way, it would not appear to be impossible. (And we may add, *ad hominem*, that Aristotle's own doctrine that time depends upon change would require us to say that time itself cannot be continuous unless some *other* change is also continuous.)

To generalize, scientific theories of this century—most notably the theory of relativity—have taught us that many common-sense assumptions about the nature of space and time are in error, and should be abandoned. The continuity of space and time might perhaps be another such common-sense assumption, which we will eventually have to abandon. One can at least roughly envisage developments of quantum theory which would point in this direction, and the rival theory of 'atomicity' is perfectly coherent. But of course we all nowadays accept Aristotle's view that space and time are continua, and that motion too is in its way continuous, and I too shall accept these views in what follows.

III States at an Instant

Much of Aristotle's discussion for the remainder of Book VI is concerned with what may be true of a thing at an instant. He begins on this during the rest of chapter 3 (from 234ᵃ24), with an argument that purports to show that there can be no motion or rest at an instant.

The claim as he first phrases it is that 'nothing moves in an instant' (οὐθὲν ἐν τῷ νῦν κινεῖται), and his first argument for it is evidently that nothing can cover any stretch of distance during an instant (234ᵃ22–31). This is so far quite compatible with our being able to say that a thing *is moving* at this or that instant, but it soon becomes clear that Aristotle wishes to rule this out too. We find that something odd is going on in the next sentence, where he claims that just as nothing moves in an instant so also nothing rests in an instant, his ground being that one can only speak of rest where motion is equally possible (234ᵃ31–4). But clearly the argument that he has just given for denying motion in an instant would not directly apply to rest. However, we have two more arguments to come. In the first, he asks us to consider a stretch of time divided by an instant into two sub-stretches, with a body that is moving throughout the whole of one sub-stretch and at rest throughout the whole of the other. Then, he says, since the dividing instant is 'in' both sub-stretches, if we could speak of the body moving or resting at that instant we should have to say that it was both moving and at rest then. This he evidently regards as an impossible result (234ᵃ34–ᵇ5). Finally, he argues that to be at rest is to be in the same state 'both now and earlier', and there is no 'earlier' in a single instant, and hence no rest then either (234ᵇ5–7).

It is clear that these arguments are wholly inadequate. In reply to the first one might say that there may perhaps be a special problem over the instant of change from rest to motion or from motion to rest, but there is no such problem about any other instant in our stretch. We may surely say that the body is moving at any instant that falls within the one sub-stretch, and is at rest at any instant that falls within the other. Moreover, the final consideration appears to give us a way of resolving the problem over the instant of change itself, for it evidently directs us to say that the instant of change from rest to motion is an instant of rest. If we may add that there is a further way of being at rest, namely to be in the same state 'both now and later', then we may similarly infer that the instant of change from motion to rest is also an instant of rest. Thus there will be first and last instants of rest, but no first or last instants of motion. All that is needed is a *decision* as to what to say about these special instants of change; and it is surely an extravagant reaction to ban *all* talk of a thing's state of motion or rest at an instant. (I add, incidentally, then the decision just suggested appears to be a perfectly reasonable one, and may be further supported by a point that I take from Sorabji.[15] If we may assume that all changes of velocity are continuous, then clearly the instantaneous velocity of an object at the instant of change from rest to motion, or motion to rest, must be zero.)

We get no further argument on this issue in *Phys.* VI (nor, I think, elsewhere in Aristotle). The arguments we have considered are repeated in chapter 8 at 239ᵃ10–22, but no new point emerges. One thing that we do find is that Aristotle intends his conclusion to apply, not just to motion proper, but to change of all kinds: one is not permitted to say, of any instant, that a thing is changing at that instant (237ᵃ14). It should also be noted that at one point in chapter 5 Aristotle himself seems to speak of a (last) instant of rest (236ᵃ18), but this must presumably be taken to be either a slip on his part, or (more probably) a claim that he thinks that his opponent in that passage—namely one who believes in time-atoms—would have to grant.[16] It is clear that his considered doctrine is that there is no rest, movement, or other change, at an instant. But what is difficult is to reconcile this

[15] R. Sorabji, 'Aristotle on the Instant of Change', *Proceedings of the Aristotelian Society*, Supp. vol. 50 (1976), 69–89.

[16] The second suggestion is due to Sorabji, in his *Time, Creation and the Continuum* (London, 1983), 415, n. 17. This is a recantation of the view put forward in his earlier article (cited in n. 15). I discuss the suggestion in section VI below.

view with some further claims that he goes on to make about what *can* be said to hold at an instant.

In chapter 5 he argues in effect that if a thing changes from being in state *A* to being in state *B*, then there is a first instant at which it *has* changed into state *B*, i.e. has completed its change. Evidently, this instant must be the same instant as that which separates its not being in state *B* from its being in state *B*, and so Aristotle's claim is that at that instant of change it counts as having reached state *B*, and therefore is then *in* state *B* (235^b6–236^a7). In contrast to this, he goes on to argue that there is no first instant at which the thing has begun its change, or has changed at least a bit (236^a7–27), and this theme is elaborated further at the beginning of the next chapter (236^b19–32). It is evident from this that the changes Aristotle is considering here are changes that take place over a whole stretch of time. Let us simply accept this for the time being. (I shall consider it more closely in my next section.)

Now the obvious difficulty with this is that a change from, say, travelling at ten miles an hour to being at rest, or vice versa, would seem to be included in the discussion. But then we must say, according to the doctrine propounded, that there is a first instant at which this change is completed, i.e. a first instant at which the thing has come to rest, or has reached the speed of ten miles an hour. And it seems very natural to add that an instant at which a thing *has* come to rest, or *has* reached a certain speed, will also be an instant at which the thing *is* at rest, or *is* moving with that speed. But of course if we do say this then we are in flat contradiction with what has been argued earlier.

It is not easy to see how Aristotle proposes to avoid this contradiction. It is not that he wishes to deny the inference from 'has come to be in state *B*' to 'is in state *B*'. Indeed chapter 5 opens with an argument *for* this inference (235^b6–30), and equally when he recapitulates the doctrine in chapter 8 of Book VIII he very clearly accepts that at the first instant at which a thing has come to be white, at that instant it is white (263^b9–264^a6). Moreover, he explicitly accepts the inference for the particular case of coming to be at rest, a topic to which he devotes chapter 8 of Book VI. Coming to rest, he first argues, is a process which takes time, though there is no smallest time in which it takes place (238^b23–239^a10). But then he adds that there is no first instant at which a thing *has* come to rest, precisely because that would imply an instant at which it *is* at rest, and this he has denied (239^a10–14). (And he goes on to remind us of his reasons, as stated earlier in chapter 3,

239^a14–17.) So he then infers that being at rest is also something that takes time, and again that there is no first time in which it occurs. If we put this together with the doctrine of chapter 5, that for *every* kind of change there is a first instant at which the thing has reached the terminal state of the change, then we can only infer that in Aristotle's view a change from moving at a given speed to being at rest (or vice versa) does not count *as* a 'change' (μεταβολή). The apparently restrictive doctrine of chapter 2 of Book v, which argues that all change can be classified under just four headings, may indeed have been intended to point forward to this result. But we may still complain that it is without adequate motivation.

I add that it is of course perfectly common usage to speak of a thing as moving with a given speed, or as being at rest, *at* this or that instant, and it is a usage which Aristotle himself falls into without noticing it. His thesis that whatever is changing has changed, i.e. has changed at least a bit, is quite naturally expressed by stressing that it has *already* changed, at that very same time at which it is changing. Moreover, Aristotle himself later proceeds to elaborate his thesis further, claiming that what is changing has *previously* changed, i.e. at a time *before* that at which it is changing (236^a34, 237^b5, 10–11, 18, 20). But we cannot take this to mean that it has changed before the *stretch* of time during which it is changing, without evidently falsifying the thesis. What is intended is clearly that for any *instant* at which it is changing there is an earlier instant at which it has changed (or, in the first version, that any instant at which it is changing is also an instant at which it has changed). One has to go a long way round about even to express the thesis that Aristotle intends if one cannot speak of instants at which a thing is changing.

It was, then, fairly clearly a mistake on his part to deny that we can speak of rest, motion, or change at an instant, and not only because the notion of instantaneous velocity is of course crucially important to Newtonian physics. (No doubt, Aristotle would not have found it easy to give an adequate definition of instantaneous velocity, but that is hardly a good reason for denying its existence.) On the contrary, it is a perfectly common-sense notion, and one which it is difficult to avoid when stating Aristotle's own position. Moreover, its denial introduces a quite needless conflict with his thesis that for any change there is always a first instant at which it is completed.

As for this latter thesis, I remark for the moment only that we have been given no reason to accept it. (All that Aristotle actually argues in

235^b30–236^a7 is that *if* there is a first time in which a thing has changed, then that time must be an instant.) It may, however, seem to be a perfectly reasonable thesis, when what we are considering is a change from being in state *A* to being in state *B*, and this change is one that takes time. For the time taken by the change will then be bounded by two instants, and it will be natural to assume that the second of these will also be the first instant of being in state *B*, and the first of them will also be the last instant of being in state *A*. (Aristotle does explicitly affirm one of these assumptions, as we have noted; as we shall see shortly, he is also committed to the other.) Hence there will be, as also seems plausible, neither a first nor a last instant at which the thing is changing from being in state *A* to being in state *B*. But all this is on the hypothesis that the change itself does last over a stretch of time. Aristotle proceeds to affirm that *all* change is of this kind.

IV All Change Takes Time

During the first half of chapter 6 Aristotle further elaborates his claim from chapter 5, that there is no first instant at which a thing has begun its change, or has changed at least a bit. He expresses this, as we have noted, as the thesis that if a thing is changing then it has changed (236^b19–237^a17), and he then proceeds to argue also for the converse, that if a thing has changed then there must have been a previous time during which it was changing (237^a17–$^b9$). Now we have already met this claim in chapter 1, as a claim specifically about motion: if a thing has walked to Thebes, then earlier it was walking to Thebes (232^a10–11). But while we may perhaps accept that this is true of motion (thus ruling out 'cinematographic' motion), what Aristotle is now urging is that the same holds of every kind of change whatever, and this clearly cannot be accepted.

His argument is this. Suppose that a thing has changed 'in an instant' from being in state *A* to being in state *B*. Then it has not changed to state *B* 'in the same instant in which it is in state *A*'. Yet if it changes (from state *A*?) in any other instant, then there will be a stretch of time between the two, and the change will have been going on throughout that whole stretch (237^a20–5). But clearly Aristotle must be assuming here that in a change from state *A* to state *B* there will not only be a first instant of being in state *B* (as has indeed been claimed) but also a last instant of being in state *A*. If this is granted, and assuming (as we may) that nothing can be simultaneously both in state *A* and in state *B*, then indeed Aristotle's conclusion follows: these two instants must be different, and hence there must be a stretch of time separating them. But why should one assume that both these instants must exist? On the contrary, if (as has been claimed) there is a first instant of being in state *B*, then at all instants previous to this the thing is not in state *B*, and we are free to suppose that it is instead in state *A*. But there is no temporal gap between a given instant and all the instants previous to it.[17]

Hoping to clarify things further, Aristotle goes on to say that his claim is clearest when we consider change of position, since the distance between one position and another cannot be indivisible (237^a28–35). And no doubt his claim is in this case perfectly sound (granting the continuity of motion). But he goes on to add that the same applies to *all* changes, even changes between opposites (ἐνάντια) and between contradictories (τὰ ἐν ἀντιφάσει) (237^a35–$^b2$). However it is perfectly clear that the same does not apply in these cases. There is no analogue of the 'distance' to be traversed between two contradictory states, and even though there may be an analogue in the case of opposites—e.g. the 'distance' between black and white that is constituted by the various shades of grey, or the 'distance' between low and high that consists of the various intervening tones—still a thing may surely 'leap' from one point on the scale to another without having to run through all the intermediate points. Aristotle's justification for his claim in these cases is that though we may not always be able to divide the 'distance' between the termini of such a change, still we can always divide the time taken (237^b2–3). But this, of course, begs the question at issue. For in these changes between opposites there need not be any time taken, and certainly in a change between contradictories there cannot be. Let us consider the implications of this latter point.

[17] There is a similar mistake later in ch. 8 of Book VIII (262^a28–$^b3$), where Aristotle argues that if we can say, of a thing in motion, that at some instant it has reached a given point, and later has left it, then we must suppose that there is a *first* instant at which the thing has left the point. (From this he infers that since the instant of reaching the point and the supposed instant of leaving it cannot be the same, they must be separated by a temporal gap during which the thing remains at the point. Consequently, if the thing simply passes the point without pausing at it, there is no instant at which it has reached the point. This, incidentally, is in support of his view in Book VIII (in conformity with the discussion of infinity in Book III) that merely passing a point does not 'actualize' it. By contrast, Book VI sees no problem with the idea that there is an instant at which the thing is level with the point (239^a33–$^b4$).)

It is a plausible principle that at any instant at which a thing *is changing* from being in state *A* to being in state *B*, it is neither (still) in state *A* nor (yet) in state *B*. But if *A* and *B* are contradictories, then at every instant a thing is either in state *A* or in state *B*, and at no instant is it in both states. It then follows that there is no instant at which the thing is changing. Thus, to use Aristotle's own example, if a thing becomes white, then there are instants at which it is (still) not white, and after all of these there are instants at which it is (already) white, but there is no instant 'in between' at which it is neither white nor not white, and therefore no instant at which it *is becoming white*. Alternatively, if we insist that there must be at least one instant at which it is changing, then this must either be the same instant as the first instant at which it is (already) white—if, as Aristotle affirms, there is such a first instant—or otherwise it must be the same as the last instant at which it is (still) not white, if instead there is such a last instant. So in the first case we must say that at the same instant a thing may both be in state *B* and be changing to it (a thesis that Aristotle explicitly denies, e.g. at 235^b25–6), and in the second that at the same instant a thing may both be in state *A* and be changing from it (a thesis which he also denies, e.g. at 234^b10–15 and at 240^b28–30). So the truth is that change between contradictories simply cannot be made to fit in with the claims that Aristotle here makes.[18]

It is, to say the least, somewhat surprising that Aristotle says nothing at all about this problem. But perhaps one can conjecture his answer to it by looking more closely at what he says in a passage towards the end of chapter 9, where he explicitly claims that change between contradictories presents no difficulty for his account (240^a19–29). He implies, it seems, that it would be a difficulty if what was changing from being not white to being white were in neither state, but he goes on to claim that he is not committed to saying this. For a thing need not be wholly white, or wholly not white, before it can be said to be white or not white respectively: we call a thing white so long as the majority of its parts, or its most important parts, are white (240^a20–5). The same applies, he concludes, to other changes between contradictories: 'for the thing will always be in one or other of the opposing states, but it will not always be in either of them as a whole' (240^a26–9). Now on the face of it this does not resolve our problem, for it reaffirms that at any

[18] It strikes me as somewhat ironical that Aristotle cites change between contradictories as the *clearest* support for his contention that there always is a first instant at which a thing has completed its change (235^b16–30).

instant the thing either is white or is not white, and so leaves no time for it to be changing from the one to the other. But at the same time the passage strongly suggests—without quite affirming it—that the change in question is 'really' between contraries and not between contradictories, namely between the initial state of being *wholly* not white, and the final state of being *wholly* white. There will be room for a whole stretch of time for this change to occupy, if—as is clearly envisaged here—the change spreads bit by bit over the surface in question (cf. 236^b6–8, 230^b32–231^a1).

To generalize, then, Aristotle's thought will be that all change can be seen as taking time, so long as we take care to describe the *termini* of the change appropriately. We may loosely describe a change as being from white to not white, or vice versa, and there is admittedly no state intermediate between these. But if we specify the termini precisely—e.g. as wholly white and wholly some other (specified) colour—then there will certainly be room for intermediate states, and hence for a time taken in passing through them. We may also add that care must be taken in specifying the termini of a change if we are to maintain the earlier doctrine that, in a change from state *A* to state *B* which does take time, there will always be a last instant of being in state *A* and a first instant of being in state *B*. For example, suppose I move an object so that its centre of gravity changes from being to the left of one end of a ruler to being to the right of the other end of it. Then undoubtedly this change takes time. But presumably there is no last instant at which the centre of gravity is to the left of the one point, nor any first instant in which it is to the right of the other. (Rather, there is a first (and only) instant at which the centre of gravity coincides with the one point, and a last (and only) instant at which it coincides with the other.) Hence to maintain the doctrine in a case such as this we must insist that the termini of the change be more precisely specified: it is not good enough to say that at the beginning the centre of gravity is 'to the left of' some given point, but we must say more precisely how far to the left it is (or must specify its precise position in some other definite way). Now it seems to me quite plausible to hold that in any change which does take (a finite[19]) time it will be possible to specify the initial state in such a way that there is a

[19] A change which takes an infinite time presumably will not have both an initial state and a final state, and the doctrine that for every change there is a first instant at which it is completed needs modifying to allow for this. (The eternal rotation of the heavens would be an example; cf. the obscure remarks on this at 265^a27–$^b8$.)

last instant of being in that state, and to specify the final state in such a way that there is a first instant of being in it. So in this revised form the doctrine that there always are both a last instant of the initial state and a first instant of the final state may perhaps be accepted, but only—of course—for changes which genuinely do take time. But we still have not been given any reason to suppose that all changes genuinely do take time, and hence that if a thing *has* changed then there must have been a time during which it *was* changing.

As a matter of fact this claim seems to be abandoned later in the *Physics*, in chapter 3 of Book VIII. For it is there claimed that a thing may change size by losing a whole particle of matter all at once—and though the particle is divisible it does not follow that it was detached bit by bit[20]—and similarly any change of quality (such as freezing) may happen all at once (253^b14–26; cf. also 186^a15–16). On the other hand Aristotle does not there, or anywhere else in the *Physics*, draw the moral that a thing may have changed without ever having been in the process of changing. It is a point that he insists upon elsewhere, frequently in the *Metaphysics*,[21] but it does not seem to appear in the *Physics* itself. It seems *almost* to be recognized in the coda to this same chapter 6 of Book VI, where Aristotle turns to apply his doctrine to coming into being and passing away. As one might expect, he claims there that if anything has come into being then at an earlier time it was coming into being, and that at any instant during that earlier time some part of it had already come into being (237^b9–22). But this of course requires that the thing coming into being has parts, and indeed is infinitely divisible, and Aristotle does preface his claim with a restriction to such things (237^b11). However he does not here say anything about the coming into being of a thing with no parts (e.g. a point), though elsewhere it is one of his favourite examples of what has changed without ever being in the process of change. On the contrary, the reader of Book VI of the *Physics* will be more likely to suppose that the restriction is otiose, for Aristotle has already urged, and will urge again, that *everything* that changes in any way must have parts. Let us take up this claim, as our final topic from Book VI.

[20] Contrast 236^b16–17.

[21] e.g. *Met. B*. 5, 1002^a28–$^b11$; *Z*. 15, 1039^b20–7; *H*. 5, 1044^b21–9; *Λ*. 3, 1070^a15–17. Cf. *Cael.* 280^b25–8, *EN* 1174^b10–13.

V Whatever Changes has Parts

The claim is argued first at the beginning of chapter 4, and the reasoning is this. When a thing changes, it changes from being in one state to being in another, say from being in state *A* to being in state *B*. Then when the thing is in state *B* its change is completed, and it is no longer changing. Equally, when it and all its parts are in state *A*, it is not yet changing. Assuming, then, that there must be a time during which it is changing—and we have just considered the argument for this assumption—Aristotle infers that during that time part of it must be in state *A* and part in state *B*, 'since it cannot be in both of the states, nor in neither of them' (234^b10–17). But it is clear that this is wholly illegitimate reasoning; if I am to change from being in Athens to being in Thebes, it is by no means necessary that during all the time that this change is going on part of me must be still in Athens and part of me must be already in Thebes. Apparently in an attempt to avoid this, and similar, objections, Aristotle proceeds to add that the state *B* is to be taken as 'that into which the thing *first* changes' in the course of its whole change. For example, if the whole change is from being white to being black, then, 'that into which the thing first changes' is being grey (234^b17–20). But here of course we may object that there need not be any such 'first' state, and there cannot be if, as with motion, the states intermediate between *A* and *B* form a continuum. Indeed, Aristotle himself is going to argue for that very point in the next two chapters. So this argument at the beginning of chapter 4 must be dismissed as worthless. (Nevertheless Aristotle proceeds to elaborate its consequences for the rest of chapter 4—often somewhat obscurely—and ends the chapter by suggesting that it is the divisibility of the changing object that is the foundation for the divisibility of time, space, and movement (235^b1–4).)

The divisibility of the changing object is assumed several times in what follows (e.g. 236^a27–35, 240^a22–9), and Aristotle comes to argue for it once more at the end of the book in chapter 10. This time he begins with a qualification. A thing without parts, he concedes, may be said to change *per accidens*, in so far as some whole that contains it is changing (240^b8–13). (And the change of the whole, he adds, need not be the same as that of its parts, illustrating this by contrasting the changes of the parts of a rotating sphere, those near the centre moving more slowly than those near the circumference (240^b13–20).) But a thing without parts, he claims, cannot change *per se*, and at first he appears simply to repeat the argument of chapter 4. For suppose such a thing does change, its initial state being *AB* and its final state *BC*. Then during the time when it is changing it must, he says, be either in state *AB* or in state *BC* or partly in one and partly in the other. But neither the first nor the second alternative can apply while the thing is changing, and the third cannot apply either if the thing has no parts. Hence a thing with no parts cannot move, or change in any other way (240^b20–31). So far this appears to be just the argument that we have already criticized, though it may be noted that the symbolism '*AB*' and '*BC*' is evidently intended to suggest states that are in some way 'adjacent'.

A little more light is shed on the issue by Aristotle's subsequent remarks. A thing without parts could move, he says, only if time were made up of instants; for then there would be no time when it *is* moving, but only times when it *has* moved. But this has already been shown to be impossible (240^b31–241^a6). Moreover, he adds, before anything can move a distance greater than itself it must first move a distance less than or equal to itself. But such a thing as a point, being indivisible, cannot move a distance less than itself; so, if it can move at all, it must be able to move a distance equal to itself, and that in turn implies that a line may be made up of points; for the one point, by moving first one and then another distance equal to itself, will eventually measure out the whole line. But we have already shown that that is impossible (241^a6–14). Elaborating these thoughts further, Aristotle adds that it must in fact always be possible for anything that moves to move a distance less than itself, so clearly a point cannot move. What is indivisible and without parts, he concludes, could move only if it were possible to move in an indivisible instant, 'for movement in an instant, and the movement of what is indivisible, depend on the same reasoning' (241^a15–26).

One can see from this what position Aristotle considers himself to be controverting. When he considers the movement of an 'indivisible thing', or a 'thing without parts', he nevertheless conceives that thing as still having some size. That is why he naturally uses the symbol '*AB*', which represents an *interval*, to stand in for its initial position, and '*BC*', which represents an adjacent interval of the same length, to stand in for its 'next' position. It is also why he simply takes it for granted that we can meaningfully speak of a distance greater than, equal to, or less than the thing itself, i.e. a stretch comparable to the stretch it occupies itself. Although he calls such a thing a point, it is perfectly clear that his argument does not in fact show that a genuine point, which occupies *no* spatial distance, is incapable of moving. At best it shows that if an extended thing can move (in the direction in which it is extended), and if motion is continuous rather than 'cinematographic', then that thing must be divisible, at least in thought. But if we consider a genuinely unextended thing, such as the point at the tip of an arrow, it is perfectly evident that this can move, even though it has no parts. (Similarly, a surface of a cube can move in a direction in which it is not extended.)

Now Aristotle no doubt thinks that he has a reply to this objection, in that he has already conceded that a thing without parts may move *per accidens*, in so far as it belongs to some whole which moves *per se*. But first, this concession in no way improves his argument, for that argument did not anywhere rely on the premiss that only motion *per se* is in question. (Hence, if the argument were successful against motion *per se*, it would also be successful against motion *per accidens*, which clearly it cannot be.) Second, the point that he has in mind seems in fact to be a point about the identification of things without parts, and not about their motion. No doubt it is plausible to say that we can only identify the point of the arrow as the *same* thing, now in one place and now in another, if we identify it by reference to the whole arrow (or at least, some extended part of the arrow) of which it is a point. (This, one might suggest, is because we identify things in the first place by what we can perceive, and unextended things do not, 'in their *own* right', have any perceptible properties.[22]) But then the correct conclusion to draw seems to be that an unextended thing cannot be identified *per se*, though there would be no difficulty in supposing it to move *per se*. For there is not, as Aristotle apparently claims, anything in the concept of motion that requires a moving thing to have parts. (Thus if, for example, the material world were wholly made up of unextended centres of force—as Boscovich suggested—there would be no conceptual impediment to their moving.)

[22] But is 'being now' a perceptible property of an instant? And can we identify 'one and the same now', as it 'moves' through time, without first identifying anything else? (Contrast 219^{b}15–33.)

VI Concluding Remarks

We have seen that quite a number of the claims that Aristotle makes in Book VI of the *Physics* are mistaken. Is there anything that can be said in explanation of these mistakes? No doubt it is sometimes right just to say that he was being careless, as with the error with which the book opens, that he fails even to state the definition of a continuum on which his first argument relies. But this seems an implausible view about most of the theses that we have disputed, for generally Aristotle offers quite extended arguments in their favour.

One suggestion might be that for much of the time he is responding, somewhat over-hastily, to issues raised by Plato in the second part of his *Parmenides*.[23] For example, the argument purporting to show that an indivisible thing cannot move is taken straight from the *Parmenides* (at 138d1–e7), though of course this does not make it any better. More significantly, Plato's main contribution in the *Parmenides* was certainly to point to the puzzling nature of an instantaneous change, and perhaps Aristotle's claim that no change is instantaneous is mainly an attempt to avoid the puzzles that Plato had revealed. Some support for this conjecture may perhaps be drawn from an earlier passage in chapter 2 of Book V, where again it is very tempting to see the influence of the *Parmenides*.

Plato had suggested that if we consider a change say from rest to motion, or from motion to rest, we can only say *when* this change occurs by positing a paradoxical 'sudden', not itself in time at all, for it to occur in. For only 'in a sudden' can a thing be neither in motion nor at rest (156c1–e7). He had then gone on to generalize this result to all other changes, and had included here not only the change from being *F* to being not-*F* but also the changes from being *F* to *becoming* not-*F*, and from there to being not-*F*. These too, he had suggested, must occur 'in a sudden', for only 'in a sudden' can a thing be neither (still) *F* nor (already) *becoming* not-*F*, and neither (still) becoming not-*F* nor yet (already) not-*F* (156e7–157b3). He thus raises a question not only about the instantaneous change from being *F* to being not-*F*, but also about the equally instantaneous change from being *F* to *changing* from being *F* to being not-*F*, and clearly there is the potential here for an infinite regress. Aristotle's response, then, is to deny in chapter 6 of

[23] I have discussed Plato's contribution in my 'Plato on Change and Time in the *Parmenides*', *Phronesis* 23 (1978), 229–42.

Book VI that the change from being *F* to being not-*F* is instantaneous, and to deny in chapter 2 of Book V that there is such a thing as the change from being *F* to being in the process of becoming not-*F*. (For clearly if there were such a change it would have to be instantaneous.) In the course of this latter response (225^b13–226^a23) he shows himself well aware of the infinite regress that would be involved (225^b33–226^a6).

(Unfortunately, to support his position in chapter 2 of Book V Aristotle falls back upon the narrow ontology of the *Categories*, according to which changes are not substances, but only substances can be the subjects of accidental predications, and hence only substances can undergo change. This leads him to generalize his response to the somewhat astounding claim that there is *no* kind of change of change—for example a change cannot increase or decrease—and so to deny any subject-matter to the study of acceleration and deceleration (225^b13–21). But the mistaken claims of Book VI do not seem to rely on any such narrow ontology; for example, there would appear to be nothing in Book VI that would conflict with the doctrine of Book IV that change is indeed the subject of which time is predicated (contrary to *Cat.* 2^a34–$^b6$).[24] At most there is a hint of the preference for substances over other things when Aristotle suggests—without any evident warrant—that it is the divisibility of the changing object that is the foundation for the divisibility of space and time (235^b1–3).)

But if it is Plato's *Parmenides* that provides the motivation for Aristotle's claim that every change takes time, one would certainly expect him to have worked out the consequences of this claim rather more carefully. In particular, one would expect him to have explained in more detail how change between contradictories could be harmonized with his thesis, for at first sight it certainly is a counter-example, and indeed the counter-example from which Plato's own argument begins. And although we have in fact been able to suggest a possible reconstruction of Aristotle's thought upon this issue, we have had to read some way between his lines to do so, which makes it altogether less probable that his account is meant as an explicit reply to Plato. It appears, rather, that he simply has not grasped the problem posed by change between contradictories. (And we may

[24] Yet Book IV would appear to claim that an instantaneous change is the subject of which the now is predicated, as I have argued in my 'Aristotle's Account of Time', *Phronesis* 25 (1980), 148–69. If so, it is of course in conflict with Book VI, which denies the existence of instantaneous changes.

note that when elsewhere he rejects the thesis that all change takes time, the counter-example that he has in mind seems not to be just any old change between contradictories, but the very special case of a change from not existing to existing, or vice versa, in the case of things that have no parts.)

Somewhat similarly, one might suggest that Aristotle's claim that there is no sense to talk of motion or rest at an instant is intended as a response to Plato's puzzle, for that puzzle begins by asking how there can be an instant at which a thing is neither in motion nor at rest, but is changing between them. But Plato generalizes his puzzle to all other changes, mentioning amongst others change of size, and change between existence and non-existence (156e7–157b3). So we may similarly ask how there can be an instant at which a thing neither exists nor does not exist, neither is of this size nor is not of this size. Aristotle, it would appear, should again be replying that there is no sense to talk of existence or non-existence at an instant, or of being a particular size just at one instant, but evidently he says no such thing. He denies that a thing either is changing at an instant or is not changing at it (counting moving as a kind of changing), but he makes no attempt to extend this to other apparent contradictories between which change may occur. So again it seems probable that this claim of his is not a response to Plato's puzzle, for if it were it would not be nearly general enough.[25]

Another possible suggestion, to explain Aristotle's denial of motion or rest at an instant, might be that he thought it was needed to resolve Zeno's paradox of the flying arrow. Certainly, chapter 8 concentrates upon the implications of the claim that nothing rests at an instant, and this then leads straight into a refutation of the arrow paradox at the beginning of chapter 9. But first, it would seem to be a somewhat exaggerated response to that paradox, and second, it appears to be only part of Aristotle's own response. Zeno apparently had argued that an arrow in its flight is always 'in some instant', and that in each instant it is 'over against what is equal to itself', from which he had inferred that in each instant it is at rest (239^b5–7). Now all that is needed by way of response to this is to point out that the inference is invalid. One can agree that at each instant a thing is always 'over against what is equal to itself' (which I take to be a way of saying that it occupies

[25] One might indeed wonder whether my presumed chronology is correct. Perhaps Plato's *Parmenides* replies to Aristotle's *Phys.* VI, rather than vice versa?

exactly the position that it does occupy), while denying that it follows from this that it is at rest (as Aristotle has just done, at 239^{a}35–^{b}1). And surely it would not be too difficult to explain this denial while still leaving open the possibility of saying that at each instant of its flight the arrow is in motion. One is not forced to say that nothing at all counts as being in motion or at rest at an instant, which is what Aristotle gives as his reason at 239^{b}1–2; and it is at first somewhat difficult to see why he should add, as a further diagnosis, that it is false to suppose that time is made up of indivisible instants (239^{b}8–9). What has that to do with it?

It may be suggested that by this last remark Aristotle means to indicate that in any case the inference from 'the arrow rests at each instant' to 'the arrow rests throughout the period' is invalid. His point is that it is invalid *because* time is not made up of instants. (For even though Aristotle has already rejected the premiss, he still needs to reject the inference too. At any rate, he himself accepts 'the arrow covers no distance in any instant', but will not infer 'the arrow covers no distance throughout the period'.) The trouble with this suggestion is, of course, that the inference would still be invalid even if time *were* made up of instants, as we now believe that it is, and as the time-atomist also believes, but in a different way. But this leads to a better suggestion, for if we construe an 'instant' as a very short *period*, as the time-atomist does, then both the premiss and the inference may indeed seem tempting. At any rate, both we and Aristotle would accept as correct 'if the arrow covers no distance in any millisecond of this minute, then it covers no distance during the whole minute'. Similarly, both we and Aristotle would agree that if the arrow were at the same spatial position throughout a whole millisecond then it could properly be said to be at rest then. I suggest, then, that Aristotle thought that the plausibility of this paradox depends upon the time-atomist's way of construing an instant as a short period, and that is why his diagnosis is that it is a mistake to suppose that time is made up of instants. He means that time-atomism is a mistake, and he takes this thesis to *be* a characterization of time-atomism.

Now in fact the consistent time-atomist should accept that the arrow covers no distance in any time-atom, though he is not bound to conclude that it therefore counts as being 'at rest' then. But in any case he should certainly reject the suggested inference, since it depends on the assumption that motion is not 'cinematographic'. Thus the paradox appears to have some bite only if one is inconsistently combining some features of the theory that time is atomic with some features of Aristotle's own theory that time is a continuum. But the significant point that emerges is that, from Aristotle's point of view,[26] it is time-atomism that lies behind this paradox. And I would suggest that it is not this paradox in particular but time-atomism in general that leads him to deny instantaneous motion or rest. In fact it is plausible to say that atomism, in one form or another, is Aristotle's main target in this book of the *Physics*. At any rate, it is certainly his target in all of chapters 1–4, and this consideration will be enough to yield a fairly plausible explanation of all our difficulties.

We may start by repeating the point just made, that Aristotle regards the thesis that time is made up of instants as characteristic of time-atomism. Similarly, when the book opens with a refutation of the claim that a line is made up of points, Aristotle regards this claim as characteristic of space-atomism. For who, in his day, had claimed that lines *were* made up of points? Not, presumably, some unknown Cantor or Dedekind, with a theory about how it really was possible to construct a line from genuinely unextended points, and who had shown how, if you take *enough* things of no magnitude you thereby obtain a positive magnitude. No. The thesis can only have seemed attractive to those who held—e.g. for reasons such as Aristotle gives at *GC* 315^b24–316^b34—that points must after all be conceived as having magnitude, and yet as being indivisible. We would more naturally say that such things should not be called points at all (but, say, 'atoms of extension'), and Aristotle would no doubt agree. Yet as we have seen he is prepared to call them points himself, in the course of elaborating a different argument against them at the end of chapter 10 (241^a6–26). It is this theory, then, that he takes himself to be refuting at the beginning of chapter 1. His argument is, in effect, that if these things really are indivisible, then they can only be points as we and he conceive points to be. And it is hardly surprising that he cannot see how *genuine* points, which cannot touch one another, could be put together, to form a gapless continuum.

The rest of chapter 1 is then concerned to maintain the continuity of (uniform, rectilinear) motion, and on that ground to deny the atomicity of space and of time. This argument continues also throughout chapter 2. It is a very nice argument, given the premiss about the

[26] Whether this was also Zeno's view is another question.

continuity of motion, but that premiss itself leads to a problem. For what argument can Aristotle use against the rival 'cinematographic' view of motion? The truth is that there is no good argument, and the only argument that he can come up with is this: we ordinarily suppose that if a thing *has* moved to a certain place, then at some earlier time it *was* moving to that place, whereas it can fairly be said that the 'cinematographic' view of motion denies this. So Aristotle has to elevate this 'ordinary supposition' of ours to an undeniable principle. But now, would there be any way of explaining, without begging the question, why this principle should hold of motion from place to place but should *not* hold of other changes? Certainly, it is difficult to think of one, and it would appear that we do ordinarily make this same supposition whatever kind of change is in question. Thus Aristotle is apparently committed to saying that in *all* cases 'has changed' implies 'was changing', and this of course is exactly what he does say in chapter 6. It is, then, his way of arguing against the atomicity of space and time in chapter 1 that commits him to his claim in chapters 5 and 6 that *all* change takes time. (He may of course also have been influenced by the view which he argues later in chapter 7 of Book VIII, that all other kinds of change depend upon motion from place to place.) But I think one can only conclude that he has not really seen the difficulties that this claim involves.[27]

Anyway, to return to our thread, chapters 1 and 2 clearly argue against 'atomist' theories of space and time, and it is evidently a continuation of the same theme when Aristotle proceeds, in chapter 3, to give further reasons for saying that 'a now' has no temporal extension. It is still, I think, part of the same programme when he proceeds to infer from this that movement and rest cannot occur 'in an instant.' For, as he conceives it, the idea that they can is an atomist idea; it is part of the idea that movement proceeds 'cinematographically'. That is why, when he uses this thesis in his reply to Zeno's arrow, he presumes that Zeno must be basing his argument on an atomist premiss. This conjecture would also go some way towards explaining an otherwise puzzling passage in chapter 5, where Aristotle appears at first sight to be rejecting his own claim that nothing can be said to be at rest at an instant.

At this point Aristotle is opening his argument for the claim that there is no first time in which a thing has changed, that is, has begun

[27] And hence that he was *not* familiar with *Parmenides* 156–7?

its change, or has changed at least a bit. For suppose, he says, *AD* is such a time. Then *AD* will be divisible into an earlier and a later part, and clearly the earlier part will equally be a time in which the thing has begun its change, or has changed at least a bit. But before he draws his conclusion, he pauses to consider the rival claim that *AD* is not divisible, and is therefore a time-atom. As one objection to this, he says that it will follow that nows (i.e. time-atoms) will be next to one another, whereas he has already argued that nows (genuine instants) cannot be ($236^{a}15$–17). But he also offers a second objection, which is puzzling. We may assume that the thing was at rest throughout a preceding period *CA*. But then, he goes on, 'it was at rest at *A*; so if *AD* is without parts it will simultaneously be at rest and will have changed, for it was at rest at *A*, but has changed at *D*' ($236^{a}17$–20). What is puzzling about this is that Aristotle appears to be assuming, as a *premiss* to his objection, that one can properly speak of the thing as being at rest at the instant *A*, although this premiss denies his own doctrine. So the explanation can only be (as suggested by Sorabji)[28] that he is not asserting this premiss in his own person, but is taking it as one that his opponent, the time-atomist, would have to grant. But why so? Well, perhaps the answer is this. The time-atomist supposes that during a now (i.e. a time-atom) everything remains unchanged. So in order to express his thesis at all, *he* must suppose that it makes sense to say that a thing is at rest in a now (i.e. a time-atom), and Aristotle—having 'proved' that a now is not extended—takes this to imply that the opponent is committed to saying that a thing may be at rest in a genuine instant. Thus his argument is that the time-atomist could be forced to accept that we can meaningfully ask whether a thing is or is not at rest *at* the instant which separates one supposed time-atom from the next, and he has already claimed that this question cannot be answered when the instant in question is one that separates a period of rest from a period of motion ($234^{a}34$–$^{b}5$).

Needless to say, this is not a good argument against time-atomism, for the time-atomist may perfectly well refuse to say what state anything is in *at* the instant that separates one time-atom from the next, and anyway the argument confuses the indivisibility that applies to a time-atom (which does not exclude divisibility 'in thought') with the indivisibility of a genuine instant. But, if this reconstruction is right, then we may note two ways in which time-atomism may be taken to

[28] See n. 16 above.

require motion and rest in an instant. It requires that a thing may *have moved*, in a genuine instant, from place *A* to place *B*, for that is how cinematographic motion occurs. It also requires that things *have remained at rest* during each time-atom, for that is precisely what time-atomism claims, and Aristotle takes this to imply that a thing has remained at rest in a genuine instant. These, then, are the theses that Aristotle really wishes to deny.

But if this is right, then we can only say that he has been taken in by an ambiguity, an ambiguity between 'nothing moves in an instant' and 'nothing is moving at an instant', for indeed the Greek for both is exactly the same: *οὐθὲν κινεῖται ἐν τῷ νῦν*. (Greek does not distinguish the so-called continuous present tense 'is moving' from the plain present tense 'moves'; and it standardly uses 'in an instant' where we would say 'at an instant'.) Undoubtedly, Aristotle's first argument for this claim is only an argument for the first version of it, for it professes to deduce a contradiction from the supposition that a thing *has* moved from *A* to *B* during one indivisible instant (234^a26–31). And there is no recognition that the sense of the claim has shifted in the following argument (234^a34–$^b5$), where however it is much more natural to suppose that what is in question is whether a thing can be said to *be* moving *at* an instant. At any rate, it is clear that he does employ the claim in this latter sense (e.g. at 237^a11–15 and 239^a10–14), though he also thinks, as we have seen, that only an atomist would wish to deny it. So there is some confusion in his thought here, and the only version of the claim that he actually needs is the less controversial one: nothing can be said either to move from *A* to *B*, or to remain at *A*, in an instant.

Finally, chapter 4, with its claim that a moving *thing* must always be divisible, is all part and parcel of this same attack on atomism, though the attack has moved now from atomism about space and time to the more familiar atomism about matter. (Though this may not be obvious from the way that Aristotle presents his reasoning in chapter 4, it is obvious from his further elaboration in chapter 10.) He wishes to claim that an atomic chunk of matter, if it really is (as its proponents say) not divisible into parts, cannot move at all. In one way he is right: if an atom is an extended chunk of matter, and if motion is continuous and not cinematographic, then it must be possible for a moving atom to be partly in and partly not in the place it occupied a little earlier. And in this sense it must have parts. But (i) it was a mistake on Aristotle's part to let this claim depend upon the hypothesis that the atom moves, for an extended atom must still have spatial parts even if it never moves, and there is nothing in the concept of motion that requires moving things to be extended; and (ii) Aristotle's objection anyway misses the point. Just as an atomist concerning space or time will conceive of his spatial or temporal atoms as extended, so that they can always be regarded as divisible *in thought*, so also one who believes in atoms of matter will equally regard them as extended, and should therefore admit that they are equally divisible *in thought*. His claim is rather that they are never divided in fact, that nothing that you can do to an atom will ever separate it into parts that move independently of one another. But against this theory Aristotle never offers any objection that carries any conviction at all.

To sum up the situation a little crudely, Aristotle has three kinds of objection to atomism about matter. One is that the Greek atomists thought of atoms as differing from one another only in size and shape, and so attempted to explain all the various properties of material objects by drawing just on these properties of their constituent atoms. Aristotle very fairly objects that such a reduction cannot be made to work (e.g. at *Cael.* III. 4, 303^a3–$^b3$, and III. 8, 306^b29–307^b24). But clearly this is not an objection of principle, since it could apparently be met by endowing atoms with further properties. His second objection is that the atomist hypothesis must also involve positing the existence of empty space, and this he thinks he can refute independently (*Phys.* IV. 6–9). But his third objection, which he appears to think of as the strongest, is what he calls the conflict with mathematics, and it is this objection that is developed in *Phys.* VI. (Cf. *Cael.* III. 4, 303^a20–4, and III. 7, 306^a26–$^b2$.) But, as we have seen, it misses the point altogether. One can be an atomist concerning space, time, or matter, and still admit the *mathematical* divisibility of one's atoms.

9

Aristotle's Mathematical Physics: A Reconstruction

EDWARD HUSSEY

I Introduction

Aristotle was not the first to suggest that the physical world exhibits mathematical relationships of various kinds. This thought is at home both in the 'Milesian' tradition of natural philosophy, and in that of the 'Pythagoreans'.[1] It is even arguable that Democritus had envisaged a thoroughly mathematical physics.[2] In any case, there existed, by the time of Aristotle, some specialized mathematical sciences, not dependent on any particular cosmology or metaphysics, but having uncontested and successful application in the actual world. These were (two- and three-dimensional) geometry, geometrical optics, harmonics (mathematical theory of music), and particularly mathematical astronomy.[3] The most mathematically developed of these sciences was astronomy, in the hands of Aristotle's contemporaries Eudoxus and Callippus. The mechanical-mathematical model developed by Eudoxus to account for the celestial motions was, I shall suggest, a source of stimulus and inspiration for Aristotle in more than one way.[4]

[1] The best short general treatment of this complex subject is S. Sambursky, *The Physical World of the Greeks* (London, 1956), chs. 1–3.

[2] An ingenious attempt to exhibit Democritus as a pioneer in this field is made by R. Löbl, *Demokrits Atome* (Bonn, 1976).

[3] On early Greek geometry and its applications: T. L. Heath, *A History of Greek Mathematics* (Oxford, 1960), vol. 1, chs. 5–7; O. Neugebauer, *The Exact Sciences in Antiquity* (New York, 1962), ch. 6 and Appendix 2; M. R. Cohen and I. E. Drabkin, *A Source Book in Greek Science* (Cambridge, Mass., 1958), esp. 33–75, 90–105. The richest single source is the writings of Aristotle himself, and the passages are collected with comment in T. L. Heath, *Mathematics in Aristotle* (Oxford, 1949). For early musical theory, see also A. Barker, *Greek Musical Writings ii: Harmonic and Acoustic Theory* (Cambridge, 1989).

[4] On Eudoxus and Callippus, and Aristotle's use of their work, see Neugebauer, *Exact Sciences*, 152–6; T. L. Heath, *Aristarchus of Samos* (Oxford, 1913, reprinted New York, 1981), chs. 16 and 17; D. R. Dicks, *Early Greek Astronomy to Aristotle* (London, 1970) chs. 6 and 7.

It is not in dispute that for Aristotle mathematics was a subject of the deepest interest. It was the prime example of an incontestably successful science having application to the actual world. Aristotle has a striking philosophical theory of the nature of mathematics and its objects.[5] He constantly calls on mathematics as a provider of examples and paradigms.[6] But it has been controversial how far, and in what ways, mathematics enters into Aristotle's substantive physical theories. In this chapter I offer what is intended to be the outline of a coherent and credible answer. In sections II and III, I consider the mathematical theory of change in general. Sections IV to VII deal with the tangled and much-discussed problems of the theory of motion in particular. Section VIII offers some general reflections in conclusion.[7]

I shall concentrate on Aristotle as a working physicist, and shall try to show that in that role Aristotle is as impressive, and as significant, as he is as working biologist. The enterprise is hampered by the lack of any works in the corpus which are to physics what the *Historia Animalium*, *Parts of Animals*, and *Generation of Animals* are to biology. We are forced to rely on evidence which emerges piecemeal and incidentally, without proper explanation.

[5] On Aristotle's philosophy of mathematics in general: I. Mueller, 'Aristotle on Geometrical Objects', *Archiv für Geschichte der Philosophie* 52 (1970), 156–71 (also in J. Barnes, M. Schofield and R. Sorabji (eds.), *Articles on Aristotle* 4 vols. (London, 1975–9), iii. 96–107); E. Hussey, *Aristotle's Physics Books III and IV* (Oxford, 1983), 176–84; (a revised and expanded version is to appear as 'Aristotle on Mathematical Objects' in a collection of essays on Greek mathematics (ed. Ian Mueller) to be published as a special number of the journal *Apeiron*); J. Lear, 'Aristotle's Philosophy of Mathematics', *Philosophical Review* 91 (1982), 161–92; J. Annas, 'Die Gegenstände der Mathematik bei Aristoteles', in A. Graeser (ed.), *Mathematics and Metaphysics in Aristotle* (Bern, 1987) 131–47.

[6] 'It was the mathematics he [Aristotle] encountered [in the Academy] that impressed him as providing the model for any well-organized science' (G. E. L. Owen, *Logic, Science and Dialectic: Collected Papers in Greek Philosophy*, ed. M. Nussbaum (London, 1986), 153). Most of the references to mathematics in the Aristotelian corpus are collected in Heath, *Mathematics in Aristotle*.

[7] This chapter constitutes a working-out (but with some radical revisions) of some of the ideas outlined in 'Additional Note B' of my *Physics III and IV*, 185–200. I am most grateful to Lindsay Judson for his constructive comments. Some works which are relevant throughout are: H. Carteron, *La notion de force dans le système d'Aristote* (reprinted New York and London, 1979), of which part appears in English translation in Barnes, Schofield, and Sorabji, *Articles On Aristotle*, i. 161–74; I. E. Drabkin, 'Notes on the Laws of Motion in Aristotle', *American Journal of Philology* 59 (1938), 60–84; Owen, *Logic, Science and Dialectic*, ch. 8 ('Aristotle: Method, Physics and Cosmology'), and ch. 18 ('Aristotelian Mechanics'); G. E. R. Lloyd, *The Revolutions of Wisdom* (Berkeley, Los Angeles, and London, 1987), esp. ch. 5; R. Sorabji, *Matter, Space and Motion: Theories in Antiquity and their Sequel* (London, 1988); R. Wardy, *The Chain of Change: A Study of Aristotle's Physics VII* (Cambridge, 1990).

II The Mathematical Theory of Physical Change: (i) Proportions and Powers

In speaking of physical change, Aristotle often mentions the 'power' ($\delta\acute{v}\nu\alpha\mu\iota\varsigma$) or 'strength' ($i\sigma\chi\acute{v}\varsigma$) of the agent of change.[8] It is fairly clear that he uses these terms interchangeably. And 'power' is, for Aristotle, a measurable quantity. This results from some places where proportions are said to hold, other things being equal, between various quantities involved in changes.

The quantities involved are four: the power of the agent, the time of the change, the amount of the changing thing, and the amount of the change. At *Phys.* VII. 5, 249^b7–250^a25, Aristotle asserts a number of proportionalities relating these quantities, which may be summed up, for convenience of reference, in a single general formula.[9] For any given type of change, and any system of units, other things being equal, there is a constant determined by the choice of units such that we have:

$$\text{(Power of agent)} \times \text{(Time of change)}$$
$$=$$
$$\text{(Constant)} \times \text{(Amount of changing thing)} \times \text{(Amount of change)}$$

This general formula includes the special cases mentioned by Aristotle. Thus, other things (including the other quantities involved) being supposed constant, (1) the power will vary in inverse proportion to the time; (2) the power will vary in direct proportion to the amount

[8] The deliberate closeness of Aristotle's terminology to that of ordinary speech is stressed by Owen, *Logic, Science and Dialectic*, 156. In many modern discussions and translations, $\delta\acute{v}\nu\alpha\mu\iota\varsigma$ in this sense is rendered by 'force'. This is doubly unfortunate: it misses the connection with 'power' in ordinary speech, and it most misleadingly suggests some kind of analogy with Newtonian force.

[9] This 'formula' is an expository device, not Aristotle's, to avoid tedious repetitions. It is admittedly dangerous to use modern short cuts in this way. Above all, one must not assume that Aristotle would have recognized the 'general formula' as a legitimate way of stating his proportionalities. Apart from anything else, it involves (whether presented in the form of a proportion or not) a relationship between four different types of quantity, and not two. Even the various particular proportionalities it entails are not all stated explicitly, but are appealed to as already known, thus: amount of changing thing inversely proportional to amount of change (250^a1–3); amount of change proportional to time (250^a3–4); amount of power proportional to amount of changing thing (250^a6–7). Nor does Aristotle mention the 'constant' and its dependence on the system of units that happens to be chosen. With these caveats, the 'formula' will be used hereafter as a convenient way of summing up the mathematical content of the proportionalities.

of the changing thing; (3) the power will vary in direct proportion to the amount of the change. It is made clear by 250^{a}9–25 that these relationships are not meant to give sufficient conditions for the actual occurrence or not of changes. So it seems that they are in the nature of necessary conditions: if there is a change of the relevant type, then, other things being equal, and in the absence of external disturbing factors, the quantities will be so related.

It is obvious that the change is here assumed to occur at a uniform rate from start to finish. This is a contrast to the flexibility of the Newtonian way of dealing with change. It also seems at first to be at odds with common sense and ordinary observation. Yet it was certainly Aristotle's standard way of thinking about change. Naturally, he was aware that actual changes hardly ever proceed at a constant rate. But it is reasonable to suppose, in interpreting these and similar passages, that Aristotle is thinking here of the simplest changes, in which there is just one agent, remaining in contact over a time with the changing thing, and free from external disturbance. A constant rate of change is here not implausible, and seems to be roughly in accord with ordinary observation. But it should also be noticed that with due caution the proportionalities allow at least the possibility of being adapted for use in the case of non-uniform change (see section VII).

We now have to ask what is meant, for any particular type of change, by (1) the 'power' of the agent, and (2) the 'amount' of the changing thing.[10] As regards the latter, one would naturally suppose that it is something like weight, not volume, when the thing is moving; but for other types of change, e.g. heating, it might be volume and not weight. Aristotle is under no requirement to establish a single kind of amount relevant to all changes. It is a matter of common sense that the more there is of the thing to be changed, the more time it needs and/or the more input from the agent.

With this we come round to the question of 'power'. Given the proportionality, which makes it proportional to the amount of change but also inversely proportional to the time, it cannot be thought of as the 'total input' from the agent, but rather as the 'rate of input'. This is,

[10] A subsidiary problem: the 'amount of change', as a measurable and continuous magnitude, is immediately intelligible in cases of motion, but puzzling in qualitative change, where there is (according to Aristotle) no continuous spectrum of qualities. This problem complicates Aristotle's attempts to assimilate the structures of spatial magnitude, change, and time: on his attempt to deal with it, see my *Physics III and IV*, 143. On the particular case of colours, see Sorabji's article mentioned in n. 56.

of course, a deliberately rough and ready, crude commonsensical way of thinking about what is going on. But it is no surprise to find it in Aristotle, just because it does correspond to our ordinary ways of thinking about physical agency. And for that very same reason, it is no surprise that it is reproduced almost exactly in Newtonian theory. The 'total input' of the agent is there represented by the quantity called, for obvious reasons, 'work'. And the rate of work is often called the 'power', corresponding thus in terminology as well as in conceptual role with Aristotelian power. The common-sense insight, enshrined in both Aristotelian and Newtonian theory, is this: most familiar agents that produce motion, particularly human beings and other animals, do have a characteristic, roughly constant, rate of producing change. The rate corresponds to what is called the 'power' of the agent in ordinary speech.[11]

Aristotle's proportionalities capture this insight. And, because the insight is correct, this part of Aristotelian theory has, for motion at least, a close equivalent in Newtonian terms, as has just been explained. Where there is a motion in one dimension at a constant rate, under a constant force balanced out by friction, the agent's power (rate of doing work) will indeed be directly proportional to the amount of motion, and to the 'amount of the changing thing', interpreted as inertial mass, and inversely proportional to the time.[12]

This point deserves to be stressed, particularly since there has been an influential line of thought which has rejected I. E. Drabkin's pioneering work, and has represented Aristotle, in physics, as out of touch with ordinary facts of observation, and working with concepts bearing no relation to Newtonian ones.[13]

A more attractive view is that of G. E. L. Owen. Owen sees that

[11] Newtonian 'power' used (in Britain at least) to be measured for some purposes in units of 'horse-power'. For this purpose, 'one horse-power' was defined as the power needed to raise 550 pounds through one foot in one second.

[12] The assimilation of δύναμις to Newtonian 'power' is suggested by Drabkin, 'Laws of Motion', 75, though he still translates by 'force', and does not extend the interpretation beyond certain special cases. In other cases he seeks an equivalent in Newtonian 'work'. Owen's objection (*Logic, Science and Dialectic*, 156, 331) to the identification of *dunamis* with Newtonian 'work' has no force against the identification with Newtonian 'power'.

[13] This line of thought is represented primarily by Carteron, *La notion de force*; also by, for example, F. Solmsen, *Aristotle's System of the Physical World* (Ithaca, NY, 1960); by A. Mansion, *Introduction à la physique aristotélicienne* (Louvain and Paris, 1946). For Drabkin's work see his 'Laws of Motion in Aristotle'; also Cohen and Drabkin, *Source Book*, 203–12. Drabkin approaches Aristotle in the same general spirit as the present chapter, though some essential parts of his account are effectively refuted by Owen.

Aristotle is relying on ordinary untechnical 'common-sense' observations as the basis of fact for his proportionalities. But he thinks that their importance is dialectical only, and that they and other mathematical relationships are not empirically based, and were 'not intended for close scrutiny'. This theory in turn contains a radical implausibility. For the standard fate of all such 'common-sense' assumptions used in Aristotelian dialectic is precisely to be scrutinized, and then either to be rejected on closer examination, or to be incorporated, perhaps with modification, into Aristotle's own account. There is no sign of any rejection of the proportionalities; on the contrary, they do Aristotle yeoman service, enabling him to banish the threat of all sorts of infinities. We must conclude that for Aristotle himself the existence of these mathematical relationships is a fundamental fact about the physical world.[14] As to the empirical basis, this is just that, roughly, power is additive in its effects: two men can shift twice the load at the same rate as one man, or the same load at twice the rate, and so on. Or (as geared mechanisms show) one man can shift twice the load at half the rate.[15]

The interpretation of *Phys.* VII. 5 has to be checked against the rest of the evidence. There are other places, as might be expected, where Aristotle speaks of (1) the power/strength of the agent as revealed in change; or (2) a proportionality between quantities involved in change. Some of these relate to the theory of natural or forced motion, and will be considered later (sections IV to VIII). Others refer to other particular types of physical change such as heating.[16]

More important for the structure of Aristotle's thinking than these are the 'thought-experiments', using *reductio ad absurdum*, on finitude and infinity in agency and change: *Phys.* VIII. 10, 266^{a}12–^{b}27; *Cael.* I. 7, 274^{b}33–275^{b}4 and 275^{b}18–25 (cf. also *Cael.* II. 12, 293^{a}9–11). In spirit and method these passages are closely interrelated. They appeal to the existence of the proportionalities, combined with the 'axiom of

[14] For Owen's views, see his *Logic, Science and Dialectic*, chs. 8 and 18. Similar views were advanced by some medieval commentators; see Sorabji, *Matter, Space, and Motion*, 145 n. 22.

[15] For gearing by pulleys, see J. G. Landels, *Engineering in the Ancient World* (London, 1978), ch. 4.

[16] Apart from the treatment of the amounts and powers of the elements in *GC* I. 10 and throughout *GC* II and *Meteor.* IV, the following isolated passages may be mentioned (the list is in no sense complete): amounts and powers of elements: *Meteor.* I. 3, 340^{a}1–19, ^{a}35–^{b}4, 341^{a}35–6; thawing: *Meteor.* II. 5, 362^{a}27–8; extinction of fire by fire: *Cael.* III. 6, 305^{a}6–13; production of noise: *Cael.* II. 9, 290^{b}34–291^{a}4; theories of lightning: *Meteor.* II. 9, 370^{a}6–7. The proportionalities involved in perception and recollection (*An.* III. 2, 426^{a}27–^{b}7; *Mem.* 2, 452^{b}7–24) cannot be straightforwardly assimilated to these cases.

Archimedes', to demonstrate the impossibility of 'co-existence' of finite and infinite quantities in changes.[17] There cannot be infinite power acting with finite effect on a finite body in a finite time; hence there cannot be infinite power in a finite body, because a finite body acting on a finite body will always produce a finite effect in any finite time (*Phys.* VIII. 10, 266^a24–$^b6$).[18] Nor can there be finite power in an infinite body (*Phys.* VIII. 10, 266^b6–24). A finite body cannot produce an infinite amount of change or a change lasting infinitely long (*Phys.* VIII. 10, 266^a12–23). There can be no acting at all of a finite body on an infinite one, or conversely (*Cael.* I. 7, 274^b33–275^b4, and 275^b18–25; II. 12, 293^a9–11). In these passages the proportionality is put to do essential work in laying the foundations of Aristotelian physics.[19]

A complication of the proportional relationships which Aristotle recognizes in *Phys.* VII. 5, 250^b7–25 is the 'threshold proviso': a certain amount of power is needed before any given thing will start changing at all. Though this is not stated in the *Physics*, it is natural to suppose that this threshold of power is constant for any given thing. It too then is a measureable quantity, which we might expect to be proportionate to the 'amount' of the object in question. This surmise seems to be confirmed by the striking thought-experiments of *De Motu Animalium*, chapters 3 and 4 (699^a27–$^b11$ and 699^b12–700^a6). The threshold is represented here by the maximum amount of power which does *not* start the object moving. Thus if just this amount of power is supplied there is a precarious equilibrium. And the amount is

[17] Proportionality is not always mentioned, and is not always the minimum assumption required mathematically for the argument to go through; but there is no indication that Aristotle was in fact working with any weaker assumption, which would have had to be of considerably greater mathematical sophistication. (The 'axiom of Archimedes' is the assumption, formulated explicitly by Eudoxus, which serves to rule out infinitesimal magnitudes of any kind, even divisible ones: it states that, for any two given magnitudes M, N of the same kind, there will be a finite multiple of M which exceeds N.)

[18] Here the 'amount' (i.e. volume) of the agent appears, a quantity not mentioned at *Phys.* VII. 5, but needed here for Aristotle's cosmological reasonings. There is, in fact, a latent assumption in this passage that the agent is freely divisible, as a magnitude is, into smaller agents, of which it can be taken to be the 'sum', both in volume and in power. The assumption is, obviously, plausible only for homogeneous bodies, and not for agents such as animals; it represents, therefore, a further specialization of the argument. Once Aristotle has proved on this basis the key propositions of *Phys.* VIII. 10, 266^a24–$^b6$ (no infinite power in a finite agent) and 266^b6–24 (no finite power in an infinite agent), he can simply treat 'finite/infinite power' and 'finite/infinite (amount of) agent' as equivalent.

[19] The fact that they are 'thought-experiments', which proceed by *reductio ad absurdum*, does not licence the conclusion that Aristotle is not committed to the assumptions they use.

proportionate to the body to be moved (699^a35–7). This shows that even a strategically important exception such as the 'threshold proviso', which might be taken to put severe limits on what can be formulated mathematically, is itself amenable to a quantitative investigation.[20]

III The Mathematical Theory of Physical Change: (ii) The Composition of Changes and Powers

A development which is structurally of immense potential importance for the development of a truly mathematical physics, is contained in two steps taken by Aristotle: (1) his recognition that two or more changes of the same kind, undergone simultaneously by the same body, will produce a compounded change that is a kind of sum of its components; and (2) his use of specific quantitative rules for determining the compounded change.

In the case of straight-line motions at constant speed, the rule is simply the one known in modern times as 'the parallelogram of velocities'. But it can be derived, and presumably was so derived by Aristotle, from a more general rule which does not assume the motions to be either necessarily straight-line or necessarily at constant speed. To find the distance gone and the direction from the starting-point, after a given time, as the combined result of the two motions, the distances and directions resulting from the two motions separately are simply added vectorially, i.e. according to the 'parallelogram' rule. The result is the same as if a mechanical model were postulated, in which body *A* moves with one of the motions, and carries along with it body *B*, which also moves with a second motion relative to *A*.[21] The

[20] *Cael.* II. 14, 297^a30–$^b14$ might seem to need to be interpreted in the light of the *MA* passage, since this too is a thought-experiment about 'moving the earth', involving some sort of equilibrium; but the thought is obscure (it may possibly—a more exotic guess—involve the principle of the lever, and 'moments' about the centre of gravity). *Cael.* II. 6, 288^b30–289^a1 (a minimum *time* for the heavens to revolve once) cannot be brought in here, since a minimum time would imply a *maximum* power for setting in motion. Wardy, *Chain of Change*, 307–8, argues against this line of interpretation on the grounds (1) that Aristotle in *Phys.* VII. 5 gives no theoretical justification for 'threshold powers'; (2) that 'a physics which treats inertia and resistance alike as undifferentiated manifestations of resistance to movement certainly does not conform to a Newtonian model for dynamics' (but then no one has ever claimed that Aristotelian dynamics conforms to a Newtonian model in every way).

[21] Aristotle grounds the distinction between absolute and relative motion, needed here, on the theory of place (*Phys.* IV. 1–4).

investigations of Eudoxus and Callippus into celestial motions as combinations of circular motions of concentric spheres, which are accepted by Aristotle, must have used this rule.[22]

The parallelogram rule is used explicitly at *Meteor.* I. 4, 342ᵃ22–8. The combination of motions here is that of an exactly upward motion and an *approximately* downward one. (An analogy is that of someone pushing roughly downwards on a large piece of wood under water. At some point the wood will wriggle sideways.) So in this passage the combined motion is roughly horizontal, but called 'along the diagonal' because it follows the diagonal of the thin parallelogram of velocities.[23]

In most types of change, there is only one dimension of change, and the rule of composition reduces to a simple sum of amounts of change undergone in a given time, or equivalently of velocities (speeds with a plus or minus sign attached), or of signed powers. Such compounded changes are mentioned by Aristotle in many different contexts.

Whenever one of the component changes is very much faster than the other(s), the resultant change will be in practice indistinguishable from the fastest component, both in speed and direction. The greater has, as it were, 'swamped' the lesser. The phenomenon of 'swamping' is denoted by such terms as 'overcome', 'stop', 'dislodge'; and in other places one change is spoken of as prevented from taking effect by a faster one in a different direction.[24]

So the mathematical rule is clear, and it is clear that it is the right rule to apply in calculating the results of the composition of changes, and that Aristotle saw this. But it is much less straightforward to determine Aristotle's logical and causal analysis of a situation involving compounded change. If two different changes of the same kind are compounded to produce a third, how many actual changes are there? In the simple case of a man in a boat, or a sphere mounted on another sphere, there are two things in motion, and two different component motions; here it is the compounded motion which appears to be not

[22] The results of Eudoxus and Callippus are used at *Met. Λ*. 8, 1073ᵇ1–1074ᵃ14, *Cael.* II. 10 and 12; *GC* II. 10, 336ᵃ14–ᵇ10.

[23] Cf. *Meteor.* I. 4 on the dynamics of shooting stars. A resultant motion 'by the diagonal' in a different sense in the movement of animals: *HA* I. 5, 490ᵇ4; II. 1, 498ᵇ5–6; *IA* 14, 712ᵇ2–12.

[24] Some instances (again not a complete list; in particular, passages in *GC* II and *Meteor.* IV not included): 'overcome' (κρατεῖν: *Phys.* VII. 2, 243ᵃ20–ᵇ2; *Cael.* I. 2, 268ᵇ26–269ᵃ2; II. 10, 291ᵃ6–10; *Som.* 3, 458ᵃ10; *MA* 10, 703ᵃ25); 'stop' (παύειν: *Phys.* VIII. 8, 262ᵃ6–12; *Meteor.* II. 6, 364ᵃ27–32 and ᵇ4–8); 'dislodge' (ἐκκρούειν: *Sens.* 7, 447ᵃ14–29, cf. 447ᵃ29–ᵇ6); one change caught up in a faster one: *Phys.* IV. 8, 215ᵃ15–17; VII. 2, 243ᵃ20–ᵇ2; *Cael.* II. 1, 284ᵃ24–6; II. 13, 295ᵃ16–21; *IA* 2, 705ᵃ9–11.

real, a sort of useful fiction. But if there is only one object, being moved in two different ways at once, the compounded motion seems to be the only actual one. If this is right, then how is this motion to be explained? It cannot be done by reference to the component motions, if these do not exist. In this case it seems that we must invoke the powers of the two agents. These agents, to be explanatory, must both be actually acting on the object. Instead of speaking of the composition of changes, we should in these cases more correctly speak of the composition of agents or powers to produce a change. This is in line with the talk of changes' 'stopping', 'overcoming', or even 'destroying' other changes.

The problem is relevant to other issues in Aristotelian physics. The problem of 'mixture' (μῖξις) of the elements, the foundation-stone of (what might have become) Aristotle's chemistry, is obviously a closely related if not identical problem. For here the powers of the component elements are somehow played off against one another in such a way that the compound apparently has resultant powers of its own. The added difficulty here is that the original possessors of the powers, the component elements, no longer exist in the mixture, except 'potentially'. The solution, if the mathematical rules are to be safeguarded, must be that though the original possessors of the powers no longer exist, the powers themselves survive, belonging now to the mixture as a whole. This has the consequence that opposed powers, e.g. heat and cold, may reside simultaneously in one and the same thing; but it is not clear that this need embarrass Aristotle, any more than, for example, two simultaneous and opposite pulls in a string.[25] A 'Heraclitean' situation, and not the only one which appears when we attempt to work out the implications of the mathematics for the explanatory structure of Aristotelian physics.[26]

IV The Theory of Motion: (i) Weight and Lightness

It is now possible to approach the most difficult area: Aristotle's theory of natural and forced motions. Again, there is nothing like a systematic

[25] On 'mixture' see the recent discussion of Sorabji, *Matter, Space, and Motion*, ch. 5 (which unfortunately treats it in isolation from other cases of composition of powers). For the powers in mixture and elemental change, see esp. *GC* I. 10, 328^{a}28–31 (there must be 'a kind of equality of powers'); II. 4, 331^{a}20–^{b}36.

[26] See below on the balance (section IV) and on projectiles (section VI).

treatment of any of this, except in the fourth book of *De Caelo*. Even there, unfortunately, Aristotle's concerns are not primarily those of the physicist, but those of the philosopher writing for a general audience.

The discussion of weight and lightness in *Cael.* IV does at least give a convenient starting-point. Here Aristotle is concerned to establish that weight and lightness can be defined absolutely, as powers involved in natural motions. He is concerned to exhibit this as a case of 'nature' as a principle of movement and rest (*Phys.* II. 1). Hence he must show these natural powers as operating so that the body is moved towards the natural goal—its natural place—regardless of the medium through which the body is moving.

Having established (311^a15–312^a8) that heaviness and lightness can be defined in this way absolutely, as powers, Aristotle (as a good physicist, dealing in observed facts) has to admit some complications in his account of the ways in which heaviness and lightness are present in actual bodies.

(1) One complication (312^a8–313^a13) is that the intermediate types of matter, water, and air, have intermediate natural places, and cannot therefore be either absolutely heavy or absolutely light. The observed heaviness or lightness of these bodies will therefore be dependent on their position. We have to take as basic the facts that, for example, water is heavy in air, but light in earth. This in turn shows that Aristotle was aware of, and could handle quantitatively, specific gravity and the more obvious phenomena associated with it.[27] For the natural behaviour, in different places, of portions of the different elements and of bodies compounded from them, will depend both upon their volumes and upon what they are made of. Leaving aside the case of compound bodies, which can be handled by averaging out the properties of their components, we are left, in effect, with sixteen different fundamental specific gravities: those of fire in fire, fire in air, air in fire, etc.

[27] Surprisingly, it is sometimes asserted that specific gravity was not known or defined before Archimedes (e.g. Sambursky, *Physical World of Greeks*, 41). It is true that Aristotle's terminology in *Cael.* IV is rather opaque, and specific gravity is not defined as such. But Aristotle moves at once to separate out, as far as possible, the complications arising from differences of specific gravity, by his distinction between absolute and relative heaviness/lightness. 'Relatively heavier than', said of equal volumes, just means 'of greater specific gravity'. And his criticisms of the 'quantity' theories of weight make essential use of the phenomena of specific gravity. His own explanation of these is at *Cael.* IV. 4, 311^a22–$^b13$. Differences of specific gravity are also recognized at *Cael.* III. 1, 299^b7–11; *Meteor.* II. 3, 359^a5–17.

(2) A complication within this complication (311^b8–10, 312^b2–19) is the principle that 'everything has weight in its own place'. This is put forward on the basis of a supposed empirical observation: an inflated wineskin weighs more than an empty one. It spoils the attractive symmetry of the system; and Aristotle seems to be in two minds about whether it applies to fire.

(3) There are also further complications (treated in IV. 6) due to the resistance of the medium to being divided: these are mostly, and rightly, abstracted from.

After these complications, designed to take care of obvious facts, are cleared out of the way, there is not much left in *Cael.* IV in the way of substantive physical theory. Aristotle is here concerned with the central fact of natural up or down movement, and much less with its mechanisms or with the mathematical laws it obeys. But we do find some notorious statements in which Aristotle, apparently in blithe disregard of very simple and easy observations, characterizes the heavier weight as 'that which moves downwards faster'. In order to approach these, we must keep a firm grasp of the empirical realities, and see whether we can understand the statements as being, like the rest of the book, firmly based on facts of observation and simple experiment.

The fundamental question to be asked is how, as a matter of practical fact, heaviness and lightness are to be measured, in any given place, as quantities. The practical answer to this question is obvious: they are determined by the balance. That is not only the natural answer, but has been directly implied by the alleged fact about the inflated and empty wineskins. So we must ask how Aristotle would interpret the action of a balance. The two weights to be compared are temporarily part of a system which may be considered as a single rigid body, fixed at its mid-point, on which two powers are pressing down at its two ends. The powers are therefore acting directly against each other, since their effect is to rotate the body in contrary senses. The stronger power will prevail. In Aristotle's terms, this means that the rotation induced by the stronger power (corresponding to the heavier body) is *faster* than the other one. What is actually being measured is the speed of the compounded motion. But since the weights-plus-balance form for the moment a single system, the component motions that are being compared do not actually exist—only the competing powers corresponding to them. It is a case of the 'Heraclitean' situation mentioned in section II.

A hypothesis can now be stated: it is the starting speeds *on the balance* that Aristotle has in mind, when he says that the heavier weight is that which 'moves downwards faster'.[28] As has just been shown, this must be true for Aristotle of the weights on a balance, provided we realize that the motions being compared are hypothetical, and are strictly speaking motions of the balance as a whole. And it is difficult to see any other sense in which it could be true. It is totally implausible to take it of speeds measured over an actual length of fall. In the first place, Aristotle had no way of measuring short time-intervals accurately, and must have known quite well that he did not. In the second place, it is a matter of common observation that heavier bodies do not in general fall noticeably faster than lighter ones. Above all, such a test was not the obvious and practical test to use in determining relative weight. The obvious practical test was weighing on the balance. The lack of explicit mention of the balance here is wholly understandable. Aristotle, here as so often elsewhere, is alluding rather lazily to a context of ideas, familiar to him (and presumably to his intended audience) and automatically present to his mind, though not necessarily easy for us to recover.[29]

The great advantage of this explanation is not just that it removes from Aristotle all appeal to non-existent facts of experiment, and thereby the imputation of laziness and dishonesty. It also shows him, positively, as a good working physicist, concerned to explain the few actual facts of experience that were available to him. Failure to reach an understanding of Aristotle in this area has perhaps always been due to a reluctance (for whatever reason) to take him seriously as a working physicist.[30]

Two passages that go naturally with those in *Cael.* IV are: (1) *Cael.* I. 6, 273^{b}29–274^{a}17, and (2) *Cael.* III. 2, 301^{a}22–^{b}31. Here we have thought-experiments designed to refute the possibility of (1) an infinite weight; (2) a body, without weight, which moves naturally either up or down. They both appeal to an inverse proportionality,

[28] *Cael.* 308^{a}6–34, 309^{b}8–16, 311^{a}16–21, 312^{b}20–32. At 312^{b}28 the text gives no good sense (perhaps read εἰσιν for ἔσται, cf. 312^{b}15). The characterization of greater weight by faster downward motion is in effect used as a definition.

[29] This interpretation seems preferable to the one given in my *Physics III and IV*, Additional Note B, 192–4, which also involves an analysis abstracting from actual observations, but one for which there is less supporting evidence.

[30] A reluctance hard to justify in view of the biological works, unless we suppose (with Owen, *Logic, Science and Dialectic*, 164) that 'Aristotle the meticulous empiricist' did not emerge from his Academic chrysalis until after *Physics* and *De Caelo* were written.

other things being equal, between the weight of a body and the time the body takes to move a certain distance. In (2) it is certainly only upwards or downwards motion that is considered; in (1) the indications are less clear.

The situation envisaged here cannot be of the type to which the proportionalities of *Phys.* VII can be straightforwardly applied, i.e. a situation in which the body is steadily acted on and in contact with an external source of power. Here, the weight is *inversely* proportional to the time, and must therefore itself be supplying the power (this is also suggested by the fact that no external agent is mentioned).[31]

These passages should probably be read not in the light of *Phys.* VII. 5 but of the *De Caelo* passages already mentioned, which deal with behaviour *on the balance*. The moral of those passages is that weight, as a measurable quantity, is revealed by the balance, and that what is really happening in weighing is that two opposed motions are being compounded. These motions are both motions of the total system of balance-plus-two-weights, not of the individual weights; they are hypothetical, but correspond to really existing and distinct powers, and may therefore be considered each in isolation from the other. For these motions, as explained, it is true that the speed is inversely proportional to the weight. The explanation I offer, then, is that Aristotle is here too talking about 'weights' as disembodied powers of locomotion, as they are revealed by the balance. There is no difficulty in applying this interpretation to either passage, and the second positively invites it, since it is concerned with natural upward or downward motion.

The exercise is then not exactly one in the physics of the real world. These are 'thought-experiments': the experiments could at least in principle be performed, but only if the physically impossible hypotheses of the discussion could be satisfied. It is a use of an abstract science of kinematics to explore the limits of the real world. This procedure is in line with other mathematical-physical reasonings about the infinite in *Cael.* I and II.

[31] In any case it is not possible to make sense of this proportionality in terms of the *Phys.* VII passage. For if the clause 'other things being equal' is strictly construed, as it must be, though the weight of the body, considered as the amount of the moving power, may vary, the 'amount' of the body, considered as a changing thing, must not do so. This would be absurd: for the 'amount' of the body, in the sense relevant for motion, is again just the weight.

V The Theory of Motion: (ii) Initial Conditions and 'Impulsion'

The hypothesis just put forward may make good sense of the proportionalities in *Cael.* IV. But there are other hurdles to be overcome by anyone who claims to find consistent mathematical reasoning, close to experience, in Aristotelian physics. There are, above all, at least two other passages, which evidently cannot refer to behaviour on the balance, and which yet have been taken as asserting that heavier weights travel (downwards, or in general) faster. The focus of difficulty is the notorious pair of proportionalities stated at *Phys.* IV. 8, 215^a24–216^a21.

In *Phys.* IV. 8, Aristotle is engaged in a thorough refutation of Atomist void. In this passage, he starts from what (he claims) is given as a fact of ordinary experience:

> We see that the same weight and body moves faster for two causes, either (A) by the differing of the medium through which it moves (e.g. through water rather than earth, or water rather than air); or (B) by the differing of the moving thing itself, when other things are equal, by the excess [ὑπεροχήν] of weight or lightness.

Note at the outset that this passage (1) makes explicit appeal to what is claimed as an uncontrovertible fact of ordinary experience, by the emphatic ὁρῶμεν ('we see'); (2) does not mention upwards or downwards motion explicitly, and therefore seems to be meant to apply to all types and directions of motion;[32] (3) compares the behaviour of 'the same weight and body' in different initial conditions. The phrase 'same weight and body' should mean that it is not only numerically the same body, but that it is thought of as keeping the same weight (which may however be 'lightness'; the word βάρος is used conveniently if loosely to cover both). Any acceptable interpretation must be able to square with all three points.

The difficulty, of course, lies with the words 'by the excess of weight or lightness'. It is hard at first to see what this could mean, if not 'by the fact of its weight or lightness being greater'. But this interpretation conflicts with not just one, but all three of the points just noted. There is no fact of ordinary experience that tells us that heavier bodies move

[32] As Owen notes (*Logic, Science and Dialectic*, 158), it also discounts the possibility of acceleration.

faster, all other things being equal; there is no obvious relevance of heaviness and lightness to sideways motion; and above all it has been stipulated in the formulation that the weight is supposed to remain the same. For the same reasons it will not help to think of specific gravity rather than weight. In the hope of throwing more light on this darkness, we must turn to 216^{a}12–21, where the second part of the original proposition is applied to a refutation of the void.

In respect of the excess [ὑπεροχὴν] of the moving things, the consequences are these. We see [ὁρῶμεν] that things that have a greater impulsion [ῥοπὴν] of weight or of lightness, if other things are equal, move faster over an equal space), and do so according to the proportion [κατὰ λόγον] that the magnitudes [τὰ μεγέθη] have to one another[33] . . . For what cause will it move faster? In plena this happens of necessity; for the greater [quantity] more quickly divides [the medium] by its strength. (For it divides by its shape, or by the impulsion which the moving or projected thing has.)

The new information from this passage is first, that the 'excess of weight or lightness' is also called a 'greater impulsion of weight or lightness'; second, that a proportionality seems to be affirmed; third, that the greater impulsion gives the body greater 'strength' to divide the medium through which it travels. Points (1) and (2) are reinforced: again an appeal to a fact of experience, and again no specific mention of up or down.

In a difficulty of this kind, it may be no bad idea for the interpreter to have recourse to common sense. Aristotle is professedly arguing on the grounds of ordinary experience and common observation. He is asking why one and the same body moves, in different circumstances, now faster, now slower. One possible explanation is the difference of the medium through which it is moving. But if that is not available, the only common-sense remark to be made is that the initial conditions must be different. To put it crudely, a body moves faster if it has been given a bigger push to start with.

The hypothesis for the explaining of this passage must be, then, that 'excess of weight or lightness' and 'greater impulsion (of weight or lightness)' refer to the same thing, something corresponding to a bigger initial push, but kept, somehow or other, for some time by the body—in other words, to something like a greater 'impetus' or momentum. In order to make such an hypothesis plausible, it is

[33] I follow the text of Ross: in particular in excluding the reference to shape in the *ceteris paribus* clause (216^{a}14–15).

necessary to show (1) that Aristotle would have felt, and did feel, the need of some physical quantity approximately corresponding to *momentum*; (2) that he could well have called this quantity 'impulsion' or 'excess of weight or lightness'; (3) that there is room for such a concept within his physics.

(1) The first task is straightforward. It needs only an ordinary acquaintance with the real world to show that the effect of a moving body on things with which it collides is dependent in some way both on its weight and on its speed. Roughly, for any fixed speed, the greater the weight the greater the effect; and for any fixed weight (as in Aristotle's discussion), the greater the speed the greater the effect. This latter point is made explicitly at *Meteor.* II. 8, 365^b31–3.[34] Therefore it is entirely necessary that Aristotle, if discussing the power of a given moving body to push aside the medium through which it is moving, should invoke some quantity which, like momentum, depends linearly on both weight and speed.[35]

(2) The linguistic side of the question. The word ῥοπή has here been translated 'impulsion'. Like most of the words which acquire the status of technical terms within Aristotle's writings, it is taken from ordinary speech. The verb ῥέπειν, in early Greek, is used primarily of the heavier weight in the balance, and thence develops a web of uses, which can be grouped not too artificially around three leading ones: (i) 'be stronger/heavier than something else, prevail'; (ii) 'come down on the one side rather than the other, decide, happen in a decisive way' (especially in a nicely poised situation); and (iii) 'come down heavily'. The noun ῥοπή has corresponding uses, and is applied to the process, or whatever causes it.[36] Within Aristotle's writings it is unproblematically applied to weight/lightness, or to the downward/upward 'push' associated with them, or to the actual resulting movements.[37] There is

[34] Perhaps also *Meteor.* III. 1, 370^b9–10, if the speed of separation out entails speed of motion. That a greater speed means greater power is also implied by the use of 'more violent' (σφοδροτέραν) at *Phys.* VII. 2, 243^a20–$^b1$.

[35] Of course Newtonian momentum depends on (inertial) mass and velocity rather than 'weight' and 'speed'; but the closeness of the correspondence is manifest. On the question of how, and how far, 'impulsion' was conserved, see the next section.

[36] The semantic situation is not well elucidated in Liddell & Scott, where the entries for ῥέπειν and ῥοπή are confusingly arranged, and do not explain the relationships between the various uses. The use of the word by later writers on physics and mechanics would be interesting to follow, but is not directly relevant here.

[37] At *Cael.* II. 1, 284^a25; II. 14, 297^a28, $^b7$, 10, 14; III. 2, 301^a23, 24; III. 6, 305^a25; *Met. I.* 1, 1052^b29. At *Cael.* IV. 1, 307^b33 the word is considered but not accepted as suitable for the actuality corresponding to the potentialities implied by the words 'weight' and 'lightness'.

no linguistic difficulty, then, in taking it here as the 'push' attributable jointly to the weight and the speed of the body.

More difficult is the phrase 'excess of weight or lightness'. 'Excess' implies a comparison. The very same body, keeping the very same weight, as stipulated at *Phys.* 215^a25, cannot be said to acquire an 'excess' of weight in the ordinary sense. The phrase will make sense in its context only if 'impulsion' is thought of here as a kind of temporary 'extra weight' possessed by a moving body, in proportion to its speed. The effect of the weight (the impulsion) is multiplied by the extra speed. Think, once again, of two weights in a balance: of two equal weights, one which is given a downward push will fall in the balance; it is as though it has a kind of temporary 'excess effective weight'. This, then, is a perfectly intelligible way of thinking of 'impulsion', though there are no other examples of this way of speaking in Aristotle.

VI The Theory of Motion: (iii) 'Impulsion', Projectiles, and Mechanisms

Task (3) still remains. It may well seem that there is no room for any concept at all like that of *momentum* within Aristotle's requirements for the causal mechanisms that are involved when bodies move. The present section will concentrate on the problem in connection with the non-natural motion of projectiles (on natural motion, see section VII).

A preliminary point: there is a prejudice, derived from Newtonian mechanics, that explanations of projectile motion, and even more of natural motion, ought to be very simple and basic, and not to involve any elaborate mechanism.[38] But this is misleading for Aristotle. Since all motion involves the continuous contact of mover with moved, the 'principles of motion' are pushing and pulling (*IA* 2, 704^b22–3) and every motion must be reducible to these, possibly in a rather complicated way. Nor do natural changes, generally, have to have simple efficient-cause explanations, merely because they are natural.[39]

The next point is that in the extant writings Aristotle does not put

[38] Though this prejudice has now been partly eroded by the progress from Newtonian mechanics to quantum mechanics, which suggests that even the simple motion of a particle in a vacuum may in fact be the much more complex motion of a packet of 'waves' of some kind.

[39] There is, for example, an enormously complex set of physical processes involved as efficient causes in the natural development of a human being.

forward any substantive theory, in terms of efficient causes, to explain the persistence of motion of projectiles. In particular, the fullest passage, *Phys.* VIII. 10, 266^{b}27–267^{a}20, does not state such a theory. In summary, it says (1) that the projectile will have to be moved by the medium through which it progresses, and which it in turn moves; (2) that the medium for this purpose will have (i) to have been acted upon by the original source of motion, and (ii) to be acting not as a whole but as a number of interacting parts; and (3) that at any moment there will have to be at least one part of the medium acting as an 'unmoved mover'. The aim of this outline is not to provide a substantive physical theory, but to respond to certain general doubts, by stating certain general constraints which any such theory will have to satisfy. Any criticism which treats this passage as a statement of 'Aristotle's theory of projectile motion' is wide of the mark.[40] (Aristotle admittedly does not help understanding by his references to the earlier theory of *anti-peristasis*, which could (but need not) be read as though he himself is endorsing it.)

The points made in *Phys.* VIII. 10 are closely paralleled in Aristotle's descriptions of other types of processes. This is as we should expect, and shows that for Aristotle the motion of projectiles is not an isolated and bizarre phenomenon, but may be sufficiently understood by attention to familiar facts of experience.

In the first place, media such as water or air have the power to propagate changes within one part of themselves to some distance. The obvious example is waves and vortices in water. The transmission of colours, sounds, and smells through air (and water) are, for Aristotle, other examples. For sounds at least, the mechanism is a series of local motions of small parts of the air, which are presumably thought of as longitudinal wave-motions.[41] Even the basic observational facts

[40] The passage does not warrant such reconstructions as e.g. that of Sorabji, *Matter, Space, and Motion*, 144: 'there is a series of pockets of air behind the projectile which acquire this power [of being a no-longer-moved mover] in turn, and they move the projectile on.' This sort of reconstruction is not even presupposed by the objection of Philoponus at *in Phys.* 641, 13–32, which Sorabji is elucidating; all that Philoponus' objection assumes is that the thrower transmits no power to the projectile, and 'it is only by pushing the air that he moves the stone'. Even these assumptions go beyond what Aristotle actually says, but it seems that Philoponus (unlike Sorabji) is not reading them into the text but making them, for the sake of argument, in order to clarify the possibilities available to an Aristotelian theory. For the substance of Philoponus' objection and its importance, see the end of this section.

[41] On transmission of sound, *An.* II. 8. In the visual (and the olfactory) case Aristotle speaks vaguely of 'changes' of the medium (*An.* II. 7, 418^{a}31–^{b}2, 419^{a}10–21, 25–35); also *Sens.* 6, 446^{a}20–447^{a}11 (note the tendency of the movements in the medium to disperse,

accessible to Aristotle would have been enough to show that, if so, the parts of the air involved would have to be very small, the wavelength very short, and the frequency very high by comparison with ordinary waves.

Next, there is a whole range of changes of different kinds, many of them natural or goal-seeking, in which the phenomenon occurs of interaction with a partly resistive and partly co-operative environment or medium. Within the class of motions, the clearest example is the locomotion of animals. The mechanics of this are examined in great detail in *De Incessu Animalium* and *De Motu Animalium*. Animal movement is like rowing: the animal has limbs equivalent to oars which it uses to lever itself forward through the resisting medium, or gets the same effect by bending its body sharply (*MA* 1–2; *IA* 3, 705^{a}3–19). For each 'oar' there is a fixed 'point' during the movement, corresponding to the fulcrum of the lever; though in the animal this 'point' is actually two organic parts, one active and one passive (*MA* 1, 698^{a}18–^{b}7; *IA* 2, 705^{a}19–23). The classification, for biological purposes, of the movement of animals attends to first, the number and nature of the 'fixed points', secondly the ways in which different movements are combined.[42] The important point for present purposes is that though the surrounding media resist the passage of the animal, as they resist any passage of any body, this very resistance is essential to the process of 'rowing'. The interaction with the environment therefore has both a 'co-operative' and a 'resistive' component.

If we go further afield, we can find more examples of the same 'Heraclitean' pattern.[43] The vital processes of the higher animals depend on breathing in and out, and on being able to take in nourishment and excrete the residues. In this they exploit the environment. But their environment also resists their attempts to live, in various ways: they already have to expend effort in breathing, feeding, and digesting; and their higher functions, in which they try to move in and act on and through their environment, meet with physical resistance. Going higher still, the activity of the craftsman in creating an artefact

and the mathematical relationships of pitch referred to different speeds of motion). Aristotle's theory of magnetic attraction is not known, but must presumably also have involved the transmission of changes by the medium.

[42] See *HA* 490^{a}26–^{b}1, 498^{b}5–10; *PA* 693^{b}8–15, 696^{a}11–16; *IA*; *MA* (esp. 1 and 3); and Owen, *Logic, Science and Dialectic*, 318.

[43] For the co-presence of collaboration and resistance in Heraclitus see Heraclitus B 8 Diels-Kranz (quoted by Aristotle, *EN* VIII. 1, 1155^{b}4–5); and A 22 Diels-Kranz (quoted by Aristotle, *EE* VII. 1, 1235^{a}25–9).

involves acting both with and against the resistance of the medium. And the good human being, in trying to live a complete life, depends on the social environment essentially, but it is also that very social environment that makes the task more difficult in some ways.[44]

All this is relevant to the problem of understanding the mechanisms of the motion of projectiles, and the way in which the concept of 'impulsion' can be accommodated to that mechanism. The relevance may be shown directly by *Cael.* III. 2, 301^{b}17–30, where the moving body is said to 'use the air as an instrument', and air is said to 'collaborate'. But these are fairly vague expressions. It is convenient to develop a clearer sense of the possibilities available to Aristotle through a series of objections about 'impulsion'.

First objection: 'If the "impulsion" of a given projectile is due to its speed, and if its speed is maintained only by the co-operative action of the medium, then its "impulsion" is not something, like its weight, that it keeps of itself, without external action.' The conclusion is correct, but is not an objection. For nothing requires that the 'impulsion' should be like the weight in *this* respect.

Second objection: 'If the "impulsion" is eventually "used up" in overcoming the resistance of the medium, all this means is that the body is slowed down by the resistance of the medium. But this is doubly contradictory: (1) It is claimed that the body cannot maintain any speed of itself, in any case; how then can it be said to be slowed down by something? (2) How can the body be slowed down by the very medium which is helping to maintain its speed?' Both of these objections can be met, provided only that we can successfully distinguish between the 'co-operative' and the 'resistive' effects of the medium. The projectile is said to be 'slowed down' by the resistive effects. The co-operative effects are counted for this purpose as belonging intrinsically to the projectile.[45] A parallel may be helpful. No animal can, strictly speaking, go on living 'of itself' for more than a few seconds: it needs at every moment to draw in air from the environment and breathe it out. But we do not count the powers and capacities of an animal as merely momentary, or as not strictly belonging to it alone, just on that account.

Third objection: 'How, if all this is true, can a greater "impulsion" be

[44] The ambivalence of 'the environment' for the life of animals (both co-operative and resisting) is explicit at *Long.* 3, 465^{b}25–7. On the related question of 'the action of time' at *Phys.* IV. 12, 221^{a}30–^{b}7, and IV. 13, 222^{b}16–27, see my *Physics III and IV*, 166–9.

[45] Persistence of the motion 'in' the projectile is implied by *Phys.* VII. 2, 243^{a}20–^{b}2.

said to be a *cause* of greater speed?' In fact, all that is said is that a greater *initial* 'impulsion' is a cause of greater speed, for in the *Phys.* IV. 8 passage Aristotle, as already remarked, is discussing initial conditions. The cause of the greater speed throughout is the greater initial push, which consists in 'impulsion', not in speed.

Fourth objection: 'Aristotle tells us that there is no sense in talk of instantaneous speed. But then there can be no sense in talk of instantaneous amounts of "impulsion" either. And yet it is supposedly the amount of "impulsion" at the moment of impact that determines the effect of a collision.' This objection highlights a very general difficulty that Aristotle had in dealing with physical quantities varying over time, and which remains whatever interpretations are adopted. It is plausible, I suggest, that he thought that bodies could always, with negligible error, be regarded as moving at uniform speed over very short periods of time.[46]

If this defence of the possibility of an 'impulsion' theory is successful, then it turns out that after all it must be something like an 'impetus theory'.[47] It also turns out that the objection of Philoponus (see n. 39) goes to the heart of the matter, but does not in the end tell against Aristotle. Philoponus points out that if the thrower communicates no power to the stone but merely sets the air in motion in some way, we would expect it to be possible to project stones more efficiently by merely making the air move. It now turns out that for Aristotle the original thrower of the stone does, in fact, communicate a power to the stone, *as well as* setting the air in motion, either directly or via the stone. *Both* are needed for the subsequent persistence of that power, the 'impulsion'. Aristotle's theory must involve a projectile and a medium in constant reciprocal interaction. The original push both gives the projectile an 'impulsion' and starts the collaborative motions of the medium. Thereafter, there is reciprocal interaction, like that between a swimmer and the water, by which the 'impulsion' helps to

[46] The need for this approximation may be one of the reasons why the laws of physics do not in general quite fit the actual world.

[47] This conclusion is also suggested by *Cael.* III. 2, 301^{b}17–30: there is some power which 'hands on' and 'as it were fastens' the principle of change in the moving object. 'Impetus theory' is a fairly elastic term. I use it here as implying that some power (in the Aristotelian sense) is communicated to the projectile originally, and that this power persists, though it diminishes, for some time, and is an essential part of any efficient-cause explanation of the motion of the projectile. On the history of the term, see M. Wolff, 'Philoponus and the Rise of Preclassical Dynamics' in Richard Sorabji (ed.), *Philoponus and the Rejection of Aristotelian Science* (London, 1987), at pp. 84–8.

maintain the collaborative motions, and these in turn maintain the 'impulsion'. But there are also two factors which slow the projectile. First, some of the 'impulsion' has to be used up irreversibly in acting to divide the medium, so that the projectile can actually make progress. Second, in the reciprocal interaction already described, some power gets irreversibly dissipated in the medium. This is a rather complicated, but in Aristotelian terms a totally natural type of theory. It also looks a more promising type of physical theory, and one more capable of fruitful development, than any of its rivals from later antiquity.[48]

The precise mechanism of the collaborative motions is, of course, still lacking. We still would like to know just how it is that the co-operative motions of the medium persist, and just how they act to maintain the 'impulsion' of the projectile. But, as already mentioned, Aristotle does not give details. The medium pushes, in some way; that is pretty well all we are told.[49] Aristotle's silence need not be taken as the silence of embarrassment; there is no place in the extant works where we would expect him to go into these details. It may be the silence of the good scientist who wants to investigate further before he commits himself. But it would be mistaken to claim that there is simply no room for any satisfactory theory along Aristotelian lines. Such a claim can be refuted by a speculation, given simply for the sake of example, which might go as follows. The original throw, like the dropping of a stone into a pond, causes transverse waves in the medium, which are not immediately dispersed but are propagated along the orbit of the projectile, and simultaneously with it. These are so phased that they pull the projectile from in front, and push it from behind. Such waves are not observable directly, but then neither are sound-waves, or the motions that explain vision at a distance, or magnetic attraction.[50] That is the *kind* of theory that we would expect, from the existing analogies.

A difficulty about this whole line of thinking is raised by the claim at

[48] On 'impetus theories' in later antiquity and the middle ages, see Sambursky, *Physical World of Late Antiquity*, ch. 3; Sorabji, *Matter, Space, and Motion*, ch. 14; Wolff, 'Philoponus and Preclassical Dynamics'; E. Grant, *Physical Science in the Middle Ages* (Cambridge, 1977).

[49] Besides the *Phys.* VIII. 10 and *Cael.* III. 2 passages already mentioned, projectile motion is also briefly touched on at *Phys.* IV. 8, 215ᵃ14–17; *Insom.* 2, 459ᵃ28–33.

[50] Compare the theory of dreams in *Insom.* and *Div.*: here waves and vortices in water are used as a model, and movements in air as an explanation (*Insom.* 3, 461ᵃ8–30; *Div.* 2, 464ᵃ6–15) for the movements in the blood of the sleeper. The parallel with projectiles is explicitly drawn (*Insom.* 2, 459ᵃ28–33).

Phys. VIII. 10, 267^a5–7, noted above, that, at any given time, there is one part of the medium that is functioning as an 'unmoved mover'. The reason why the claim is made is clear enough. There will be an infinite regress, if the small bits of the medium which move and are moved by one another are going themselves to be treated in effect as projectiles. But if not, then they will remain each in contact with the next while they move. But they must not be required to all move at once, which would make the whole process impossible. Hence some bit will always be being moved by another bit which is at rest. The argument of *Phys.* VIII. 10 is quite general, and will apply to absolutely any propagation of motion, or any other change, through any continuous medium or body. Hence there must be a very large number of cases of 'unmoved movers', quite apart from the case of projectile motion.

The difficulty here may be divided into a general part and a part specific to the case of motion. The general difficulty is that these 'unmoved movers' seem to be something mysterious and exceptional within Aristotelian physics. The analysis in the earlier part of *Phys.* VIII is sometimes thought to need the principle that 'everything that causes change, does so when it is itself changing'. This general problem can be resolved, if at all, only by analysis of the elaborate argumentation of *Phys.* VIII. This will not be attempted here.[51] The specific difficulty is to explain how something no longer moving can still have the power to cause motion, if we think of this power as being 'impulsion'; for if 'impulsion' like momentum depends linearly on speed, the 'impulsion' of a body at rest must be zero. Once again, in the face of such a problem, it is difficult to see any satisfactory solution for Aristotle short of a 'field-theory', in which a medium at rest can contain tensions and pressures capable of causing motion.

VII The Theory of Motion: (iv) Natural Motion

The natural motions have been left till last; being natural, they are not merely the paradigm and most important case, but the most complex one.

If *Phys.* IV. 8 has been rightly understood above, Aristotle must have thought that all heavy bodies, let go without a push, fall initially, in the

[51] On *Phys.* VIII and its arguments see particularly S. Waterlow, *Nature, Change, and Agency in Aristotle's Physics* (Oxford, 1982), ch. 5.

same medium, at the same speed. Since there is no 'excess' in either of them, and the medium is by hypothesis the same, there is no reason for them not to fall, at least to begin with, equally fast. (This, of course, refers to 'free' fall, *not* to behaviour on the balance, which is different for the reasons explored in section IV).

The same result may be arrived at by considering the powers and the proportionality involved. The powers are here simply the weights; for Aristotle considers weight (as a special case of 'impulsion') as a power which acts on bodies to move it downwards (*Cael.* III. 2, 301^{b}25; IV. 1, 308^{a}3). But the 'amount' of the moving body, in the relevant sense, is also the weight. Hence from the five-term equation the weight, which appears on both sides, may be cancelled out, and the result is, again, that all heavy bodies fall at the same speed initially.

A possible objection: the conclusion at *Phys.* IV. 8, 216^{a}20, is that all bodies will move with equal speed in a void; which is taken as a *reductio ad absurdum* of the hypothesis that there is a void. But this refers, as the surrounding argument makes clear (see section V), to the absurdity of supposing that all bodies all move with equal speed regardless of *all* of the initial conditions of projection.

Aristotle's laws of proportionality therefore successfully account for the fact that all bodies fall at the same speed, if we abstract from the resistance of the medium. The ultimate test for the laws is the problem of explaining the acceleration in natural motion. For this is a fundamental and well attested fact about one of the most fundamental kinds of change; and as an increase in a quantity it calls for quantitative treatment. Could the system meet this challenge?

As has often been pointed out, Aristotle did not have, already developed, the mathematical resources to deal satisfactorily with non-uniform speeds. But it is fairly clear what his first move would have been in the attempt to handle them. Various remarks in *De Caelo* show that, on some questions of physics, the physicist may safely allow himself to be guided by mathematics, since the truths of mathematics are closely related to physical truths.[52] Now mathematics shows that we may analyse the more complex into its simple components. Just as Archimedes created an approximative method for dealing with areas

[52] Mathematics is used as a guide to fundamental properties of the physical world at: *Cael.* I. 2, 268^{b}17–20 (analysis of motions into straight-line and circular, the only simple motions); *Cael.* I. 5–7 (finitude guaranteed by the fact that infinite quantities cannot enter into proportions); *Cael.* III. 7, 306^{a}26–30 (divisibility of bodies ad infinitum guaranteed by mathematics).

and volumes, which did most of what was needed, so Aristotle could have created a way of analysing non-uniform changes, at least to a first approximation, as finite sums of uniform ones.[53] And this would have carried with it the corollary that the physical situation corresponded to the mathematical one: therefore, on this hypothesis, Aristotle should have explained the acceleration in physical terms by the repeated input of extra 'impulsions' from some source.

We come back, then, to the basic physical question: what is really going on in accelerated fall? There are indications that Aristotle did not arrive at any answer that he found satisfactory, or was prepared to put forward confidently. It is not just that no place in the extant works actually offers any answer, for there is no place where a detailed theory of accelerated fall is actually called for. But those passages in which Aristotle does discuss aspects of the physical theory of natural motions are, by contrast even with the discussions of projectiles, so circuitous and cautious as to raise the suspicion that he was anxious not to commit himself.

In the first place, nothing can safely be built on *Cael.* I. 8, 277ᵃ29–33. The thought is condensed and the text suspect. It is argued that the possibility of unlimited acceleration to arbitrarily large speeds entails the possibility of arbitrarily large weight; but it does not follow that Aristotle thought that the acceleration was to be explained by an increase in weight. Such an explanation would in any case require a further explanation of the increase in weight.

Little else of physical significance emerges, except that *Cael.* III. 2, 301ᵇ23–6 tells us that the body pushes the air, so as to make the body itself move. *Cael.* IV. 3, and *Phys.* VIII. 4, 254ᵇ33–256ᵃ3, are concerned with the existence of a cause external to the whole process, which triggers it. *MA* 4, 700ᵃ11–21 reaffirms the existence of a trigger, so that it can be said (also 700ᵇ6) that all non-living things are moved by something else. The most informative passage is *Cael.* III. 2, 301ᵇ18–31. Here the acceleration of natural motion seems to be attributed to an extra 'helping hand' from the medium, which has been set in motion just as in the case of a projectile. This leaves a non-accelerated component, to be attributed somehow to the nature of the body.

So far as this goes, it is compatible with the mathematical analysis suggested. The repeated extra inputs are from the medium. These are

[53] This analysis is supported by *Phys.* V. 4, 228ᵇ15–18: a non-uniform change can be 'divided up', i.e. into uniform changes, and is in that way 'not one'.

not needed, as in the case of a projectile, to maintain the original 'impulsion', since that is done in this case by the body's nature. So the contributions from the medium are here of the nature of bonuses.

This still leaves the question of how the non-accelerated natural component works physically. It cannot be via the medium again, or it would interfere with the foregoing explanation. We need a persistent contact with some agent outside the body: this is supported by the claim that the 'nature' of the body in this case involves a *passive* power, a power of being moved, not one of initiating motion (*Phys.* VIII. 4, 255^b30–1). Aristotle may have thought to close the explanatory gap here by invoking the power of place. This can be no more than a speculation, since we have no more than hints. *Phys.* IV. 1, 208^b8–12 asserts as given in experience that the natural motions show that 'place has some power'. The arguments on place and against void again suggest that the distinctions of up and down, and the location of a place along the up-down dimension, which are impossible in a void, must be regarded as a cause of natural motion (*Phys.* IV. 2, 210^a2–5; IV. 8, 214^b13–16, 214^b32–215^a1, 215^a1–14). This then would again be something like a field-theory; another foreshadowing of a Stoic solution.[54]

All this is an exploration of possibilities, at best. If Aristotle fails to give any clear account of the mechanisms of natural motion, that is perhaps because he had found none that satisfied his high standards. There is no need to think that the failure, if such it was, was on the mathematical side, though more sophisticated mathematical machinery might have been a help. It is a failure of physical insight, presumably connected with the lack of any notion corresponding to Newtonian force. But where Aristotle failed, no one before Newton succeeded.[55]

VIII Conclusion

Unfortunately it is not possible here to look at other branches of physics: optics, acoustics, and particularly 'thermodynamics' (if we may so call Aristotle's theories about the behaviour of heat and cold). These subjects, for which the evidence is scrappy and scattered, and

[54] On discussion of this after Aristotle, see Sorabji, *Matter, Space and Motion*, ch. 11.

[55] In regard to this problem, no later 'impetus theory' was in any better position than Aristotle's.

particularly 'thermodynamics', which is full of problems, cannot be treated concisely. Yet they were an important part of Aristotle's vision of the physical world, and it is clear that here too he thought that the proportionalities were applicable to the study of powers and their effects.[56]

I have tried to show that when we take Aristotle seriously as a working physicist, a number of sheer implausibilities and stubborn puzzles vanish almost of themselves. Some obscurities persist, which is not surprising considering the scarcity of evidence, and the intrinsic difficulty of the subject. We find 'Aristotle the meticulous empiricist' at work, carefully working out the implications of the data available, and looking for explanations which are intelligible in terms of ordinary experience.

Some guiding principles of these explanations are clear. One is that there must always be a 'mechanism' involving agents acting by contact (usually pushing or pulling) on other things. Another is that there is no lurking 'microstructure' in apparently homogeneous bodies. Another is the 'continuity' in various senses of bodies and other physical quantities, and the rejection of atomism in any form. Sometimes it seems difficult to reconcile all these constraints with the data in a plausible way. When the pressure is on, it often seems that the ultimate implication of Aristotle's approach to the physical world is something like a 'field-theory' of Stoic type, which can find room, in its Heraclitean tensions and oscillations, for the co-presence of opposed powers. Within the homogeneous media or the homogeneous compounds we do find a sort of 'microstructure' after all, but it is one not of concealed substances but of concealed motions and powers.

All of this does not seem particularly to encourage the giving of quantitative explanations or mathematical patterns for physical processes. Aristotle's enthusiastic attempt, made in spite of the practical difficulties of quantitatively accurate observations, to build up a definitely *mathematical* physics, could not have been predicted from his biology, and many have thought that such an attempt is precluded by the type of explanations Aristotle requires. We must, to start with, set

[56] See n. 16. On the mathematical treatment of one area in particular R. Sorabji, 'Aristotle, mathematics and colour', *Classical Quarterly*, NS 22 (1972), 293–308. For the mostly non-quantitative treatment of heat and cold in the (possibly not Aristotelian) book *Meteor.* IV, see D. Furley, *Cosmic Problems: Essays on Greek and Roman Philosophy of Nature* (Cambridge, 1989), ch. 12: 'The Mechanics of *Meterologica* IV: A Prolegomenon to Biology'. *EN* V. 5 seems to envisage a mathematical theory of prices and monetary values.

aside anachronistic expectations: we must not demand that the mathematics and the explanations be integrated together in the same intimate way as in modern physics (in which sheer mathematical elegance sometimes seems to fill the role of both efficient and final cause).

Aristotle's quest for mathematical laws in the physical world is, like everything else in his physics, closely grounded on the experience of ordinary life. The central 'laws of proportionality' must be so understood. But the governing assumption that the mathematical relationships are there to be discovered must be due to the realization that in basic physical processes it is only the quantities and their relationships that hold out some prospect of a reasonably precise explanation; a purely qualitative law would inevitably be full of unexplained exceptions. But a law relating quantities is necessarily a mathematical one.

No doubt this realization was reinforced by the apparent success of Eudoxus and Callippus in accounting for the celestial motions—the most important changes in the world—in terms of concentric spheres and uniform rotations. If we want to look for a single source of inspiration and example for Aristotle's mathematical physics, Eudoxus—who was also the creator of the 'general theory of ratios' on which Aristotle's general treatment of magnitude relies—is the best candidate. In creating his astronomical theory, Eudoxus also showed how the composition of two motions could be handled mathematically.[57] And a method of composition is also a method of analysis.

Mathematical analysis, of compounded changes and powers into simple basic ones, could therefore seem to provide a key to the complexities of the physical world, the more so as the importance of proportions in this connection had long been realized. Not that Aristotle was led thereby into the Platonist mysticism about mathematics which has appealed to some mathematicians and physicists. For Aristotle, mathematics was a study of certain particularly basic properties of physical bodies. In pure mathematics these were studied in abstraction from others; in applications, they were reinserted into the physical world from which they had been abstracted. That the physical world should possess such an exact and beautiful mathematical structure at all was, of course, not explained thereby.[58]

[57] In so doing he must have implicitly treated distances traversed as functions of time, and thereby may have given a clue for the investigation of physical continua and their interrelationships in *Phys.* IV. 12–14 and *Phys.* VI.

[58] See n. 52 on the 'appeal to mathematics' in cosmology and physics.

There is much more, then, to 'Aristotle's mathematical physics' than has been explored in this chapter. Even within the restricted area covered, it has not been possible to discuss, for instance, the question of the resistance of the medium (it is clear that the proportionalities abstract from this: could they have been extended to handle it?). Some implications of this reconstruction for more general questions have been pointed out but not explored—the implications for (e.g.) the questions of different levels and types of explanation (efficient and final causes), necessity and determinism, identity-conditions of changes, unmoved movers and self-movers, and the nature of place. The aim has been only to show that Aristotle did deliberately formulate mathematical 'laws' of physics which were based firmly on observed facts.

Aristotle's mathematical physics is not a failure; it is a triumph of creative theorizing married to physical insight and respect for the observable facts.[59] In spite of a lack of accurate observations and of sophisticated mathematical techniques, he seized the importance of 'power' and 'impulsion', and incorporated them in mathematical relationships which were entirely satisfactory as first approximations. He grasped the need for mathematical laws of composition of powers and changes, and formulated these in the elementary cases. He also realized (a subject not considered here) the central importance for this kind of physics of the mathematical theory of the continuum provided by Eudoxus, and investigated its properties and the functional relationships between different physical continua. All this with applications not only to the theory of heavy and light bodies, but to the theory of light, sound, and heat as well. One may well doubt whether more than a few physicists in the whole history of physics could have gone as far from the same starting-point.

[59] There is a striking contrast between Aristotle, the working physicist, and his Neoplatonist critics, who came to physics from philosophy. They philosophized about the need to 'mathematize' physics, but produced no new mathematical theories themselves (and no new observations). Their positive contributions to physics took the form of isolated suggestions, which they never made the effort to work up into coherent theories capable of standing up to criticism. It is strange to find them, rather than Aristotle, being hailed as forerunners of modern physics. (The notion of 'impetus', as has been shown, is already present in Aristotle (not to mention Hipparchus).)

10

Aristotle on Self-Motion

MARY LOUISE GILL

In Book VIII of the *Physics* Aristotle discusses self-motion in order to show that particular changes caused by one body acting on another can be explained without citing an infinite number of previous changes. In VIII. 5 he argues that an explanatory regress can be avoided because ordinary changes are caused, either directly or indirectly, by something that moves itself, and self-motions by an unmoved mover within the self-mover. Typical changes thus finally depend on a mover that is not itself changed and so requires no further cause. This solution must be questioned, however, because in the very next chapter, *Phys.* VIII. 6, Aristotle suggests that animals—the paradigmatic self-movers—do not after all initiate their own motions but derive that impetus from features in their environment (259^b1–16).[1] In what sense, then, are animals self-movers, and how does their dependence on external factors affect the proposed solution in VIII. 5?

In explaining animal motion, Aristotle seems interested in questions that expect two types of answer. First, when an animal moves, why is its motion of one sort rather than another, why is the motion sustained in a particular direction, and why does it stop at a certain limit? To answer these questions he appeals to an unmoved source within the animal. This internal principle legitimates calling the animal's motions self-motions. Second, what triggers the motion, and why is it sometimes impeded? An unmoved mover cannot explain why motion starts in a body at rest or why it stops prematurely in a body in motion, because if it could the mover would itself be moved in providing the impetus or obstruction. Although an unmoved mover can be moved

[1] Cf. *Phys.* VIII. 2, 253^a11–21. For solutions to the conflict, see esp. David Furley, 'Self-Movers', in G. E. R. Lloyd and G. E. L. Owen (eds.), *Aristotle on Mind and the Senses*, Proceedings of the Seventh Symposium Aristotelicum, (Cambridge, 1978), 165–79; and Martha Nussbaum, *Aristotle's De Motu Animalium* (Princeton, 1978), Essay 2. On Aristotle's theory of self-motion and its role in the argument of *Phys.* VIII, see Sarah Waterlow, *Nature, Change, and Agency in Aristotle's Physics* (Oxford, 1982), ch. 5.

accidentally (κατὰ συμβεβηκός) when it is carried along with the body it moves,[2] such a mover cannot be moved in itself (καθ' αὑτό) and still be an unmoved mover. Aristotle appeals to independent triggering factors to avoid two unacceptable alternatives. On the one hand, if an unmoved mover could supply a sufficient reason for motion to start on a given occasion, it would need to alter from a previous state to furnish the stimulation, and would thus fail to be unmoved. On the other, if the object moved could respond to an unmoved source without an impetus, the response would be uncaused, and Aristotle would be forced to accept a thesis that he firmly rejects, that motion can start *ex nihilo*.[3] Except for the motions of the heavenly spheres, which are eternal, and so require no instigation and suffer no interruption, all others originate or digress on account of external factors. But if so, how do the self-motions of animals differ from enforced changes caused by one body acting on another? And how can Aristotle escape an explanatory regress of movers?

In the main part of my paper I shall spell out Aristotle's general theory of change, involving a distinct mover and moved, and show why he thinks that ordinary motions can be explained by a self-motion, and self-motion by an unmoved cause. The discussion of first causes will also establish his justification for distinguishing self-movers from other objects that experience motion, and hence his reason for thinking that animals are responsible for their own behaviour. Unlike most interpreters who address the problem of self-motion in Aristotle, I shall focus on first movers as efficient causes.[4] The final cause—often specified as an object of desire—though central to his account in *Metaphysics Λ*, and relevant to his discussion of intentional action in *De Motu Animalium* and *An.* III. 10, plays no explicit role in his analysis of motion in *Phys.* VIII or *De Caelo*. My aim is to determine how a first mover operates in contexts that do not involve desires.

In focusing on first movers as efficient causes, I am not suggesting that such movers are causally efficacious, injecting new forces into a causal chain. I have already indicated that an unmoved mover cannot play such a role and be unmoved. It is a mistake in any case to assume that every Aristotelian efficient cause brings about its effects, because he distinguishes first movers from instrumental causes,[5] and as

[2] *Phys.* VIII. 6, 259^{b}16–20.

[3] *Phys.* VIII. 6, 259^{b}1–16; cf. *Phys.* I. 8, 191^{a}23–31; *GC* I. 3, 317^{b}11–18.

[4] For a similar perspective, see the excellent paper by Susan Sauvé, 'Unmoved Movers, Form, and Matter', *Philosophical Topics* 15 (1987), 171–92.

[5] See e.g. *Phys.* VIII. 5, 256^{a}21–^{b}3; 256^{b}14–27; *An.* II. 4, 416^{b}20–7.

examples of first movers he regularly cites the form of the agent or the art that informs the agent's soul.[6] Since he also designates such movers as unmoved,[7] their causal role will be of some special type. Part of my task will be to explain what it means to class such entities as moving causes. This understanding should, in addition, help to clarify the nature and role of the final cause, or goal, in non-intentional contexts. Although I shall largely ignore the problem of intentional action, I shall offer some preliminary suggestions about how Aristotle's analysis of motion might apply to this case.

The last part of my discussion will treat the initiation of animal motion, and I shall argue that even though one can trace the origin of such motions to features in the environment, this fact does not lead to a regress of the sort Aristotle tried to avoid in *Phys.* VIII. 5 when he appealed to an unmoved mover. On the contrary, it indicates that all motions, including the self-motions of animals, depend on the eternal rotation of the heavenly spheres and ultimately on the Prime Mover.[8]

One final preliminary concerns my use of the term 'self-motion'. In *Phys.* VIII Aristotle focuses mainly on locomotion, or change of place, and in VIII. 6 he suggests that animals move themselves in only one way—and, indeed, not strictly even in that way. He is thinking of locomotion.[9] Yet he typically treats changes (κινήσεις) as occurring in three categories—quantity, quality, and place—and as examples of self-motion he sometimes mentions quantitative and qualitative changes, for instance, a doctor curing himself.[10] Thus, while he regards locomotion as the proper self-motion that animals display in respect of their perfected natures,[11] he evidently envisages other self-motions as well, especially those enabled by certain acquired capacities. In addition, self-caused motions seem to be of two general types: first, those, like the ones we have just been discussing, that lead the animal into a new location or state, which I shall call strictly 'self-changes'; and second, those that express a state that the animal has already acquired or naturally possesses and that, instead of leading it into a new condition, preserve or perfect the one it is already in. Aristotle typically calls such motions ἐνέργειαι, or 'activities', and as we shall see, such

[6] See e.g. *Phys.* II. 3, 195^b21–5; *Met.* Z. 7, 1032^b21–3; *Λ*. 4, 1070^b28–35.

[7] *GC* I. 324^a26–$^b14$.

[8] Cf. the interpretation of Simplicius, *in Phys.* 1258,7–1259,4.

[9] Although the reference to locomotion is not explicit at VIII. 6, 259^b6–7, it is explicit in a parallel passage at VIII. 2, 253^a11–21.

[10] See *Phys.* V. 1, 225^b7–9 and *MA* 5; for the example, see *Phys.* II. 1, 192^b23–7.

[11] *Phys.* VIII. 7, 260^b30–2; 261^a13–26; *MA* 5, 700^a31–2.

motions are presupposed by other typical changes. I shall use the term 'self-motion' as a generic expression to apply to both self-changes and activities.

Passive and Active Δυνάμεις

Changes—both enforced and self-changes—involve a mover and a moved, each characterized by a special sort of δύναμις, or potency. In *Met.* Θ. 1 Aristotle describes the δύναμις that causes change as 'the source of change in another thing or [in the thing itself] *qua* other' (ἡ ἀρχὴ μεταβολῆς ἐν ἄλλῳ ἢ ᾗ ἄλλο) (1046a10–11), and *Met.* Δ. 12 specifies the δύναμις more precisely as an active source, an ἀρχὴ μεταβλητική (1019b35–1020a6). The δύναμις responsible for a change thus typically belongs to an entity other than the object changed or, in the special case of self-change, as when a doctor cures himself, to the moved itself considered as other. He characterizes the δύναμις of the moved as a source of passive change (ἀρχὴ μεταβολῆς παθητικῆς) by another thing or by the thing itself *qua* other.[12] How should we understand these two principles?

First, a δύναμις, whether active or passive, is always directed toward a definite end or actuality. When a δύναμις is such that its realization requires a change, the actuality is a particular state of an appropriate object. For example, a doctor has an active, an invalid a passive δύναμις for health; and these δυνάμεις differ from those of a teacher and student, which are directed toward a particular type of knowledge. Second, as these examples suggest, pairs of active and passive δυνάμεις correspond in their goal: both members are directed toward a particular state to be realized in the passive object.[13] A doctor's active δύναμις for health enables him to cause health in a suitable patient; and an animal's passive δύναμις for the same state enables it, when properly affected, to become healthy.

Although corresponding active and passive δυνάμεις coincide in their goal, they none the less differ in being active or passive. We need

[12] *Met.* Θ. 1, 1046a11–26; Δ. 12, 1019a20–3.

[13] In *Phys.* III. 3 Aristotle speaks of the goal of the agent as a 'deed' (ποίημα) and that of the patient as an 'affection' (πάθος) (202a22–4), but the two goals are evidently identical, despite the difference in description, because he later objects to the idea that doing and suffering are two distinct motions that occur in the patient on the ground that there would then be two alterations of one thing to one form (εἰς ἓν εἶδος) (202a31–6).

to consider in what this difference consists. Let us begin with Aristotle's familiar discussion of the levels of passive δύναμις in *An.* II. 5.[14] We are told that an individual has knowledge potentially in one way if, although lacking the expertise, the individual is the sort of being that regularly does have knowledge. By means of a change she can acquire the positive state. An individual has knowledge potentially in a higher way if, having learned, she can use her ability when she wishes if not prevented. This second-level δύναμις enables an entity to engage in certain activities, and we can ignore it for the moment, since we are presently concerned only with changes. A change is the gradual realization of a first-level passive δύναμις—a subject's transition from lacking the property for which it has the δύναμις to possessing the property.

Since an object's lack can be rectified by means of a change, the privation is not just any state other than the goal but one on a path leading to it. The healthy comes to be from the sick, the white from the black or grey, and so forth: the privation and form must be properly opposed on a range falling under a common genus within one of the categories of quantity, quality, or place—as black and white under colour, or sweet and sour under flavour.[15] Furthermore, according to the passage in *An.* II. 5, the deprived object must belong to a kind whose members sometimes possess the form. For example, a man is potentially musical because human beings are sometimes musical, and he is potentially healthy because living things are sometimes healthy. A final point is clarified by our passage in *Met.* Θ. 1. A passive δύναμις is described as 'the source, in the very thing that suffers, of being passively changed by something else or [by itself] *qua* other' (1046^{a}11–13). A passive δύναμις depends for its realization on a source external to the passive object or, in the case of self-change, within the object itself considered as other. Accordingly, an object has a first-level passive δύναμις for some property ϕ, if it belongs to a kind whose members sometimes have the property, if it is presently not-ϕ, and if its privation can be rectified by something else or by itself *qua* other. As we shall now see, in turning to the active δύναμις, the realization of a passive δύναμις requires a form whose identity

[14] Cf. *Phys.* VIII. 4, 255^{a}30–^{b}5.

[15] At *Phys.* I. 5, 188^{a}35–^{b}3 Aristotle denies that something white can come to be from something musical, except in the incidental sense that something not-white happens also to be musical.

corresponds to the goal of the passive object, and thus one opposed to its initial lack.

On several occasions Aristotle describes movers responsible for changes. In *Phys.* III. 2 he says, 'The mover will always bring a particular form, either substantial, qualitative, or quantitative, which will be the source and cause [ἀρχὴ καὶ αἴτιον] of the change, when it produces change; for instance, a man in actuality produces a man from that which is a man potentially' (202^a9–12). And in VIII. 5 he says: 'The mover is already in actuality, for example, the hot thing heats and generally that which has the form generates' (257^b9–10). In natural contexts the changer possesses the form that it produces: a man replicates his form in matter that potentially has the form; similarly, a hot object induces heat in an object that is potentially hot. In artificial contexts the form imposed is present in the agent's soul. For instance, a builder who constructs a house has the form of a house in mind.[16] According to *GC* I. 7, the properties that enable an agent to act and a patient to suffer are like in kind but contrary and unlike in form: one body is affected by another, one flavour by another, and so on (323^b29–324^a5). By means of a change the agent assimilates the patient to itself (324^a10–11). Thus, once the change has been accomplished, the agent and patient are not only like in kind but also like in form. The agent guides the patient from a state unlike its own to one like its own.

Despite the sameness in content between the productive form of the agent and the resulting form of the patient, these forms, too, still typically differ in being active or passive. What distinguishes one form as active and another as passive, when their content coincides? While the form of the agent enables its own replication and so is somehow responsible for the agent's action, that induced in the patient, though it sometimes—as in organic reproduction—transforms the patient into a new agent, often merely adapts its possessor to respond to its surroundings in particular ways. For instance, the person who has become healthy usually cannot duplicate his health;[17] nor can a house reproduce itself. The acquired form is both a first-level passive actuality, as goal of the patient's change, and a second-level passive δύναμις. As such, it equips its possessor to behave in certain ways in response to stimulation but does not enable its own reproduction.

Passive forms are not movers but depend for their expression on an

[16] *Met.* Z. 7, 1032^a32–$^b6$; 1032^b11–14.
[17] See *GC* I. 7, 324^b14–18.

active source. Consider an artefact, such as an axe. Let us specify the form of an axe as its function, its capacity to chop. When a craftsman produces an axe, this function, which informs his soul and which he aims to embody, guides his choice of materials and tools, and the organization that he imposes. Once the axe has been completed, the artefact and its maker are like, because the axe has the capacity to chop, which the smith had in mind in carrying out his project. But the form in the artisan's mind is active, while that possessed by the axe is passive. Not only can the axe not duplicate its form, it cannot even perform its function without someone who knows how to use an axe. In a telling passage in *An.* II. 1, Aristotle contrasts the form of an axe with that of an animal:

> Suppose that some instrument, such as an axe, were a natural body. The being for an axe would be its substance, and this would be its soul. And if this were removed, there would no longer be an axe, except in name. As it is there is an axe. For the soul is not the essence and formula of such a body, but of the sort of body that has in itself a source of motion and rest. (412ᵇ11–17)

If an axe were a living organism, its form would be a soul—an active δύναμις—and the axe would be destroyed if the soul were removed. In fact, however, the 'soul' of an axe belongs, not to the axe, but to the person who uses it, and this 'soul' can be removed and the axe remain intact. The form of an axe is a passive δύναμις, an ability to respond in certain ways if acted upon by an appropriate agent. The survival of the axe depends only on the retention of this passive capacity. On the other hand, since the active δύναμις belongs to the handler and not to the artefact, the axe depends for the exercise of its capacity, and hence for the full realization of its being, on an external mover.

An active δύναμις, by contrast, is a source of motion and rest. Aristotle describes such a source in *Met.* Z. 7: 'The maker [τὸ ποιοῦν] and that from which the motion of becoming healthy begins [ὅθεν ἄρχεται ἡ κίνησις τοῦ ὑγιαίνειν], if [the production] is artificial, is the form in the soul' (1032ᵇ21–3). The active δύναμις—the 'maker' (τὸ ποιοῦν)—is the form of health that guides the doctor's production. How is such a form productive? I have already explained that a first mover, if unmoved, cannot inject forces into a causal chain. And although the active δύναμις determines the goal to be achieved, it is not simply a final cause, because, as we have seen, the outcomes of many changes, though comparable in content with their active source, are passive rather than active forms. So the active δύναμις as source of motion, though it may include the final cause, is evidently something more.[18]

An. II. 4 offers some clues about how we should understand the productivity of an active form. In discussing nutrition and growth, Aristotle first takes Empedocles to task for suggesting that such operations are fully explained by material factors. According to Empedocles, the roots of plants grow downward because earth naturally tends in that direction, and their stems and branches grow upward because fire naturally tends in that direction. Aristotle protests: Why, if this explanation is correct, do plants not fly apart, their roots in one direction, their stems and branches in another? What prevents this dissolution? It is prevented, he claims, because soul is the cause of nutrition and growth. Next he remarks that some people regard fire as the simple cause of these proceedings. To this proposal he responds,

> [Fire] is in a sense the helping cause [συναίτιον], but it is certainly not responsible without qualification. Rather the soul [is responsible]; for the growth of fire is unlimited [εἰς ἄπειρον], as long as there is something that can be burned. But all things constructed naturally have a limit [πέρας] and proper proportion [λόγος] of magnitude and growth. And these things are [the responsibility] of soul, but not of fire, and of the formula [λόγος] rather than matter. (416ᵃ13–18)

Although the soul is the cause of nutrition, there is no indication that it initiates the activity. These passages propose that the soul takes an active role in nutrition and growth by regulating the motions of the simple materials so that they serve a higher purpose. Aristotle suggests that the materials if left to themselves would realize their own δυνάνεις and so destroy the objects they constitute. Earth and fire would move off to their own places, earth toward the centre of the cosmos, fire toward the periphery of the sublunary sphere, and so dissolve the compound. And fire would burn indefinitely as long as it had stuffs to consume. But if their natural motions are adapted to a particular end, fire and other materials can help to bring about the positive outcome.

An active form does more than preserve an object composed of simpler stuffs. Passive forms, too, integrate material constituents into objects of greater complexity. The passive form that makes an object an axe, or one that makes an inorganic material iron, unifies the

[18] Note that Aristotle sometimes says that formal, final, and moving causes can coincide; see *Phys.* II. 7, 198ᵃ24–8; *An.* II. 4, 415ᵇ8–12.

ingredients into a compound with properties distinct from those of the components. As long as the passive form survives the object remains intact. Furthermore, such forms, as we shall see later when we discuss the four elements, can account for some simple types of behaviour. What a passive form cannot do, and thus what distinguishes an active form from it, is to direct motions that replicate the form in a suitable passive object, or to co-ordinate distinct, and often successive, motions toward a goal whose realization depends upon that co-ordination, as in an organism's nutrition and growth.[19]

In his biological works, Aristotle appeals to physical mechanisms, such as heating and cooling, to explain the generation of animals and their parts, but he insists that these mechanisms cannot explain why a process yields a particular outcome, for instance, flesh or bone,[20] or why flesh or bone is formed by the application of heat to an appropriate material, at an appropriate time, and in an appropriate place.[21] The nature of the product and the sequence of interactions that yields it are determined by an active δύναμις—by the soul of the male parent during the early stages of embryonic development and later, once the central organ (the heart or analogous part) has been articulated, by the foetal soul, which is located in the heart.[22] Given its regulatory role, we might compare the active δύναμις to a list of instructions that determines the materials and tools needed to realize a particular end, and the order, timing, and extent of operations that are carried out by the instrumental materials.[23]

I have argued that an active δύναμις determines the proper ordering

[19] In his theory of animal generation, Aristotle attributes an active δύναμις to the male, and a passive δύναμις to the female (*GA* I. 2, 716^{a}4–7; I. 20, 729^{a}9–11). A new animal is generated when the directive powers of the male, transmitted by the semen at conception, co-ordinate motions that take place in materials contributed by the female (*GA* II. 3, 737^{a}18–22; II. 4, 739^{b}20–30). The situation is more complicated than this, however. The female, too, has active, as well as passive, nutritive powers that enable her to process food for her own maintenance and development; it might therefore seem that she, too, should be able to reproduce her form. Aristotle recognizes this problem for his theory and points out that females of some species are to a certain extent reproductive: they can generate wind eggs. But because wind eggs lack the soul principle contributed by the male, they never become animate beings (*GA* II. 3, 737^{a}27–34; cf. I. 21, 730^{a}4–32; II. 5, 741^{a}13–32). Apparently, then, the active δύναμις of the female, though adequate for her own purposes, is simply too weak to be successfully reproductive (*GA* II. 5, 741^{b}2–7).

[20] *GA* II. 1, 734^{b}27–735^{a}5; cf. *Meteor.* IV. 12, 390^{b}2–14.

[21] *GA* II. 6, 743^{a}1–26.

[22] *GA* II. 1, 735^{a}12–21; II. 4, 740^{a}1–9; II. 4, 740^{b}18–741^{a}3. I thank James Lennox for insight into this topic.

[23] Cf. the role assigned to an animal's nature at *MA* 10, 703^{a}29–^{b}2.

of motions but cannot, as an unmoved mover, introduce new forces into a causal chain, whether by initiating a causal sequence or by imposing physical constraints. Since such a δύναμις does not originate the behaviour that it directs, the impetus must derive from other factors inside the organism or in its environment.

Does the account of active and passive δυνάμεις that I have described apply only to low-level organic behaviour, like nutrition and growth, whose origin Aristotle explicitly attributes to external factors,[24] or does it also explain psychic involvement in such high-level motions as intentional actions?[25] In his discussion of voluntary action in the *Nicomachean Ethics*, Aristotle locates the ἀρχή of action within the voluntary agent, and he arguably thinks that an internal ἀρχή (such as a choice) originates a causal series.[26] Although Aristotle's theory of intentional action must be a topic for another occasion, some doubts about the scheme's flexibility can be removed, because there is evidence that he extends his model to at least one type of mental activity—contemplation. In this situation the active δύναμις appears to play a role analogous to that in the contexts on which we have focused. Let us, then, briefly consider this topic.

In a short and tantalizing chapter, *An.* III. 5, Aristotle divides the rational soul into two components, identifying part of the intellect as active—a 'maker' (ποιητικόν)—and part as its passive material (430^{a}10–15). One might initially think that the productive intellect is introduced as ἀρχή (430^{a}19) to explain why thinking starts on a given occasion, but this is not the question that Aristotle asks. The question that motivates the distinction between active and passive mind is why the mind sometimes fails to think (III. 4, 430^{a}5–6). The failure is due, not to the active intellect, which always thinks (430^{a}22), but to the passive intellect, which is subject to change.[27]

The active intellect cannot initiate an act of thinking in the passive mind because to do so would require providing a sufficient reason for the activity to start. Since the active mind is described as ἀπαθής (430^{a}17–18; 430^{a}24), it should be immune to variation, and so unqualified to activate the thinker to episodes of thinking. Aristotle uses several analogies to clarify the ἀρχή, comparing it to a state, like light,

[24] *Phys.* VIII. 2, 253^{a}11–13; VIII. 6, 259^{b}8–16.

[25] This question was put to me by David Furley and Lindsay Judson.

[26] See e.g. *EN* III. 1, 1111^{a}22–4. This position was defended by David Furley in a commentary on my paper at the University of Pittsburgh.

[27] *An.* III. 5, 430^{a}10–11; 430^{a}14–15; 430^{a}24–5.

which enables colours to be actually visible (430^a15–17), and to a craftsman's art (430^a11–14). An art, as we have seen, is an unmoved mover that, by establishing the aim of an artistic production, regulates motions toward its achievement. If this analogy is apt, the active mind, too, should be an unmoved mover, determining the object of contemplation and controlling the mental response to it. Since the active mind cannot explain why thinking is intermittent, the passive mind is also required. Because the passive intellect is subject to change, it accounts for the fact that the response sometimes occurs and sometimes does not.

If the active mind cannot originate a causal sequence, does the passive mind do so by means of a choice or some other mental event? If Aristotle's theory of rational activity accords with his physics, he should reject this possibility, because it would involve creation *ex nihilo*. Any event must be caused, even if only by an accidental trigger. If rational activity is episodic, some change in the animal or in its environment must trigger the response.

This synopsis of *An.* III. 5, though only sketching an interpretation of that difficult chapter, shows that Aristotle applies his scheme of active and passive δυνάμεις to psychic functions besides the nutritive, and so suggests that actions involving intentions may admit a similar analysis. For example, given the treatment of active νοῦς as ἀρχή in contexts of contemplation, it seems likely that the internal ἀρχή involved in voluntary action is an active δύναμις, which directs the activity, rather than an uncaused origin of it.[28] Furthermore, since uncaused events are excluded by Aristotle's physics, mental events, like choices, even if they appear to initiate an intentional activity, should themselves be triggered by factors inside the person or in his environment.[29] So

[28] In his discussion of voluntary action in *MA* 9, for instance, Aristotle identifies the soul as the ἀρχή of action and locates it in the heart.

[29] At *Met.* Θ. 5. 1048^a8–15 Aristotle claims that desire or choice will be the deciding factor in contexts in which an agent could, in the circumstances, take either of two courses of action (for instance, when a doctor in treating a patient could improve or aggravate the condition). But Aristotle does not say that the choice between the alternatives is uncaused, and in *EN* III. 3 he treats choice as the result of a process of deliberation (1113^a2–12). Pushing the question back, one can ask what initiates the deliberation or the desire that motivates the deliberation. *MA* 8, which spells out the steps that lead to animal motion, locates the ἀρχή that originates the process in the object of pursuit or avoidance, which the animal perceives in its environment (701^b33–5). Desire results from imagination, which in turn results from thought or perception (702^a17–19). The process thus seems to originate either in the appearance of an external object that triggers the agent's perception or in some internal event that triggers the agent's thought.

although my discussion of self-motion is confined to its simplest manifestation, I suggest that Aristotle's scheme of active and passive δυνάμεις will also help to explain self-motions involving higher parts of the soul.

Enforced Change and Self-Motion

Enforced changes involve a distinct agent and patient characterized by correlative active and passive δυνάμεις, and the states of the two objects are like in kind but unlike in form. A change occurs, which replaces the patient's privation with the positive form, if (barring interference) the two objects come into contact. Since all motions, including those caused by an unmoved mover, involve some sort of contact, let us now consider this aspect of Aristotle's theory.

In *GC* I. 6 Aristotle introduces a strict notion of contact requiring that mover and moved be discretely located magnitudes that have their extremities together ($322^{b}32$–$323^{a}12$). When contact occurs in this sense, the two objects are physically contiguous, and the agent not only acts on the patient but also experiences a reaction caused by the body acted upon. For instance, a knife that cuts is blunted by the object cut, and something that heats is cooled by the object heated. He then loosens the conditions for contact to account for further examples. First, there are cases of action and passion, such as teaching and learning, in which the agent and patient, though discretely located magnitudes, normally perform their transaction without having their extremities together. Even so, there must be an appropriate relation between the mover and moved such that the one acts and the other responds ($323^{a}23$–5).[30] He then extends the notion still further to provide for cases in which the mover is entirely unaffected by its contact with the moved. It is possible, he says, for that which causes a motion only to touch what it moves, but not to be touched in return; for instance, a grieving man 'touches' us but we do not 'touch' him

[30] Commentators do not usually distinguish this type of contact from the first (see e.g. Harold Joachim, *Aristotle: On Coming-to-Be and Passing-Away* (Oxford, 1922), 147; C. J. F. Williams, *Aristotle's De Generatione et Corruptione* (Oxford, 1982), 117–19). Elsewhere I argue that contact of this sort, which typifies the relation between an intentional agent and the object it moves, differs significantly from the other two types (see my *Aristotle on Substance: The Paradox of Unity* (Princeton, 1989), 195–8). Nothing turns on this distinction for my purposes here.

(323^a28–33).[31] This final type of contact characterizes the relation between an unmoved mover and moved.

Enforced changes often involve a series of movers the last of which undergoes direct physical contact with the object moved and the first of which is unmoved. In *GC* I. 7 Aristotle discusses such an example. A doctor in treating a patient often uses an instrument, such as bread or wine, to effect his cure. Such instruments, designated as 'last' movers, actually touch the patient and so are themselves reciprocally changed. The doctor, too, is a moved mover, whose action depends on a first cause—the art of medicine, or source of health—which informs his soul. The doctor and invalid experience a relational contact (the second type I mentioned earlier); and the first mover and its objects experience a contact that is one-way, such that they are affected while it remains wholly immune.

Aristotle describes this type of chain in *Phys.* VIII. 5, when he argues that all changes have their source in a self-motion, and that self-motions are ultimately caused by an unmoved mover. Many self-motions themselves involve such a sequence. For instance, a doctor who cures himself—by acting on himself *qua* other—is both agent (as doctor) and patient (as invalid). His instruments, such as bread or wine or his medical tools, are last movers, which produce the cure, and the first cause is the art of medicine, which informs his soul. But even if there is such a series, Aristotle argues that a self-motion requires only the first pair in the sequence. He says, 'That which moves itself must contain something that causes motion but is unmoved and something that is moved but need not move anything else, and either both components are in contact with each other or one with the other' (258^a18–21); in the next line he affirms the second alternative: between the unmoved mover and moved the contact is one-way (258^a21–2).

Consider such a self-mover, for instance, a doctor who responds to his medical art—his knowledge of health. We are here ignoring the doctor's possible relation to an ailing patient, whether himself or someone else, and so disregarding the change that his motion may cause. Instead, we are considering only the motion displayed by the doctor (as moved) in response to his active δύναμις (as mover). This

[31] Although the touching in this example might seem to be merely metaphorical (see e.g. Joachim, *Aristotle: On Coming-to-Be*, 147–8), according to Aristotle's theory of the emotions (πάθη), the visible form of the grieving man is impressed upon us and thus affects us. For a helpful discussion of such 'one-sided relationships' in Aristotle, see Williams, *Aristotle's De Generatione*, 118–19.

motion is not a change, because it does not lead to a new condition; it is an activity that expresses the state that the doctor is already in. When such motion occurs, the active and passive δυνάμεις are not initially opposed and unlike, but like.

In *Met.* Θ. 8 Aristotle modifies the account of δύναμις offered in Θ. 1 to accommodate these special motions. He says,

> I mean by δύναμις not only the one that has been defined, which is called an active source of motion in another thing or [in the thing itself] *qua* other, but generally every active source of motion and rest. For nature [φύσις] is also in the same genus as δύναμις; for it is an active source of motion—not, however, in another thing but in the thing itself *qua* itself. (1049^{b}5–10)

The crucial claim is that a δύναμις can be a source of motion in the thing itself *qua itself.* A δύναμις can operate not only on a body *qua* other, as when someone cures himself, but also on a matching passive δύναμις, as like on like. The situation can be compared to that of a person who uses an axe, where user and implement share the same active and passive capacities, except that this situation involves a pair of capacities that inform a single individual. Although Aristotle's mention of a φύσις suggests that he is primarily concerned with natural contexts, the same configuration should occur in the case of some acquired capacities.

A doctor possesses not only an active δύναμις—his knowledge (a first-level active actuality)[32]—which enables him to control his bodily and psychic motions to serve his action, but also a second-level passive δύναμις (or first-level passive actuality), which determines the physical and mental organization that promotes such activity. The motion (traditionally called a second-level actuality)[33] is the mutual realization of these active and passive capacities. This motion does not yield a new organization, but rather secures, and often improves, the one that is already there. In *An.* II. 5, in contrasting this motion with a typical change involving the replacement of a previous state, Aristotle says that it is 'rather a preservation [σωτηρία] of what is in potentiality by what is in actuality and like it, in the way that potentiality is [like]

[32] *An.* II. 1, 412^{a}10–11; 412^{a}19–28. Aristotle does not explicitly distinguish actualities as active or passive, as I am doing, but such a distinction indicates whether we are re-describing an active δύναμις or a second-level passive δύναμις.

[33] For Aristotle's distinction between levels of actuality, see the texts cited in the previous note. He refers to the soul (an active δύναμις) as a first actuality at *An.* II. 412^{a}27–8 and 412^{b}4–6.

actuality' (417^b3–5).[34] Activity, the motion caused by an active δύναμις on a comparable passive δύναμις, as like on like, preserves the state that the agent is already in. The activity involved when a doctor responds to his medical art does not alter the agent but rather maintains his capacity (both active and passive) for such activity. On this view, the final cause—the goal to be achieved—is the activity that expresses and thereby preserves the agent's current condition.

The following picture emerges. Enforced changes, such as a person's recovery of health, result—often through a sequence of intervening motions—from a self-motion that is an activity of an unmoved mover and moved. Aristotle thinks that this account avoids an explanatory regress involving an infinite chain of typical movers, because the unmoved mover proper to the self-moving agent can explain why the change, though typically composed of distinct components, displays a sustained direction and terminates at a certain limit.

With this understanding of self-motion, let us now turn to the question I raised at the outset. If self-motions, as well as other motions, depend on an impetus in order to start, does this dependence result in an explanatory regress, and does it subvert the status of animals as genuine *self*-movers? Moreover, how does this dependence affect Aristotle's thesis in *Phys.* VIII. 5 that motions are ultimately caused by an unmoved mover?

Self-Motion, Heavenly Motion, and the Prime Mover

One of these questions seems easily answered. Even if self-motions fail to be fully explained by appeal to an internal cause, self-movers merit their special status because they contain an active source—a principle of motion and rest—and so, unlike commonplace objects that experience motion, they can sustain and direct their own, often intricate, behaviour. Even so, since the internal active principle cannot trigger the motion, the impetus derives from the creature's environment. Does this reliance on external factors undermine Aristotle's argument in *Phys.* VIII. 5?

[34] For a defence of the claim that the 'preservation' describes the activity rather than the switch to activity (as often assumed), see my *Aristotle on Substance*, 222–6. For the other view, see esp. L. A. Kosman, 'Aristotle's Definition of Motion', *Phronesis* 14 (1969), 40–62, esp. 54–6; and 'Substance, Being, and *Energeia*', *Oxford Studies in Ancient Philosophy* 2 (1984), 121–49, esp. 129–32.

I think not. Aristotle can preserve the main point of that argument by appealing to the mechanics of the cosmos as a whole. He locates the continuity and variety of sublunary changes in the motions of the heavenly spheres, and particularly in the twofold motion of the sun (*GC* II. 10, $336^{a}31$–$^{b}15$).[35] The sun's daily westward rotation with the sphere of the fixed stars and yearly eastward motion along the ecliptic, resulting in longer and shorter days, translates itself down to the sublunary sphere in the elemental change observed in seasonal variation.[36] Without the sun's distinctive motion, the four common elements would be permanently sorted into layers; all the earth would be settled in a ball at the centre and surrounded by all the water, then by all the air, and finally by all the fire (*GC* II. 10, $337^{a}7$–15). There would be no coming-to-be and passing-away.[37] As it is, there is abundant activity, with portions of the elements dislodged from their places and undertaking to regain them again. In their attempt to sort themselves out, they frequently interact, and such interaction sometimes results in the transformation of one element into another, and sometimes in the combining of elements into uniform compounds, such as minerals, natural metals, and various liquids.

In their continuous cyclical transformations the four elements are said to 'imitate' the eternal circular motion of the heavens, and these transformations guarantee the continuity and diversity of sublunary changes (*GC* II. 10, $337^{a}1$–7). So an animal is constantly subject to stimulation, whether by simple climatic reversals caused directly by the sun's motion or by other changes that the sun regulates more indirectly, such as animal generation itself (*Met.* Λ. 5, $1071^{a}14$–16) and length of life (*GC* II. 10, $336^{b}10$–15).[38] An animal responds to some stimulants and not others because of its nature (φύσις)—its inner active principle—which should also determine how the response is manifested.[39] Therefore, in explaining why an animal starts moving on

[35] Cf. *Met.* Λ. 6, $1072^{a}9$–17; *Phys.* VIII. 6, $258^{b}26$–$259^{a}6$.

[36] On the mechanics of this process, see *Meteor.* I. 3, $341^{a}13$–32; cf. *Cael.* II. 7, $289^{a}28$–34.

[37] *Met.* Λ. 6, $1072^{a}10$–12; *Cael.* II. 3, $286^{b}1$–9; *GC* II. 10, $336^{a}31$–$^{b}10$.

[38] Cf. *GA* IV. 10, $777^{b}16$–$778^{a}9$. I thank Allan Gotthelf for bringing this passage to my attention.

[39] On the role of an animal's nature in establishing its length of life, see *GA* IV. 10, $777^{b}6$–8; cf. *Long.* 5, $466^{a}17$–$^{b}4$. A creature's nature also controls its response to external factors, for instance, whether it lives or dies in air or water; see *Resp.* 14, $477^{b}31$–$478^{a}4$; 16, $478^{a}29$–35; 19, $479^{b}8$–13. I thank James Lennox for calling my attention to these passages.

a particular occasion, one need not track down an infinite number of previous changes; instead one can appeal, first, to the animal's nature, which determines why a stimulus of a particular sort arouses a response of a particular kind, and second, to the sun's motion, which accounts for the patterns of interaction in the sublunary region.

Having elevated the explanation of animal motion to the superlunary realm, we should now ask what accounts for the sun's distinctive course. Unfortunately, Aristotle says relatively little about the mechanics of celestial motion, and some of his claims seem inconsistent with one another. In particular, he appears to vacillate on the question of whether the heavenly spheres are self-movers or not. In *De Caelo* he asserts that the heaven is alive (ἔμψυχος) and compares the motions of some of its parts to those of animals.[40] In *Met. Λ*. 7 he claims that the Prime Mover causes motion by being desired or loved (1072^b3). If the spheres are moved by their desire, they should be ensouled. These claims suggest that the spheres resemble ordinary self-movers in possessing both internal active and passive sources of motion.

Yet on several occasions Aristotle insists that the mover responsible for the eternal rotation of the outermost sphere is external to it,[41] and *De Motu Animalium*, after arguing that the heaven is moved by an external unmoved mover, contrasts the heaven with animals, which require, in addition to an external unmoved factor, an internal unmoved component as well (700^a6–11).[42] Since the author applauds certain theorists who deny that any part of a rotating sphere can remain fixed and gives as his reason that otherwise the whole sphere would have to remain still or be deprived of its continuity (699^a17–20), he apparently thinks that one cannot attribute to the heavenly spheres both internal moving and moved components without undermining their intrinsic unity.[43] These claims suggest that the sphere of the fixed

[40] *Cael.* II. 2, 285^a29–30; II. 12, 292^a18–$^b25$. But see n. 44 below.

[41] *Phys.* VIII. 10, 267^b6–9; *MA* 3–4; *Met. Λ*. 7, 1073^a3–5.

[42] Martha Nussbaum, *Aristotle's De Motu Animalium*, 3–12, has argued for the authenticity of this treatise. Even if one questions its genuineness, the text is presumably an early Peripatetic work and the main tenets Aristotelian.

[43] Cf. *Phys.* VIII. 4, 255^a12–18. Although this passage primarily concerns the four sublunary elements, whose natural unity and continuity prevent the distinction between moving and moved components—which would enable their classification as self-movers—Aristotle extends the claim to other continuous bodies as well. This passage, together with the one in *MA*, suggests that the continuity of the heavenly spheres prevents them from being self-movers.

stars, and perhaps the lower spheres as well, has only an intrinsic passive source of motion.[44]

The motions of the heavenly spheres might therefore be better compared to the natural motions of the four sublunary elements than to the self-motions of animals.[45] According to *De Caelo*, the spheres are composed of a fifth element, aether.[46] Like the simple bodies in the lower region, whose natures enable them to move upward toward the sphere of the moon or downward toward the cosmic centre, the celestial element naturally rotates around the centre (*Cael.* I. 2). The nature of aether apparently accounts for the circular motion of the heavenly spheres in the way that the natures of earth, water, air, and fire account for the natural motions of these bodies. Before pressing

[44] The thesis in *Met. Λ.* 7, that the spheres are moved by their desire, might accord with this possibility, because Aristotle views the faculty of desire as a passive capacity of soul—a moved mover that responds to an unmoved mover (*An.* III. 10, 433^b^13–18; *MA* 6, 700^b^32–701^a^1). The crucial question is whether the account in *Met. Λ* requires that the spheres resemble animals in possessing an internal active δύναμις as well. Since *Met. Λ* does not mention an internal active principle, it could be that he regards the spheres as ensouled, in virtue of their special passive capacity, yet not as self-movers, because they lack an internal directive δύναμις. One might object that Aristotle's silence in *Met. Λ* tells against rather than for this suggestion, since the proposed disanalogy between the spheres and animals is sufficiently important to merit clarification. But since the disanalogy is mentioned in *MA* 4 (700^a^6–11, cited above), it seems more likely that *Met. Λ* is silent because such internal ἀρχαί are not part of the theory. The passages cited above from *De Caelo* can also be reconciled with this proposal. From the claim in II. 2 that the heaven is alive and has an inner source of motion, Aristotle concludes that it has various parts (upper and lower, right and left). This inference suggests that οὐρανός here refers to the cosmos as a whole rather than to the outermost sphere (cf. *Cael.* I. 9, 278^b^9–21 on the various meanings of οὐρανός). If so, this passage does not bear on our question about the spheres (though it is quite Platonic in suggesting that the whole cosmos resembles an animal). The passage in II. 12 is not about the spheres either, but about the heavenly bodies. The claim that they partake of action and life (292^a^18–22) does not conflict with the proposal that the spheres have only passive souls. Furthermore, the analogy between the planets and animals (292^a^20–^b^25) concerns only the number of motions required for them to attain the good; nothing is said about the presence of an inner active principle.

[45] For this proposal, cf. James Bogen and J. E. McGuire, 'Aristotle's Great Clock: Necessity, Possibility and the Motion of the Cosmos in *De Caelo* I. 12', *Philosophy Research Archives* 12 (1986–7), 387–448.

[46] Although Aristotle says a good deal about the fifth element in *De Caelo*, he is curiously reticent about it in *Phys.* VIII and *Met. Λ*. Even so, the heavenly spheres are made out of some material, and *Met. Θ*. 8 carefully distinguishes the incorruptible matter of the heavenly bodies, which enables them to move from one location to another (ποθὲν ποί), from the matter of bodies in the sublunary region, which accounts (among other things) for their perishability (1050^b^6–28). A similar distinction occurs at *Λ*. 2, 1069^b^24–6. In addition, *Λ*, 8 mentions τὸ κύκλῳ σῶμα, which is eternal and unresting (1073^a^31–2). Given these passages, I assume that the doctrine of a fifth element, which naturally rotates, is presupposed in the relevant discussions in the *Physics* and *Metaphysics*.

the analogy further, however, we need to untangle Aristotle's account of elemental natural motion, which is itself problematic.

In *Phys.* VIII. 4 Aristotle denies that the elements are self-movers, arguing that although they possess a passive source, which accounts for why, if unimpeded, they move in one direction rather than another, they lack an internal active principle (255^b29–31). The active source lies outside. VIII. 4 lists various items responsible for elemental natural motion, including that which generates the element or removes an impediment to its natural progression; but these causes are rightly identified as accidental (255^b24–9), because they merely trigger the motion but do not explain why it is sustained in a particular direction or terminates at a certain limit. We are left asking what serves as the active principle of elemental natural motion comparable to that in other typical changes.

Cael. IV. 3 takes up the question again,[47] and here Aristotle says that even more than things that change on their own in response to slight stimulation, the elements seem to possess an internal principle. Yet he does not reject the conclusion of *Phys.* VIII. Instead he makes a remarkable claim: the elements appear to possess an internal source because, he says, 'their matter is closest to substance' (310^b31–3). If my earlier suggestion is granted, that an active principle integrates distinct, and often successive, events into a complex motion that promotes a higher goal, the elements should need no internal active source because they display only a single simple behaviour, which requires no co-ordination. Elemental matter is 'closest to substance' because the elements, if unimpeded, automatically exercise their δύναμις in seeking their proper places.[48] They are not genuine substances, however, because, in lacking an internal active source, they cannot limit their natural motions at their natural termini. Fire is not programmed to stop at the periphery—it would proceed upward indefinitely if it were not confined by the sphere of the moon;[49] and water would progress downward indefinitely were it not eventually stopped by the heavier mass of earth below.

Aristotle claims in *Cael.* IV. 3 that the motion of an element to its proper place is motion 'to its own form' (310^a33–$^b1$); and he identifies

[47] At the end of the chapter (311^a9–12) Aristotle mentions the accidental causes listed in *Phys.* VIII. 4 and refers to his earlier discussion.

[48] *Phys.* VIII. 4, 255^b5–12.

[49] Recall Aristotle's claim in *An.* II. 4 that fire would burn 'without limit' as long as it had stuffs to consume.

the boundaries—the periphery and the centre—as in some sense the form of the body contained at that location (310^b7–12); he further states that the adjacent body above stands to the one below 'as form to matter' (310^b14–15). These pronouncements suggest that whatever limits an element's natural motion—the centre in the case of earth, the sphere of the moon in the case of fire, and the accumulated mass of the adjacent element in the case of water and air—counts as the element's form, and thus plays the limiting role of an active δύναμις for the simple bodies.

I have suggested that the aetherial spheres may be only natural movers, like the four elements, rather than self-movers. If so, they possess an inner passive source, which accounts for their rotation, but the governing principle lies outside the sphere that it directs. A passage in *Phys.* VIII. 6 supports this proposal. Aristotle claims that the ἀρχαί of the heavenly bodies that experience more than one motion are moved accidentally by something else, while those of perishable things alone are moved accidentally by themselves (259^b28–31). This division excludes the ἀρχή of the outermost sphere, because the stars display one motion only; and he has already argued that the first mover responsible for the eternal motion is unmoved even incidentally (259^b20–8). Since the first mover is immune even to accidental motion, it lies outside the entire rotating system. In VIII. 10 Aristotle locates the first mover on the circumference of the cosmos (267^b6–9). Although the location of the ἀρχαί of the lower spheres is more problematic, this passage at least establishes that the principle of each sphere is situated within the rotating system but outside its proper sphere. For if the lower spheres each possessed an internal active source, that source would be moved accidentally by itself, like that of animals, because it would be carried around with its own sphere. Since he claims instead that the principle in each case (excepting the stars) is moved accidentally by something else, the ἀρχή must inhabit some sphere (to be moved accidentally) yet not the sphere of which it is the active principle. The ἀρχαί of the lower spheres could be located in the sphere of the stars (and so be carried around with it),[50] or more likely, the principle of each sphere could occupy the next sphere above.[51] Either way,

[50] Aristotle thinks that all the spheres exhibit the motion of the outermost sphere as well as their own; see *Met.* Λ. 6, 1072^a10–17; cf. Λ. 8, 1073^b25–6 (on Eudoxus' theory). Cf. Simplicius, *in Phys.* 1261, 17–19.

[51] I owe this suggestion to Steven Strange.

the active principle is external to the sphere it moves.[52]

On this interpretation, the celestial spheres resemble the four sublunary elements, with their simple rectilinear motions, in moving in simple concentric circles around the earth. The apparently complex motions of the sun, moon, and planets are explained by a combination of homocentric spheres, each rotating in a westward or eastward direction around its own axis.[53] Since Aristotle attributes an ἀρχή to each sphere, not to the luminary whose motion is the product of several spheres (*Met. Λ.* 8, 1074^a14–16), the active δύναμις does not co-ordinate motions but need only determine the inclination of a sphere's axis, its proper orientation (to west or east), and the speed of its rotation. To control these factors, the ἀρχή of each sphere should be located in the next above, at the point at which the axis of the lower sphere makes contact with its higher neighbour. The Prime Mover, located outside the rotating system, should account for the axis of rotation of the outermost sphere, its westward orientation and period.[54] But as we shall see, the Prime Mover must also perform a more vital role.

Although the evidence strongly suggests that the celestial spheres are natural movers rather than self-movers, for my general argument the decision one way or the other is not crucial. Since the motion of the celestial spheres persists eternally, the phenomena can be

[52] Since the ἀρχή occupies a sphere other than the one it directs, it is accidental to the sphere in which it resides (because it is not part of its defining account). For the idea that an active principle can be present in matter accidentally, we might usefully compare Aristotle's account of the transmission of soul in animal generation. Although the semen and then the inseminated female material serve as vehicles of soul, the soul is not the actuality *of* these materials. (Only once the heart of the new creature has been articulated is there a body to which the soul belongs *per se*.) On this topic, see Alan Code, 'Soul as Efficient Cause in Aristotle's Embryology', *Philosophical Topics* 15 (1987), 51–9, esp. 56–8.

[53] Aristotle's account in *Met. Λ.* 8 appears to transform a geometrical theory of homocentric spheres, which Eudoxus and Callippus used to explain the motions of the sun, moon, and planets, into a mechanical system of spheres in contact with one another. According to Aristotle, Eudoxus and Callippus invoked a set of spheres to explain the motions of each heavenly body and included in each case the sphere of the stars. Aristotle claims that additional counteracting spheres are required if all the spheres combined are to explain the phenomena (1073^b38–1074^a1). The additional spheres are thus introduced to establish a mechanical system; they counteract the motions proper to each planet so that the motion of the outermost sphere alone is transferred to the heavenly body below. On this topic I have profited from reading Thomas Heath, *Aristarchus of Samos: The Ancient Copernicus* (New York, 1981), 190–224; and Thomas S. Kuhn, *The Copernican Revolution* (Cambridge, Mass., 1957), 55–9.

[54] For the dependence of the outermost heaven on the first mover, see *Met. Λ.* 8, 1073^a23–5.

explained by treating the spheres either as animals, whose active δύναμις is internal, or as inanimate objects, like the four elements, whose active source lies outside. On either conception, however, a sphere's rotation should be an activity—an ἐνέργεια—the mutual expression of a mover and moved whose active and passive δυνάμεις are like.[55] In rotating, a sphere does not acquire a new location but moves continually from and to the same location.[56] By means of this constant activity, the spheres maintain their position within the cosmic hierarchy. If in fact Aristotle thinks that the active principles occupy their own spheres, the celestial self-movers will independently guard their proper places within the system. The question, then, is why these autonomous agents adapt themselves to a co-operative venture that benefits the cosmos as a whole. The Prime Mover will play this organizational role. If instead, as I have argued, each sphere depends for its orientation on the next sphere above, the nested spheres are parts of a mechanical network. But on this construction, too, the Prime Mover is needed to explain why the celestial spheres are so adapted to foster the continued life and interaction of bodies in the lower region.

In *Met. Λ*. 10 Aristotle asks whether the good of the whole lies in the order of its parts or in something separate by itself—whether, as in an army, the good lies in its organization or in its leader. He claims that the good lies in both but more in the leader, because the order depends on him (1075ᵃ11–15).[57] The Prime Mover must integrate the proceedings into an organized system that promotes the good of all its members (1075ᵃ16–25). Like an animal's active δύναμις, which regulates simpler motions to serve a higher purpose, the Prime Mover, too, is an efficient cause, controlling the motions of the universe as a whole for the sake of the whole operation. On this interpretation, the good of the cosmos—its οὗ ἕνεκα—lies both in the Prime Mover

[55] Waterlow, *Nature, Change, and Agency*, 249–57, is concerned that eternal rotation fails to satisfy Aristotle's definition of change in *Phys.* III. 1 (recalled at VIII. 1, 251ᵃ9–10 and 5, 257ᵇ8–9), because such motion is a complete rather than an incomplete actuality. She argues that it is a change none the less because it has a distinct agent (pp. 254–5). On my view, the distinction between change and activity does not turn on the presence or absence of a mover: both require a mover, and in either case the mover can be internal or external. So, for instance, when an axe is used to chop, its behaviour is an activity even though the mover is external. If this analysis is correct, Waterlow's criterion is inadequate to establish eternal rotation as a κίνησις. Of course, such motion need not satisfy the definition of change if, as I am suggesting, it is an ἐνέργεια. On this topic, cf. Bogen and McGuire, 'Aristotle's Great Clock', 417–22.

[56] *Phys.* VIII. 8, 264ᵇ9–19; *Cael.* I. 4, 271ᵃ19–22.

[57] Cf. *GC* II. 10, 337ᵃ16–22.

himself, whose power is expressed in his eternal activity, and in the functioning system, whose order and continuity he maintains by means of the regular motions of bodies acting according to their natures.[58]

[58] This double meaning of οὗ ἕνεκα to specify, on the one hand, the good for the sake of which (τὸ οὗ ἕνεκα τινός) and, on the other, the beneficiary for the sake of which (τὸ οὗ ἕνεκα τινί) is pointed out by Aristotle on several occasions; see *Met.* Λ. 7, 1072^b1–4; *An.* II. 4, 415^b2–3; 415^b20–1; *Phys.* II. 194^a35–6. I presented versions of this paper at a conference entitled 'Self-Motion: From Aristotle to Newton' at the University of Pittsburgh, and at a session of Alan Code's seminar on Aristotle's Metaphysics at the University of California, Berkeley. I thank those audiences for their helpful questions. I am particularly grateful to David Furley for his challenging commentary at the Pittsburgh conference, to Lindsay Judson for his perceptive objections, which caused me to rethink my argument at a number of points, and to James Lennox for many valuable criticisms. I am also indebted to Susan Sauvé, Heike Sefrin-Weis, and Steven Strange for suggestions on particular points.

BIBLIOGRAPHY

The bibliography contains all works cited, as well as other important work on the subjects discussed in the essays in this volume.

ACKRILL, J. L., *Aristotle the Philosopher* (Oxford, 1981).

ALLAN, D. J., 'Causality, Ancient and Modern', *Proceedings of the Aristotelian Society* Supplementary Vol. 39 (1965), 1–18.

ANNAS, JULIA, 'Aristotle, Number and Time', *Philosophical Quarterly* 25 (1975), 97–113.

— 'Aristotle on Inefficient Causes', *Philosophical Quarterly* 32 (1982), 311–26.

— 'Die Gegenstände der Mathematik bei Aristoteles', in Graeser (ed.), *Mathematics and Metaphysics in Aristotle*.

ARMSTRONG, D. M., *What is a Law of Nature?* (Cambridge, 1983).

BAILEY, CYRIL, *The Greek Atomists and Epicurus: A Study* (Oxford, 1928).

BALME, D. M., 'Greek Science and Mechanism I: Aristotle on Nature and Chance', *Classical Quarterly* 33 (1939), 129–38.

— 'Development of Biology in Aristotle and Theophrastus: Theory of Spontaneous Generation', *Phronesis* 7 (1962), 91–104.

— 'Aristotle's Use of Teleological Explanation', Inaugural lecture, Queen Mary College, University of London (London, 1965).

— 'Teleology and Necessity', in Gotthelf and Lennox (eds.), *Philosophical Issues in Aristotle's Biology*.

BARKER, ANDREW, *Greek Musical Writings*, ii. *Harmonic and Acoustic Theory* (Cambridge, 1989).

BARNES, JONATHAN, 'Aristotle's Theory of Demonstration', *Phronesis* 14 (1969), 123–52; reprinted in revised form in Barnes, Schofield, and Sorabji (eds.), *Articles on Aristotle* i. *Science*.

— *Aristotle's Posterior Analytics, Translated with Notes*, Clarendon Aristotle Series (Oxford, 1975).

— *The Presocratic Philosophers*, 2 vols. (London, 1979).

— *Aristotle* (Oxford, 1982).

— (ed.), *The Complete Works of Aristotle: The Revised Oxford Translation*, 2 vols. (Princeton, 1984).

— SCHOFIELD, MALCOLM and SORABJI, RICHARD (eds.), *Articles on Aristotle*, 4 vols. (London, 1975–9).

BERTI, ENRICO (ed.), *Aristotle on Science: The Posterior Analytics*, Proceedings of the Eighth Symposium Aristotelicum (Padua, 1981).

BOGEN, J. and MCGUIRE, J. E., 'Aristotle's Great Clock: Necessity, Possibility and the Motion of the Cosmos in *De Caelo* I. 12', *Philosophy Research Archives* 12 (1986–7), 387–448.

BOLTON, ROBERT, 'Essentialism and Semantic Theory in Aristotle: *Posterior Analytics* II. 7–10', *Philosophical Review* 85 (1976), 514–44.

—— 'Definition and Scientific Method in Aristotle's *Posterior Analytics* and *Generation of Animals*', in Gotthelf and Lennox (eds.), *Philosophical Issues in Aristotle's Biology*.

—— 'The Epistemological Basis of Aristotelian Dialectic', in D. Devereux and P. Pellegrin (eds.), *Biologie, logique et metaphysique chez Aristote* (Paris, 1990).

—— 'Aristotle's Conception of Metaphysics as a Science', in David Charles *et al.* (eds.), *Aristotle's Metaphysics Today* (forthcoming).

BOSTOCK, DAVID, 'Aristotle, Zeno and the Potential Infinite', *Proceedings of the Aristotelian Society* 73 (1972–3), 37–51.

—— 'Plato on Change and Time in the *Parmenides*', *Phronesis* 23 (1978), 229–42.

—— *Logic and Arithmetic*, ii. *Rational and Irrational Numbers* (Oxford, 1979).

—— 'Aristotle's Account of Time', *Phronesis* 25 (1980), 148–69.

—— 'Aristotle on the Principles of Change in *Physics* I', in Schofield and Nussbaum (eds.), *Language and Logos*.

—— 'Time and the Continuum', *Oxford Studies in Ancient Philosophy* 6 (1988), 255–70.

BRADIE, MICHAEL and MILLER, FRED D. 'Teleology and natural necessity in Aristotle', *History of Philosophy Quarterly* 1 (1984), 133–46.

BRENTANO, FRANZ, *Philosophical Investigations on Space, Time and the Continuum*, trans. Barry Smith (London, 1988).

BRODY, B. A., 'Towards an Aristotelian Theory of Scientific Explanation', *Philosophy of Science* 39 (1972), 20–31.

BURNET, J., *Early Greek Philosophy*, 4th edn. (London, 1930).

BURNYEAT, MYLES, 'Aristotle on Understanding Knowledge', in Berti (ed.), *Aristotle on Science: The Posterior Analytics*.

CALLAHAN, J. F., *Four Views of Time in Ancient Philosophy* (Cambridge, Mass., 1948).

CARTERON, HENRI, *La Notion de force dans le système d'Aristote* (Paris, 1923; reprinted New York and London, 1979); part of the Introduction is translated under the title, 'Does Aristotle have a Mechanics?' in Barnes, Schofield, and Sorabji (eds.), *Articles on Aristotle*, i. *Science*.

CHARLES, DAVID, *Aristotle's Philosophy of Action* (London and Ithaca, NY, 1984).

—— 'Aristotle on Hypothetical Necessity and Irreducibility', *Pacific Philosophical Quarterly* 69 (1988), 1–53.

CHARLTON, WILLIAM, *Aristotle's Physics, Books I and II: Translated with Introduction and Notes*, Clarendon Aristotle Series (Oxford, 1970).

CLARK, STEPHEN, R. L., *Aristotle's Man: Speculations upon Aristotelian Anthropology* (Oxford, 1975).

CODE, A., 'Aristotle's Response to Quine's Objections to Modal Logic', *Journal of Philosophical Logic* 5 (1976), 159–86.

—— 'The Persistence of Aristotelian Matter', *Philosophical Studies* 29 (1976), 357–67.

—— 'Soul as Efficient Cause in Aristotle's Embryology', *Philosophical Topics* 15 (1987), 51–9.

COHEN, MORRIS R. and DRABKIN, I. E., *A Source Book in Greek Science* (Cambridge, Mass., 1948; reprinted 1958).

COOPER, JOHN M., 'Aristotle on Natural Teleology', in Schofield and Nussbaum (eds.), *Language and Logos*.

—— 'Hypothetical Necessity', in Gotthelf (ed.), *Aristotle on Nature and Living Things*.

CORISH, D., 'Aristotle's Attempted Derivation of Temporal Order from that of Movement and Space', *Phronesis* 21 (1976), 241–51.

DAVIDSON, DONALD, 'Causal Relations', in *Essays on Actions and Events* (Oxford, 1980).

DESCARTES, RENÉ, *Œuvres*, 12 vols., ed. C. Adam and P. Tannery (Paris, 1897–1912; reprinted 1964).

DICKS, D. R., *Early Greek Astronomy to Aristotle* (London, 1970).

DRABKIN, I. E., 'Notes on the Laws of Motion in Aristotle', *American Journal of Philology* 59 (1938), 60–84.

DRISCOLL, J., 'The Platonic Ancestry of Primary Substance', *Phronesis* 24 (1979), 253–69.

DÜRING, INGEMAR (ed.), *Naturphilosophie bei Aristoteles und Theophrast*, Proceedings of the Fourth Symposium Aristotelicum (Heidelberg, 1969).

EDMUNDS, LOWELL, 'Necessity, Chance and Freedom in the Early Atomists', *Phoenix* 26 (1972), 342–57.

ENGBERG-PEDERSEN, T., 'More on Aristotelian *Epagoge*', *Phronesis* 24 (1979), 301–19.

EVANS, M. G., 'Causality and Explanation in the Logic of Aristotle', *Philosophy and Phenomenological Research* 19 (1958–9), 466–85.

FINE, GAIL, 'Aristotle on Determinism: A Review of Richard Sorabji's *Necessity, Cause and Blame*', *Philosophical Review* 90 (1981), 561–79.

—— 'Forms as Causes: Plato and Aristotle', in Graeser (ed.), *Mathematics and Metaphysics in Aristotle*.

FREDE, DOROTHEA, 'Aristotle on the Limits of Determinism: Accidental Causes in *Metaphysics E*. 3', in Gotthelf (ed.), *Aristotle on Nature and Living Things*.

FREDE, MICHAEL, 'The Original Notion of Cause', in Malcolm Schofield, Myles Burnyeat, and Jonathan Barnes (eds.), *Doubt and Dogmatism: Studies in Hellenistic Epistemology* (Oxford, 1980); reprinted in Michael Frede, *Essays in Ancient Philosophy* (Oxford, 1987).

FREELAND, CYNTHIA A., 'Aristotelian Actions', *Noûs* 19 (1985), 397–414.

—— 'Aristotle on Possibilities and Capacities', *Ancient Philosophy* 6 (1986), 69–89.

—— 'Scientific Explanation and Empirical Data in Aristotle's *Meteorology*', *Oxford Studies in Ancient Philosophy* 8 (1990), 67–102.

FRIEDMAN, MICHAEL, 'Explanation and Scientific Understanding', *Journal of Philosophy* 71 (1974), 5–19.

FRITZ, K. VON, 'Das ἄπειρον bei Aristoteles', in Düring (ed.), *Naturphilosophie bei Aristoteles und Theophrast*.

— 'Die ἐπαγωγή bei Aristoteles', in von Fritz, *Grundprobleme der Geschichte der antiken Wissenschaft* (Berlin, 1971).

FURLEY, DAVID, 'Aristotle and the Atomists on Infinity', in Düring (ed.), *Naturphilosophie bei Aristoteles und Theophrast*; reprinted in Furley, *Cosmic Problems*.

— 'Self Movers', in G. E. R. Lloyd and G. E. L. Owen (eds.), *Aristotle on Mind and the Senses*, Proceedings of the Seventh Symposium Aristotelicum (Cambridge, 1978); reprinted in Furley, *Cosmic Problems*.

— 'The Greek Commentators' Treatment of Aristotle's Theory of the Continuous', in Kretzmann (ed.), *Infinity and Continuity*.

— 'The Rainfall Example in *Physics* II. 8', in Gotthelf (ed.), *Aristotle on Nature and Living Things*; reprinted in Furley, *Cosmic Problems*.

— *Cosmic Problems: Essays on Greek and Roman Philosophy of Nature* (Cambridge, 1989).

GAISER, KONRAD, 'Das Zweifache Telos bei Aristoteles', in Düring (ed.), *Naturphilosophie bei Aristoteles und Theophrast*.

GEACH, P. T., 'What Actually Exists', *Proceedings of the Aristotelian Society* Supplementary Vol. 42 (1968), 7–16.

GILL, MARY LOUISE, 'Aristotle's Theory of Causal Action in *Physics* III. 3', *Phronesis* 25 (1980), 129–47.

— 'Sorabji and Aristotle against Determinism', *Ancient Philosophy* 2 (1982), 122–33.

— 'Aristotle on the Individuation of Changes', *Ancient Philosophy* 4 (1984), 9–22.

— *Aristotle on Substance: The Paradox of Unity* (Princeton, 1989).

GLYMOUR, C., 'Explanation and Realism', in Paul M. Churchland and Clifford A. Hooker (eds.), *Images of Science: Essays on Realism and Empiricism, with a reply from Bas C. van Fraassen* (Chicago, 1985).

GOTTHELF, ALLAN, 'Aristotle's Conception of Final Causality', *Review of Metaphysics* 30 (1976–7), 226–54; reprinted with a Postscript in Gotthelf and Lennox (eds.), *Philosophical Issues in Aristotle's Biology*.

— 'The Place of the Good in Aristotle's Natural Teleology', in John J. Cleary and D. C. Shartin (eds.), *Proceedings of the Boston Area Colloquium in Ancient Philosophy* iv: 1988 (Lanham, Md., 1989).

— 'Teleology and Spontaneous Generation in Aristotle: a Discussion', in Terry Penner and Richard Kraut (eds.), *Nature, Knowledge and Virtue: Essays in Memory of Joan Kung*, *Apeiron* 22.4 (December 1989).

— (ed.), *Aristotle on Nature and Living Things: Philosophical and Historical Studies Presented to David M. Balme on his Seventieth Birthday* (Pittsburgh and Bristol, 1985).

— and LENNOX, JAMES G. (eds.), *Philosophical Issues in Aristotle's Biology* (Cambridge, 1987).

GRAESER, ANDREAS (ed.), *Mathematics and Metaphysics in Aristotle*, Proceedings of the Tenth Symposium Aristotelicum (Bern, 1987).

GRAHAM, DANIEL W., *Aristotle's Two Systems* (Oxford, 1987).

GRANT, E., *Physical Science in the Middle Ages* (Cambridge, 1977).

GUTHRIE, W. K. C., *Aristotle: On the Heavens*, Loeb Classical Library (London and Cambridge, Mass., 1939).

— *A History of Greek Philosophy*, 6 vols. (Cambridge, 1962–81).

HADGOPOULOS, DEMETRIUS J., 'A Note on Aristotle's Notions of Universality and Necessity', *Logique et Analyse* 18 (1975), 171–4.

HAMLYN, D. W., 'Aristotelian *Epagoge*', *Phronesis* 21 (1976), 167–84.

HARDIE, R. P. and GAYE, R. K., *Physica*, in W. D. Ross (ed.), *The Works of Aristotle Translated into English*, ii (Oxford, 1930).

HARRÉ, ROM, *The Principles of Scientific Thinking* (London, 1970).

HEATH, D. D., 'On some Misconceptions of Aristotle's Doctrine on Causation and τὸ αὐτόματον', *Journal of Philology* 7 (1877), 97–115.

HEATH, T. L., *Aristarchus of Samos: The Ancient Copernicus* (Oxford, 1913; reprinted in New York, 1981).

— *A History of Greek Mathematics*, 2 vols. (Oxford, 1921; reprinted Oxford, 1960 and New York, 1981).

— *Mathematics in Aristotle* (Oxford, 1949).

HEINAMAN, ROBERT, 'Aristotle on Accidents', *Journal of the History of Philosophy* 23 (1985), 311–24.

HINTIKKA, JAAKKO, *Time and Necessity: Studies in Aristotle's Theory of Modality* (Oxford, 1973).

HOCUTT, MAX, 'Aristotle's Four Becauses', *Philosophy* 49 (1974), 385–99.

HUSSEY, EDWARD, *Aristotle's Physics, Books III and IV: Translated with Notes*, Clarendon Aristotle Series (Oxford and New York, 1983).

IRWIN, T. H., 'Aristotle's Concept of Signification', in Schofield and Nussbaum (eds.), *Language and Logos*.

— *Aristotle's First Principles* (Oxford, 1988).

JOACHIM, HAROLD H., *Aristotle: On Coming-to-Be and Passing Away: A Revised Text with Introduction and Commentary* (Oxford, 1922).

JUDSON, LINDSAY, 'Eternity and Necessity in *De Caelo* I. 12', *Oxford Studies in Ancient Philosophy* 1 (1983), 217–55.

KAHN, CHARLES H., 'The Role of *Nous* in the Cognition of First Principles in *Posterior Analytics* II. 19', in Berti (ed.), *Aristotle on Science: The Posterior Analytics*.

— 'The Place of the Prime Mover in Aristotle's Teleology', in Gotthelf (ed.), *Aristotle on Nature and Living Things*.

KIM, JAEGWON, 'Explanatory Realism, Causal Realism, and Explanatory Exclusion', in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (eds.), *Realism and Antirealism*, Midwest Studies in Philosophy 12 (Minneapolis, 1988).

KIRWAN, CHRISTOPHER, *Aristotle's Metaphysics, Books Γ, Δ, and E: Translated with Notes*, Clarendon Aristotle Series (Oxford, 1971).

KOSMAN, L. A., 'Aristotle's Definition of Motion', *Phronesis* 14 (1969), 40-62.

—— 'Substance, Being, and *Energeia*', *Oxford Studies in Ancient Philosophy* 2 (1984), 121-49.

—— 'Animals and Other Beings in Aristotle', in Gotthelf and Lennox (eds.), *Philosophical Issues in Aristotle's Biology*.

KRETZMANN, NORMAN, 'Aristotle on the Instant of Change', *Proceedings of the Aristotelian Society* Supplementary Vol. 50 (1976), 91-114.

—— (ed.), *Infinity and Continuity in Ancient and Medieval Thought* (Ithaca, NY and London, 1982).

KUHN, THOMAS S., *The Copernican Revolution* (Cambridge, Mass., 1957).

KULLMAN, WOLFGANG, 'Different Concepts of the Final Cause in Aristotle', in Gotthelf (ed.), *Aristotle on Nature and Living Things*.

KUNG, JOAN, 'Aristotle on Essence and Explanation', *Philosophical Studies* 31 (1977), 361-83.

LANDELS, J. G., *Engineering in the Ancient World* (London, 1978).

LEAR, JONATHAN, 'Aristotelian Infinity', *Proceedings of the Aristotelian Society* 80 (1979-80), 187-210.

—— 'A Note on Zeno's Arrow', *Phronesis* 26 (1981), 91-104.

—— 'Aristotle's Philosophy of Mathematics', *Philosophical Review* 91 (1982), 161-92.

—— *Aristotle: The Desire to Understand* (Cambridge, 1988).

LE BLOND, J. M., *Logique et méthode chez Aristote: Étude sur la recherche des principes dans la physique aristotélicienne*, 2nd edn. (Paris, 1970).

LEIBNIZ, GOTTFRIED WILLHELM, *Die Philosophischen Schriften*, 7 vols., ed. C. I. Gerhardt (Berlin, 1875-90; reprinted Hildesheim, 1965).

LENNOX, JAMES G., 'Teleology, Chance, and Aristotle's Theory of Spontaneous Generation', *Journal of the History of Philosophy* 20 (1982), 219-38.

—— 'Aristotle on Chance', *Archiv für Geschichte der Philosophie* 66 (1984), 52-60.

LEWIS, FRANK A., 'Accidental Sameness in Aristotle', *Philosophical Studies* 42 (1982), 1-36.

—— 'Teleology and Material/Efficient Causes in Aristotle', *Pacific Philosophical Quarterly* 69 (1988), 54-98.

LLOYD, G. E. R., *The Revolutions of Wisdom: Studies in the Claims and Practice of Ancient Greek Science* (Berkeley, Los Angeles, and London, 1987).

LÖBL, R., *Demokrits Atome* (Bonn, 1976).

LONGINO, HELEN E., 'Evidence and Hypothesis: An Analysis of Evidential Relations', *Philosophy of Science* 46 (1979), 35-56.

MCCARTHY, TIMOTHY, 'On an Aristotelian Model of Scientific Explanation', *Philosophy of Science* 44 (1977), 159-66.

MACHAMER, PETER K., 'Aristotle on Natural Place and Natural Motion', *Isis* 69 (1978), 377-87.

McGuire, J. E., 'Philoponus on *Physics* ii 1: *Φύσις, Δύναμις*, and the Motion of Simple Bodies', *Ancient Philosophy* 5 (1985), 241–67.

— and Strange, Stephen K., 'An Annotated Translation of Plotinus, *Ennead* iii. 7: *On Eternity and Time*', *Ancient Philosophy* 8 (1988), 251–71.

McKirahan, R., 'Aristotelian *Epagoge* in *Prior Analytics* II. 21 and *Posterior Analytics* I. 1', *Journal of the History of Philosophy*, 21 (1983), 1–13.

McTaggart, J. McT. E., 'The Unreality of Time', *Mind* 17 (1908), 457–74.

— *The Nature of Existence*, 2 vols. (Cambridge, 1921 and 1927).

Madigan, Arthur, '*Metaphysics E*. 3: A Modest Proposal', *Phronesis* 29 (1984), 123–36.

Mansion, Augustin, *Introduction à la physique aristotélicienne*, 2nd edn. (Louvain and Paris, 1946).

Mansion, Suzanne (ed.), *Aristote et les problèmes de méthode*, Proceedings of the Second Symposium Aristotelicum (Louvain, 1961).

Matthews, Gareth B., 'Accidental Unities', in Schofield and Nussbaum (eds.), *Language and Logos*.

Mignucci, Mario, '*Ὡς ἐπὶ τὸ πολύ* et nécessaire dans la conception aristotélicienne de la science', in Berti (ed.), *Aristotle on Science: The Posterior Analytics*.

— 'Puzzles about Identity: Aristotle and his Greek Commentators', in Jürgen Wiesner (ed.), *Aristoteles: Werk und Wirkung*, i. *Aristoteles und seine Schule* (Berlin and New York, 1985).

Miller, Fred D., 'Did Aristotle have the Concept of Identity?', *Philosophical Review* 82 (1973), 483–90.

— 'Aristotle on the Reality of Time', *Archiv für Geschichte der Philosophie* 56 (1974), 132–55.

— 'Aristotle Against the Atomists', in Kretzmann (ed.), *Infinity and Continuity*.

Millikan, Ruth, *Language, Thought and other Biological Categories: New Foundations for Realism* (London, 1984).

Modrak, D. K. W., 'Aristotle on the Difference between Mathematics and Physics and First Philosophy', in Terry Penner and Richard Kraut (eds.), *Nature, Knowledge and Virtue: Essays in Memory of Joan Kung*, *Apeiron* 22.4 (December 1989).

Moore, A. W., *The Infinite* (London, 1990).

Moraux, Paul, *Aristote: Du ciel: texte établi et traduit* (Paris, 1965).

Moravcsik, Julius M., 'Aristotle on Adequate Explanations', *Synthese* 28 (1974), 3–17.

— '*Aitia* as Generative Factor in Aristotle's Philosophy', *Dialogue* 14 (1975), 622–38.

Moreau, Joseph, *L'Espace et le temps selon Aristote* (Padua, 1965).

— 'Le Temps et l'instant selon Aristote' in Düring (ed.), *Naturphilosophie bei Aristoteles und Theophrast*.

Mourelatos, A. P. D., 'Aristotle's "Powers" and Modern Empiricism', *Ratio* 9 (1967), 97–104.

MUELLER, I., 'Aristotle on Geometrical Objects', *Archiv für Geschichte der Philosophie* 52 (1970), 156–71; reprinted in Barnes, Schofield, and Sorabji (eds.), *Articles on Aristotle*, iii. *Metaphysics*.

NEUGEBAUER, O., *The Exact Sciences in Antiquity*, 2nd edn. (Providence, Rhode Island, 1957; reprinted New York, 1962).

NUSSBAUM, MARTHA, *Aristotle's De Motu Animalium, Text with Translation, Commentary and Interpretive Essays* (Princeton, 1978).

— 'Saving Aristotle's Appearances', in Schofield and Nussbaum (eds.), *Language and Logos*.

OWEN, G. E. L., 'Zeno and the Mathematicians', *Proceedings of the Aristotelian Society* 58 (1957–8), 199–222; reprinted in Owen, *Logic, Science and Dialectic*.

— 'Τιθέναι τὰ φαινόμενα', in Suzanne Mansion (ed.), *Aristote et les problèmes de la méthode*; reprinted in Owen, *Logic, Science and Dialectic*.

— 'Aristotle: Method, Physics, and Cosmology', in Charles Coulston Gillispie (ed.), *Dictionary of Scientific Biography*, i (New York, 1970); reprinted in Owen, *Logic, Science and Dialectic*.

— 'Aristotle on Time', in Peter K. Machamer and Robert G. Turnbull (eds.), *Motion and Time, Space and Matter: Interrelations in the History of Philosophy and Science* (Columbus, Ohio, 1976); reprinted in Owen, *Logic, Science and Dialectic*.

— 'Aristotelian Mechanics', in Gotthelf (ed.), *Aristotle on Nature and Living Things*; reprinted in Owen, *Logic, Science and Dialectic*.

— *Logic, Science and Dialectic: Collected Papers in Greek Philosophy*, ed. Martha Nussbaum (London, 1986).

PELLETIER, FRANCIS JEFFRY, 'Sameness and Referential Opacity in Aristotle', *Noûs* 13 (1979), 283–311.

PICKERING, F. K., 'Aristotle on Zeno and the Now', *Phronesis* 23 (1978), 253–7.

PREUS, ANTHONY, 'Aristotle's "Nature uses . . ."', *Apeiron* 3 (1969), 20–33.

REID, THOMAS, *Essays on the Intellectual Powers of Man* (Glasgow, 1785).

ROBIN, LÉON, 'Sur la conception aristotélicienne de la causalité', *Archiv für Geschichte der Philosophie* 16 (1909–10), 1–28 and 184–210.

ROSS, W. D., *Aristotle's Physics: A Revised Text with Introduction and Commentary* (Oxford, 1936).

— *Aristotle's Prior and Posterior Analytics: A Revised Text with Introduction and Commentary* (Oxford, 1949).

RUSSELL, PAUL, 'Sorabji and the Dilemma of Determinism', *Analysis* 44 (1984), 166–72.

SALMON, W. C., *Scientific Explanation and the Causal Structure of the World* (Princeton, 1984).

SAMBURSKY, S., *The Physical World of the Greeks* (London, 1956).

— *The Physical World of Late Antiquity* (London, 1962).

— and PINES, S., *The Concept of Time in Neoplatonism: Texts with Translation, Introduction and Notes* (Jerusalem, 1971).

SAUVÉ, SUSAN, 'Unmoved Movers, Form, and Matter', *Philosophical Topics* 15 (1987), 171–92.

— Review of David Charles, *Aristotle's Philosophy of Action*, *Philosophical Review* 97 (1988), 411–16.

SCHOFIELD, MALCOLM and NUSSBAUM, MARTHA (eds.), *Language and Logos: Studies in Ancient Philosophy Presented to G. E. L. Owen* (Cambridge, 1982).

SCRIVEN, MICHAEL, 'Causation as Explanation', *Noûs* 9 (1975), 3–16.

SEECK, GUSTAV ADOLF (ed.), *Die Naturphilosophie des Aristoteles* (Darmstadt, 1975).

SOLMSEN, FRIEDRICH, *Aristotle's System of the Physical World: A Comparison with his Predecessors* (Ithaca, NY, 1960).

— 'Nature as Craftsman in Greek Thought', *Journal of the History of Ideas* 24 (1963), 473–96.

SORABJI, RICHARD, 'Aristotle, Mathematics and Colour', *Classical Quarterly* 22 (1972), 293–308.

— 'Aristotle on the Instant of Change', *Proceedings of the Aristotelian Society* Supplementary Vol. 50 (1976), 69–89; reprinted in Barnes, Schofield, and Sorabji (eds.), *Articles on Aristotle*, iii. *Metaphysics*.

— *Necessity, Cause, and Blame: Perspectives on Aristotle's Theory* (London and Ithaca, NY, 1980).

— *Time, Creation, and the Continuum: Theories in Antiquity and the Early Middle Ages* (London, 1983).

— *Matter, Space, and Motion: Theories in Antiquity and their Sequel* (London, 1988).

STRAWSON, P. F., *Individuals: An Essay in Descriptive Metaphysics* (London, 1959).

STRIKER, GISELA, 'Notwendigkeit mit Lücken: Aristoteles über die Kontingenz der Naturvorgänge', *Neue Hefte für Philosophie* 24–5 (1985), 146–64.

TORSTRIK, A., '*Περὶ τύχης καὶ τοῦ αὐτομάτου*: Aristot. *Phys.* B 4–6', *Hermes* 9 (1875), 425–70.

VAN FRAASSEN, BAS C., *The Scientific Image* (Oxford, 1980).

— 'A Re-Examination of Aristotle's Philosophy of Science', *Dialogue* 19 (1980), 20–45.

WAGNER, HANS, *Aristoteles: Physikvorlesung* (Berlin, 1972).

WARDY, R., *The Chain of Change: A Study of Aristotle's Physics* VII (Cambridge, 1990).

WATERLOW, SARAH, *Nature, Change, and Agency in Aristotle's Physics: A Philosophical Study* (Oxford, 1982).

— *Passage and Possibility: A Study of Aristotle's Modal Concepts* (Oxford, 1982).

— 'Instants of Motion in Aristotle's *Physics* VI', *Archiv für Geschichte der Philosophie* 65 (1983), 128–46.

— 'Aristotle's Now', *Philosophical Quarterly* 34 (1984), 104–28.

WEISS, HELENE, *Kausalität und Zufall in der Philosophie des Aristoteles* (Basel, 1942).

WHITE, MICHAEL J., 'Aristotle on "Time" and "A Time"', *Apeiron* 22 (1989), 207–24.

WHITE, NICHOLAS, P., 'Aristotle on Sameness and Oneness', *Philosophical Review* 80 (1971), 177–97.

WIELAND, WOLFGANG, *Die Aristotelische Physik: Untersuchungen über die Grundlegung der Naturwissenschaft und die sprachlichen Bedingungen der Prinzipienforschung bei Aristoteles*, 2nd edn. (Göttingen, 1970).

— 'Aristotle's *Physics* and the Problem of Inquiry into Principles', in Barnes, Schofield, and Sorabji (eds.), *Articles on Aristotle*, i. *Science*.

— 'The Problem of Teleology', in Barnes, Schofield, and Sorabji (eds.), *Articles on Aristotle*, i. *Science*.

WILLIAMS, C. J. F., *Aristotle's De Generatione et Corruptione: Translated with Notes*, Clarendon Aristotle Series (Oxford, 1982).

WITT, CHARLOTTE, *Substance and Essence in Aristotle: An Interpretation of Metaphysics VII–IX* (Ithaca, NY and London, 1989).

WOLFF, MICHAEL, *Fallgesetz und Massebegriff* (Berlin, 1971).

— 'Philoponus and the Rise of Preclassical Dynamics', in Richard Sorabji (ed.), *Philoponus and the Rejection of Aristotelian Science* (London, 1987).

WOODFIELD, ANDREW, *Teleology* (Cambridge, 1976).

WOODS, MICHAEL, *Aristotle's Eudemian Ethics, Books I, II, and VIII: Translated with a Commentary*, Clarendon Aristotle Series (Oxford, 1982).

WRIGHT, LARRY, 'Functions', *Philosophical Review* 82 (1973), 139–68.

INDEX LOCORUM

INDEX OF NAMES

9 780198 236023

Aristotle's *Physics*
A Collection of Essays

Edited by
LINDSAY JUDSON

CLARENDON PRESS · OXFORD

ARISTOTLE'S *PHYSICS*

This book has been printed digitally and produced in a standard specification in order to ensure its continuing availability

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

Reprinted 2003

ISBN 0-19-823602-6

ACKNOWLEDGEMENTS

I should like to thank my wife, Jenny Tyler, for all her help in the preparation of this volume, and my daughter Anna Penelope for sometimes being quiet enough for me to hear my wife's comments.

LJ

CONTENTS

NOTES ON CONTRIBUTORS

ROBERT BOLTON is Professor of Philosophy at Rutgers University. He is the author of a number of articles on Aristotle.

DAVID BOSTOCK is a Fellow of Merton College, Oxford. His publications include *Logic and Arithmetic*, *Plato's Phaedo*, and *Plato's Theaetetus*.

DAVID CHARLES is a Fellow of Oriel College, Oxford. He is the author of *Aristotle's Philosophy of Action*.

WILLIAM CHARLTON is Senior Lecturer in Philosophy at the University of Edinburgh. His publications include *Aristotle's Physics, Books I and II* in the Clarendon Aristotle Series and a translation of Philoponus' commentary on *De Anima* III. 4–8, as well as articles on ancient philosophy.

CYNTHIA A. FREELAND is Associate Professor of Philosophy at the University of Houston. She has published articles on Aristotle's science, metaphysics, and ethics, and has also written on topics in aesthetics and feminist theory.

MARY LOUISE GILL is Associate Professor of Classics and Philosophy at the University of Pittsburgh. She is co-editor of *Ancient Philosophy*, and the author of *Aristotle on Substance: The Paradox of Unity* as well as of a number of articles on Plato and Aristotle.

EDWARD HUSSEY is University Lecturer in Ancient Philosophy at Oxford and a Fellow of All Souls College. He has published *The Presocratics* and *Aristotle's Physics, Books III and IV* in the Clarendon Aristotle Series.

MICHAEL INWOOD is Fellow and Tutor in Philosophy at Trinity College, Oxford. He is the author of *Hegel*, and of a number of articles on Hegel.

LINDSAY JUDSON is Official Student and Tutor in Philosophy at Christ Church, Oxford. He is Associate General Editor of the Clarendon Aristotle Series, and has published a number of articles on the history of philosophy.

JULIUS M. MORAVCSIK is Professor of Philosophy at Stanford University. He has written numerous articles on ancient philosophy, has edited several anthologies of essays, and is the author of *Thought and Language*.

ABBREVIATIONS

Cat.	*Categories*
Int.	*De Interpretatione*
An. Pr.	*Prior Analytics*
An. Post.	*Posterior Analytics*
Top.	*Topics*
Soph. El.	*Sophistici Elenchi*
Phys.	*Physics*
Cael.	*De Caelo*
GC	*De Generatione et Corruptione*
Meteor.	*Meterologica*
An.	*De Anima*
Sens.	*De Sensu*
Mem.	*De Memoria et Reminiscentia*
Som.	*De Somno*
Insom.	*De Insomniis*
Div.	*De Divinatione Per Somnum*
Long.	*De Longitudine et Brevitate Vitae*
Resp.	*De Respiratione*
HA	*Historia Animalium*
PA	*Parts of Animals*
MA	*De Motu Animalium*
IA	*De Incessu Animalium*
GA	*Generation of Animals*
Met.	*Metaphysics*
EN	*Nicomachean Ethics*
EE	*Eudemian Ethics*
Rhet.	*Rhetoric*

I

Aristotle's Method in Natural Science: *Physics* I

ROBERT BOLTON

I Aristotle's Concern with Method

Aristotle begins his *Physics*, and thus his studies in natural science generally, with a chapter on method. Further remarks on this subject are added at crucial junctures later in the work, particularly in the subsequent chapters of Book I and in Book IV. This concern with method in the *Physics* is just one prominent example of the explicit and self-conscious attention to method which is displayed in nearly all of Aristotle's major works. Among other things, this interest reflects the lively concern, and the lively disagreement, among philosophers and scientists in Aristotle's own day not only about positive doctrine in various areas but also about the proper method or methods to use to reach such doctrine.[1]

An equally lively concern, and equally lively disagreement, about how to *understand* Aristotle's views on method has figured prominently in recent Aristotelian scholarship. The *Physics* itself has served as a *locus classicus* for this disagreement. Traditionally scholars have found the notion congenial that Aristotle's intended method in his works on natural science is empirical, even as they have criticized him for failures on this count. The current generation has reversed this verdict entirely. The *Physics* in particular is now standardly taken as a paradigm of Aristotle's use of dialectical method, understood as a largely conceptual or a priori technique of inquiry appropriate for philosophy, as opposed to the more empirical inquiries which we,

[1] Of many relevant passages, one not usually noticed, *GC* I. 2, 316^a5–14, nicely illustrates Aristotle's stand on one of these important disagreements about method which will be of particular interest here.

these days, now typically regard as scientific.[2] This sharp difference of interpretation raises a question which takes us, clearly enough, to the heart of Aristotle's conception of the nature of philosophy and of science. Does Aristotle follow Plato in thinking of philosophy and science as, in important respects, non-empirical in its method, or does he oppose Plato on this as on so many other points? This is, in effect, the main question with which we will be concerned here.

II Scientific Method in *Physics* I. 1: Some Puzzles

Though Aristotle begins his *Physics* with a chapter on method, recent studies have largely ignored this chapter and have concentrated mainly on Book IV. This may be due, in part, to the comparative obscurity of the initial discussion but, clearly enough, further study should take full account of Aristotle's own point of departure and its application in Book I.[3] The initial chapter is compact enough to include here in full.

> Knowledge and scientific knowledge, in every inquiry where there are principles or causes or elements, comes from the grasp of these things. For we think we know a thing when we grasp its primary causes and its first principles and we are down to its elements. It is evident, then, that in the science of nature too the primary task must be to settle the things which concern the principles.
>
> The natural procedure is to go *from* things which are more intelligible and clearer to us *to* things which are clearer and more intelligible by nature. For what is intelligible to us is not the same as what is intelligible simply. Therefore, it is necessary to proceed in this manner: from what is less clear by nature, though more clear to us, to what is clearer and more intelligible by nature. The things which are first evident and clear to us are things which are rather jumbled up. Later from them their elements become intelligible and the

[2] A main stimulus for this approach is G. E. L. Owen's influential article '*Tithenai ta phainomena*' (1961, final version in his *Logic, Science and Dialectic*, London and Ithaca, 1986). A variant of the approach is developed in W. Wieland, *Die Aristotelische Physik* (Göttingen, 1962; 2nd edn., 1970); see e.g. pp. 216 ff. Owen's reading of Aristotle's method in the *Physics* is now standard: see, e.g., W. Charlton, *Aristotle's Physics I–II* (Oxford, 1970), pp. x–xii, 66, *et passim*; E. Hussey, *Aristotle's Physics III–IV* (Oxford, 1983), p. ix *et passim*; and T. H. Irwin, *Aristotle's First Principles* (Oxford, 1988), sections 34–7. For other references and general discussion of the issues see R. Bolton, 'Definition and Scientific Method in Aristotle's *Posterior Analytics* and *Generation of Animals*', in A. Gotthelf and J. Lennox (eds.), *Philosophical Issues in Aristotle's Biology* (Cambridge, 1987).

[3] For discussion of the issues raised by Book IV see Bolton, 'Definition and Scientific Method', 120–30, 164–6.

principles serve to separate these [jumbled up] things. That is why it is necessary to proceed from the comprehensive to specifics; for the whole is more intelligible in relation to perception and the comprehensive is a sort of whole since it embraces many things as parts.

This same thing happens also, in a certain way, to names in relation to their account. For a name also signifies a certain whole in an indefinite way; for instance the circle [is so signified], while the definition of it distinguishes [the whole] into its specifics. Also, children at first call all men father and all women mother; later they distinguish each of them. (184^a10–$^b14$)

This chapter begins with material which is quite familiar and intelligible to us from other passages in Aristotle, particularly from the early chapters of the *Posterior Analytics*. The first paragraph introduces the doctrine of the *Analytics* that unqualified scientific knowledge in any area requires an understanding of the most fundamental principles, principles which explain why the other facts hold (*An. Post.* I. 2; cf. II. 1–2). The second introduces the distinction made in the *Analytics* between the things which are more intelligible to and first known by us on some subject and the things which are first known and more intelligible by nature, where the latter things are the fundamental principles which are explanatory of the other things (I. 2, 71^b29–72^a5). However, Aristotle's further description of the process of scientific inquiry in this passage is not so clear in the light of other familiar passages. In particular, two features of his description of the starting-points for scientific inquiry are, notoriously, unusual and not easy to understand.

First, these things—the things which are first evident and clear to us—are described as things which are jumbled up or compounded (συγκεχυμένα 184^a22). Second, they are described as comprehensive (καθόλου $^a23$–4) and whole (ὅλον $^a24$) because at this level our understanding embraces many as yet undistinguished things as parts (μέρη $^a26$, with $^b2$ καὶ ἀδιορίστως and $^b5$). Aristotle offers two illustrations to make clearer what he means: (1) a name, he says, also signifies a certain whole or complex in an indeterminate way, that is, without clearly distinguishing what a proper definition would distinguish; and (2) little children also cannot distinguish things, e.g. their mother from other women, which they later can distinguish. However, it is not easy to understand just what the precise content and point of these examples is. They do indicate, clearly enough, that the initial facts which we grasp about some object of scientific inquiry do not enable us adequately to distinguish that object from some others which we may also be aware of. When we find the principles which explain these facts we can make the distinctions. This does not take us very far, however, in our attempt to understand in detail the *process* of inquiry itself or Aristotle's two unusual descriptions of its starting-points.

In addition to these two unusual descriptions of the starting-points, however, there is a final one which is again familiar. The starting-points are further described as what is more intelligible (or more knowable) in relation to perception. This too fits with what we find at the beginning of the *Posterior Analytics*, where Aristotle identifies what is first known by and more intelligible to us with what is 'closer to perception' (I. 2, $72^{a}1$–5). In *Top.* VI. 4 what is more intelligible to us is similarly described as what is most accessible to perception ($141^{b}5$–14). However, in the passage in the *Analytics* such items are said to be most opposed to and furthest from what is καθόλου (i.e. universal). This flatly contradicts what is said in *Phys.* I. 1 unless, as most commentators reasonably conclude, the use of the term καθόλου in *Phys.* I. 1 is peculiar and not the typical use displayed, for instance, in the passage at the beginning of the *Analytics*. (I have followed the commentators here and used 'comprehensive' not 'universal' to translate καθόλου in *Phys.* I. 1.)

This standard move serves, however, only to deepen the problem of how to understand the peculiar descriptions in the *Physics*. But one close and clear point of correspondence between the *Physics* and the *Analytics* suggests a route to try in order to resolve this problem. Both works, as already noted, identify what is most knowable to us, i.e. the starting-points for scientific inquiry, with what is most knowable, or is knowable most directly, from perception. The famous final chapter of the *Posterior Analytics* (II. 19) attempts, in more detail than does the opening chapter of the *Physics*, to describe just what the process is by which we move from what we are most directly provided with by perception to a grasp of fundamental scientific principles. This connection suggests that attention to that chapter may be useful for understanding the opening chapter of the *Physics*.

III *Physics* I. 1 and *Posterior Analytics* II. 19

The aim of *An. Post.* II. 19 is the same as that of *Phys.* I. 1, to explain how we come to know first principles in science ($99^{b}17$–18). These are the first principles of demonstration ($100^{b}8$–14; cf. $99^{b}26$–9). Two constraints are imposed by Aristotle at the beginning of his discussion in the chapter on this explanation. First, since knowledge of the principles is not innate the explanation must show how the principles are *learned* (99^b26–9). Second, since we learn them, the explanation of how they are learned must show how we infer the principles from what is previously known. This is what learning of the sort in question requires, as Aristotle has already said at the beginning of the *Posterior Analytics* ($^b28$–30; cf. I. 1, 71^a1 ff., $^b29$–33). The requirement of *previous* knowledge for learning, as the earlier passages make clear, is the requirement of this knowledge as a basis for inference to what is learned.[4] This shows that the account in II. 19 is not, as some have supposed, merely a genetic account of the psychological preconditions for the generation of knowledge of first principles. It is an account of the inferential process by which such knowledge is reached from previous knowledge.[5] This previous knowledge cannot, however, be more exact than the principles, that is, it cannot serve as a basis for *demonstrating* the principles, since they are immediate and indemonstrable. So it must be less exact (99^b30–4). Thus, we must have some power which enables us to acquire initial knowledge of what is less exact than the principles from which the principles can be inferred and thus learned ($^b30$–4). This power, Aristotle says, is perception ($^b34$–5). Perception, via memory, provides us with items of experience (ἐμπειρία) which serve as the starting-point for learning the principles.[6] Aristotle describes these items of experience, in the *Analytics*, in just the way in which he describes the starting-points for scientific inquiry in *Phys.* I. 1. They are the things 'from which' the fundamental principles are learned (100^a6–10, cf. 184^a22–3), without being more knowable or intelligible (by nature) than the principles (100^a10–11, cf. 184^a18–21). Thus they are acquired by perception, which is a procedure for acquiring information that is more knowable to us (100^a11; cf. 184^a24–5).

[4] J. Barnes (*Aristotle's Posterior Analytics* (Oxford 1975), 251) criticizes Aristotle for invoking his earlier doctrine from I. 1 about 'intellectual' or inferential learning. His reason is that Aristotle does not believe that the learning of the principles is intellectual, that is, 'knowledge of principles is not deduced knowledge'. But Aristotle's invocation of the earlier doctrine shows that he does believe that the learning of principles is inferential. The inferential process may not be deductive but, according to I. 1, not all intellectual learning is deductive. It may also be inductive (71^a5–9), which is what Aristotle has in mind here (100^b2–4).

[5] A main source for the alternate approach is W. D. Ross, *Aristotle's Prior and Posterior Analytics* (Oxford, 1949) 49. See, further, section VIII.

[6] 'From perception memory comes, from memory . . . experience [comes], . . . from experience . . . a first principle of art or science [comes]' (100^a3–8).

These connections establish a very close correspondence between the two passages and lead directly to the question whether there are further connections which illuminate Aristotle's thought in the *Physics*. In particular, does anything in the *Analytics* correspond to Aristotle's descriptions of the starting-points for inquiry in the *Physics* as ὅλον καθόλου and as συγκεχυμένα which contain μέρη that are ἀδιόριστα? Interestingly enough, the first basic account that Aristotle offers of the starting-points, i.e. of 'experience', in *An. Post.* II. 19 is that it is 'the πᾶν καθόλου which is stabilized in the soul' (100ᵃ6–7). This is later described as a 'first καθόλου', which is a ἓν τῶν ἀδιαφόρων (100ᵃ15–16). This could easily mean that it is 'a unity composed of undifferentiated things', i.e. a συγκεχυμένον. This first καθόλου is then said to be succeeded as learning advances by things which are ἀμερῆ, 'without parts' (as well as being καθόλου, 100ᵇ1–2). This language, clearly, is extremely close to the language which Aristotle uses in *Phys.* I. 1 to describe the starting-points, and the further process of scientific inquiry leading to knowledge of principles. This establishes, I think, at least a prima-facie presumption that Aristotle is using the same language to say the same thing in the two passages where, after all, he is discussing the same topic.

This, however, runs counter to the mainstream of interpretation of the *Analytics* passage. The πᾶν καθόλου in the soul, which Aristotle identifies with 'experience' there, is standardly taken to be a 'universal' in the more familiar sense, that is, a universal proposition (or concept), since experience is described there as a grasp of 'the one apart from the many, what is one and the same in all those things' (100ᵃ7–8). But there are very serious difficulties with this interpretation. First of all, Aristotle says clearly that experience is constituted *simply* by 'many memories' of particular cases of the same thing grouped together (100ᵃ5–6), not by something which covers *all* cases of the same kind of thing. In the parallel passage in *Met. A.* 1, in fact, Aristotle denies that experience involves a grasp of a 'universal' of this latter sort. His example there of something one knows from experience is that: 'In many individual cases, namely the cases of Socrates, Callias, etc., this treatment was beneficial for this sickness.' But, Aristotle expressly says, this does *not* involve a 'universal judgement' that in 'all cases distinguished *as one in kind*' (κατ' εἶδος ἓν ἀφορισθεῖσι) this treatment was beneficial for this sickness (981ᵃ5–12). This fits perfectly with what Aristotle says in *Physics* I. 1 where also, at the initial stage of our understanding of certain things, those things are undifferentiated (ἀδιόριστα, 184^b2) in kind from other related ones.

This does not conflict, moreover, with the description of experience in the *Analytics* as a grasp of what is 'the same in all *those things*', that is, all of the 'many things' remembered (100^a7–8). To take the example from the *Metaphysics*, one is aware there, just as the *Analytics* says, of one thing which is 'the same in all' the particular cases, i.e. the cases of a certain illness which we recall, namely that this treatment works as a cure. We just don't know how to mark off definitively these cases as a *kind* (εἶδος) distinct from other similar ones. So it is quite easy to make the initial treatment of experience in the *Analytics* internally consistent, and consistent with the treatment in the *Metaphysics*, if the πᾶν καθόλου which we grasp in experience is the ὅλον καθόλου of *Phys.* I. 1, rather than a strict universal.

IV Further Support for a Unified Account

There are equally good reasons for treating the second discussion of 'experience' in the *Analytics* passage (100^a14 ff.) along the same lines. Aristotle goes on to redescribe the 'first καθόλου' which is present in the soul as a ἓν τῶν ἀδιαφόρων (100^a15–16). Here again the καθόλου is normally understood as a standard universal, since Aristotle says of the perceptual state in which it is grasped that it is, for example, 'of man, not of the man Callias' (100^b1). But this, again, faces serious difficulties. Earlier in *An. Post.* I. 31, Aristotle makes the same point that 'perception is of a certain sort of thing not of the individual'. But he goes on immediately to add that, nevertheless, one cannot perceive 'what is universal and holds in every case' of a given kind. This sort of universal is grasped '*from* our seeing', that is, by reasoning from what we 'often observe' or what we 'see in each individual case'—in other words, by reasoning from experience. It is the object of scientific knowledge (ἐπιστήμη) not perception or experience (87^b28–31, 88^a2–17). So the καθόλου of II. 19 (100^a17) which *is* the object of experience, or of what we often perceive in many individual cases without further reasoning, cannot be a universal of the sort which 'holds in every case' of a given kind.

Furthermore, if we understand the object of experience as a universal of that sort it is hard to make sense of Aristotle's description of it as a ἓν τῶν ἀδιαφόρων (100^a15–16). This cannot mean that grasp of it involves an awareness of just *one* of the many particulars referred to earlier, on which experience is based (100^a3–6), since Aristotle explicitly says that experience has to be based on many of these. So commentators typically interpret it to mean one of the *infimae species*, which are ἀδιάφορα in the sense of not being further sub-dividable. This, then, is understood to commit Aristotle to a certain view of concept acquisition, according to which we start from particulars then through abstraction we grasp, first, *infimae species*, then progressively higher genera. But in this case the chapter turns out to be about how we acquire general or abstract concepts, not about what it is supposed to be about, namely how we acquire knowledge of *all* first principles (99^b17–18). Acquiring general concepts may have something to do with coming to know adequate scientific definitions, though they are hardly the same thing.[7] But the acquiring of general concepts is not at all the same thing as coming to know the first principles of the types other than definitions, namely the axioms and hypotheses (see *An. Post.* I. 10).

In addition, even if we can somehow ignore the problem about the other types of principles (as we will see below we cannot), and even if we can somehow treat the acquiring of general concepts and the acquiring of definitions as much the same thing, this interpretation still commits Aristotle to the curious idea that we acquire genuine first principles, that is, scientific knowledge, of species before we acquire scientific knowledge of the genera under which they fall. This conflicts with his standard doctrine that a species is defined, and known *as* a complex of genus and differentia.[8] One might respond to this (though the main tradition would not) by saying that Aristotle is talking here simply about rough initial definitions or concepts of species and genera, not the adequate definitions or concepts where knowledge of a species presupposes knowledge of its genus. But there is no good reason to believe he thought that we always acquire rough definitions or concepts of more specific things before we acquire rough definitions or concepts of more generic things. His examples earlier in the *Analytics* of our initial or rough definitions and concepts often include, in the *definiens*, the genera of the things defined, for instance the initial

[7] It is clear from Aristotle's earlier discussion in *An. Post.* II. 8–10 that in his view one can have a general concept adequately formulated in a *nominal* definition, e.g. of man, without having an adequate scientific definition.

[8] See *Top.* VI. 4, 141^b29–34. This picture of definition by division is present also in the *Posterior Analytics* (II. 13, 97^a23–8).

rough definition of thunder as a certain noise in the clouds and of man as a certain kind of animal (see II. 8).

These difficulties are avoided if the ἓν τῶν ἀδιαφόρων is understood, in the light of *Phys.* I. 1, as a unity composed of (as yet) undifferentiated things. That is, the content of perception, at the level where we have only experience, is a καθόλου in the sense of *Phys.* I. 1. This content is still, 'for example, man, not the man Callias' (100^{a}17–18) because, as Aristotle has already said, perception has for its content 'a certain sort of thing not the individual' (87^{b}28–31). To take the case from *Met. A*. 1, when we are aware through perception of many cases where a certain treatment cured fever, we are aware of a certain sort of thing; that is, we know of a certain treatment (in many cases) curing fever, and we are not simply aware of the individual cases. However, we cannot distinguish the cases where this works 'as a kind'. So too in the example from the *Analytics* we are aware of (many cases of) a certain sort of animal, Aristotle says, and thus of man. We just cannot, at this stage, precisely delimit the kind. Further attention to what we experience enables us to move from this grasp of man, i.e. of a certain sort of animal, to the point where we can clearly distinguish the kind, animal, and within it man and other distinct species. At this point, Aristotle says, we have καθόλου which are ἀμερῆ, that is, not composed of as yet undifferentiated parts and, thus, we have καθόλου which are now universals in the familiar sense. They cover all cases of a genuine distinct kind. (100^{b}1–3: 'In turn, among these [undifferentiated things] a stand is made to the point where καθόλου without [further undifferentiated] parts stand, e.g. [we go from] this sort of animal to the point where [the kind] animal stands [distinct] and within it [the sub-kind man] equally.')

The comparison, then, between the parallel passages in the *Physics* and the *Analytics*, particularly in light of *Met. A*. 1, supports the suggestion that Aristotle uses the same unusual language in the two passages to say the same thing. This has its pay-off, as we have just seen, in enabling us to see how to avoid certain inconsistencies and solve certain puzzles in the interpretation of the *Analytics*. Of more interest for present purposes is the pay-off for the interpretation of the *Physics*.

V *Posterior Analytics* II. 8 and the Puzzles in *Physics* I. 1

The use of the *Analytics* to understand the puzzling details of *Phys.* I. 1 is facilitated if we remember an important fact about the *Analytics* which is often overlooked. The final chapter of the *Posterior Analytics* is not the first or only place in that work where Aristotle describes in some detail how we come to know first principles. He concludes *An. Post.* II. 8 with the remark, 'We have now said how what a thing is [i.e. how a basic definition] is grasped and comes to be known' (93^b15–16). So Aristotle is talking about the same subject as in II. 19, namely, 'how the first principles come to be known' (99^b17–18), with special attention to those first principles which are definitions. This indicates that we should use II. 8 together with II. 19 to understand *Phys.* I. 1.

As we might expect, the discussion in II. 8 follows the same pattern as in the later chapter. Aristotle supposes that we start from some initial items of information about some subject and come to know the principles as the things which account for our initial information. The discussion, again, is not a piece of genetic psychology but a description of the scientist's actual method. In fact, one of the examples which Aristotle mentions in II. 8 of our initial knowledge on the basis of which a basic definition becomes known is just the one he uses prominently in II. 19, namely our initial grasp of man as 'a certain kind of animal' (93^a23–4; cf. 100^b2–3). Unfortunately Aristotle does not say more about that case in II. 8, but his other examples there are helpful. We are first familiar with a lunar eclipse, for instance, as 'a certain loss of light' by the moon (93^a23). At this stage, 'from which' we begin to learn the first principle which gives the fundamental definition of eclipse, what we are aware of, as II. 19 makes clear, is 'many cases' of loss of light by the moon (cf. I. 31, 88^a2–7). It is easy to see, moreover, how we might at this stage be only aware of a καθόλου or a συγκεχυμένον containing as yet undistinguished things, as *Phys.* I. 1 says we are. For there are other cases of loss of light by the moon which we may well not be able initially to distinguish in kind from the cases of eclipse that we have experienced. For instance, the moon loses its light when it wanes and we may not know initially whether the waning of the moon is a different kind of thing from an eclipse, or perhaps the same kind of thing. In this case the common thing which we grasp the genuine cases of eclipse *as*—namely a loss of light—contains many as yet undistinguished types of things. (In early Greek astronomy some did treat eclipses and the phases of the moon as the same kind of thing.[9]) The principles, however, as the *Physics* says, 'serve to distinguish these [previously undifferentiated] things' (184^{a}23). When we learn, for example, that the basic definition and cause of all the repeated familiar cases of eclipse which we grasp in experience is the blockage by the earth of the moon's light, then we can distinguish eclipses universally 'as a kind', since the causes of the other phenomena are not the same as this (See *An. Post.* I. 31, 87^{b}39–88^{a}6). This use of *An. Post.* II. 8 together with II. 19 enables us to see then, in further detail, what Aristotle has in mind in his peculiar descriptions of the starting-points, and of the causes and principles which serve to differentiate their 'jumbled up' elements, in *Phys.* I. 1.

VI Some Conclusions about Aristotle's Method in Natural Science

The correspondence between the *Physics* and the *Analytics* enables us now to draw a general conclusion about the method which Aristotle recommends in the *Physics* for natural science, namely that this method is exactly the one he describes in the *Posterior Analytics*. It is an immediate consequence of this that Aristotle's method in natural science is not dialectic; it is not the method which he describes in the *Topics*, as opposed to the one which he lays out, particularly in the second book, in the *Posterior Analytics*. The most obvious point of difference has to do with the difference in the starting-points. Dialectic starts and reasons exclusively from ἔνδοξα, that is, from the standing convictions either of everyone, or of most people, or of the wise—either all of the wise, or most of them, or the most famous (*Top.* I. 1, 100^{a}29 ff.). Scientific inquiry on the other hand starts from experience; and an item of experience as Aristotle describes it in *An. Post.* II. 19, and elsewhere, need not be an ἔνδοξον—a standing conviction of the many or the wise. Items of experience may be ἔνδοξα, but they need not be.

[9] See J. Burnet, *Early Greek Philosophy*, 4th edn. (London, 1930), 78, and references given there. There is another clear case of this sort, which Aristotle introduces himself, in the *Generation of Animals*. He points out there that from our initial understanding of *sperma*, as an originating source of animal generation, it is not clear to us whether the *menses* is a type of *sperma* or not (721^{a}30–^{b}6 with 724^{a}14–20 and 726^{a}28–^{b}1).

What is important for scientific starting-points is that they come from proper experience, not who accepts them.[10]

Since it seems quite clear in *Phys.* I. 1, even independently of the problems of interpretation which we have discussed, that the starting-point for scientific inquiry is, as Aristotle says, simply what is 'more intelligible to (or more knowable by) perception' (184^{a}24–5), one might well ask why interpreters have been so quick to suppose that Aristotle's recommended method in the *Physics* is dialectic. The evidence typically cited to support this claim from *Phys.* I. 1 is Aristotle's remark that scientific inquiry moves from what is more intelligible (or knowable) to us to what is more intelligible by nature.[11] It is widely supposed that this is a description of dialectic. But where does Aristotle ever describe dialectic in this way? The closest that he comes to this is in *Top.* VIII. 5 where he lays it down as a rule of dialectical reasoning that the premisses must be more intelligible or better known than the conclusion (159^{b}8–9). It is plausible enough that he means by this that the premisses must be more intelligible to us than the conclusion, which is then less intelligible to us than the premisses.[12]

But, clearly, just because an argument or an inquiry proceeds from what is more intelligible to us to what is less intelligible to us does not suffice to show that it proceeds from what is more intelligible to us to what is more intelligible by nature. So Aristotle's description of scientific inquiry in *Phys.* I. 1 is not as such a description of dialectic. Nor can we say that what he describes there must at least count as one type of dialectical reasoning, since it is at least one type of reasoning from what is more intelligible to us to what is less intelligible to us. Nothing in Aristotle's discussions of the method of dialectic marks out for us any *type* of dialectical reasoning whose rules guarantee or make it likely that the conclusion of some piece of reasoning is not only less intelligible to us than the premisses but also more intelligible by nature. What guarantees this result according to *Phys.* I. 1, and according to the corresponding material in the *Analytics*, is that the conclusion of our reasoning and our inquiry gives us a principle which *explains* (and gives us a firm delineation of) the perceptual phenomena

[10] For an elaboration of this point see Bolton, 'Definition and Scientific Method', 120–30.

[11] See Irwin, *First Principles*, 66.

[12] For discussion of this see R. Bolton, 'The Epistemological Basis of Aristotelian Dialectic', in D. Devereux and P. Pellegrin (eds.), *Biologie, logique et metaphysique chez Aristote* (Paris, 1990), section 7.

which we use to reach it. But no rule of general dialectic or of any type of dialectic which Aristotle discusses is designed to guarantee conclusions of this sort. So if dialectic does reach conclusions of this type it is accidental and not due to the method of dialectic itself.[13]

VII Further Discussion of Method in *Physics* I. 2

If the method recommended for inquiry in natural science in *Phys.* I. 1 is not dialectic but the method of the *Analytics*, we should expect this to be confirmed by Aristotle's later discussion, particularly in the later chapters of Book I where his preaching in chapter 1 ought most of all to be guiding his practice.[14] Since these chapters have seemed to many interpreters to present us with a paradigm of dialectical inquiry, they are all the more in need of attention here. In these chapters Aristotle is engaged in a search for at least some of the principles of nature, just as he indicates he will be in chapter 1. He begins, in chapter 2, by raising the question of how many principles there are and by pointing out that various positions on this topic have been taken by his predecessors. Before pursuing this question further, however, Aristotle stops to consider the position of the Eleatics, since according to their view there are *no* principles of nature because there is no natural world—no world of naturally changing things—at all. He prefaces his discussion of their view with some new methodological remarks:

To inquire whether what exists is one and unchanging is not an inquiry in the science of nature. For just as there is no further proof for the geometer to give to someone who has rejected his principles, but this proof will proceed from some different type of knowledge or from knowledge which is common to everyone, so also neither is there [a proof to give to one who denies principles] for anyone [whose inquiry is] governed by principles . . .

To inquire whether what exists is one in this [alternate, non-scientific] way is like discussing [dialectically] any other position put forward for the sake of discussion, such as Heraclitus' thesis . . . or it is like exposing as fallacious an eristic argument—which in fact the arguments of both Parmenides and

[13] It will become clear below that the type of reasoning which Aristotle has in mind in *Phys.* I. 1 and *An. Post.* II. 19 cannot be dialectical, even accidentally.

[14] It ought to guide his practice throughout his studies in natural science (184^a14–16). Aristotle conceives of scientific inquiry quite generally as the search for principles. (See *An. Post.* II. 1–2, *Met. A.* 1–2) And this chapter prescribes the method for this search in natural science. But Aristotle is least likely to lapse from his intended method in the immediately succeeding discussion.

Melissus are . . . For our part [as scientists], however, we proceed through the hypothesis that all, or some, of the things which exist by nature change, and this [hypothesis] is made evident by induction.

Nor is it proper [for the scientist] to expose every error, but only as many as someone falsely demonstrates from his principles; other errors, not. For example, it is the geometer's task to expose as fallacious the procedure of squaring [the circle] by use of segments, but not the geometer's task to expose Antiphon's procedure. Nevertheless, though these people [the Eleatics] are not doing inquiry in the science of nature, they do incidentally discuss problems in natural science. So perhaps it is proper to engage in a bit of [dialectical] discussion with them. For this investigation has use for [natural] philosophy. (184^b25–185^a20)

There are a number of points which Aristotle makes in this important passage which are directly relevant to our topic. First, Aristotle explicitly says that the study and the refutation of the position of the Eleatics is not a job for someone engaged in scientific inquiry in a natural science. Rather, it is a job for someone who argues from 'knowledge which is common to everyone'.[15] What does Aristotle mean by this? His remarks on this point are strikingly similar in every detail to remarks which he makes in *Soph. El.* 9 and 11. There Aristotle also distinguishes argument, and in particular refutation (ἔλεγχος, λύσις), which is the proper job of the scientist, from argument and refutation based on 'the common things' (κοινά), which are matters of common knowledge and not known only by specialized scientists in a given area (170^a36–$^b11$; 171^b4–7; 172^a21 ff.). He even uses the same main example there as he uses here. He contrasts the refutation of the *scientific* attempt to square the circle through segments, which is proper to the geometer and thus employs *demonstration*, with the

[15] The alternative (185^a2), that the principles are demonstrated by some special science, super-ordinate to natural science, is not relevant here. So the refutation of the Eleatics must proceed from πασῶν κοινῆς (ἐπιστήμης), 185^a2–3. This phrase might be translated 'knowledge (or science) common to all the sciences'. (So Irwin, *First Principles*, 67 and n. 52, following most commentators.) Ross (*Aristotle's Physics* (Oxford, 1936), 337, 461) takes the phrase, so understood, to mean 'universal science', i.e. metaphysics. But metaphysics is universal because it studies what is primary, not because its contents (excepting the common axioms and the like) are in general common to all sciences, or to everything (*Met.* *E.* 1, 1026^a29–32). Some, including Irwin, take the phrase to refer to dialectic, since dialectic is an art useful for all sciences (*Top.* I. 2). But dialectic also does not typically prove things 'from *knowledge*' common to all sciences (or to everything). Aristotle does not describe metaphysical knowledge (or science) or dialectic as 'knowledge common to' all sciences or to everything. He does, however, describe dialectic, or a certain form of dialectic, as based on knowledge which is common *to all people* in contrast to knowledge which is based on mastery of a science, as we shall directly see.

refutation of Antiphon's procedure (171^b34–172^b4). The latter mode of argument is based on the common things in the relevant area which 'everyone, even the unlearned ... know themselves no less than the scientists' (172^a30–4), and is the province of dialectic, Aristotle says, and in particular of the special type of dialectic which he calls peirastic (172^a27–36).[16]

So the first point that Aristotle is making here in *Phys.* I. 2 is that the natural scientist cannot use a scientific, that is a demonstrative, argument to refute someone who denies that the natural world of changing things exists. In natural science it is an indemonstrable first principle that the natural world of changing things exists (cf. *Met. E.* 1, 1025^b7–28, quoted below). This principle is the basis for all demonstration in natural science, so no demonstration of it, in order to refute scientifically someone who denies it, is possible. One can only refute this denial dialectically, or peirastically.

VIII How does the Scientist know Principles?

What, however, is the epistemic status of the principle, the hypothesis that the natural world exists, for the scientist? If the dialectical procedures for dealing with the principle lie outside the bounds of strict scientific inquiry, how does the scientist as such come to know the principle? Aristotle does not claim or imply, as some have argued, that this principle is, for the scientist, self-evident or an object of epistemically 'immediate apprehension'.[17] Such a claim would in any case, of course, be quite absurd. How could it be self-evident that a natural world of changing things exists? Aristotle does not make this claim; rather he says quite specifically that 'for our part', i.e. for the natural scientist, this first principle becomes 'evident through induction' (185^a13–14). This is, of course, just what we would expect, given what Aristotle says about how we come to know first principles in *Phys.* I. 1, particularly in light of the parallel discussion in *An. Post.* II. 19. He specifically claims in the latter passage that we come to know *all* first

[16] In a parallel passage in *Rhet.* I. 1, Aristotle uses the same language and contrast which he employs in *Phys.* I. 2 (185^a2–3) and describes the basis for certain dialectical arguments as 'what it is common to everyone to know (κοινὰ ἁπάντων γνωρίζειν) and does not require any special scientific knowledge' (1354^a1–3). For further discussion of this matter see Bolton, 'Basis of Dialectic', section 10.

[17] See Irwin, *First Principles*, 68, 118 for a recent statement of this traditional view, taken also by Ross. See also n. 19.

principles 'by induction', that is, by inductive argument based on information provided by perception (100^b3–5). He repeats this claim in *Met. E.* 1 in a passage quite similar in its language to our passage in the *Physics*.

[The sciences] do not provide any proof (λόγος) of *what* [the object of their study] is. But starting from [an account of] what it is which some sciences make evident by perception and some set out as an hypothesis, they in this way demonstrate, more or less strictly, the things which belong in their own right to the kind with which they deal. Thus it is quite clear that there is no demonstration of the essence or of what something is, by such induction, but some other [non-demonstrative] way of making it evident. They similarly also say nothing [by way of demonstrative proof] as to whether the kind with which they deal exists or not, because it is the task of the same kind of intellectual reasoning [i.e. induction] to make it evident what something is and whether it exists. (1025^b10–18)

This passage says quite generally that the first principles of the special sciences, both definitions and existential principles, are typically made evident, to the scientist, by a procedure which treats them as hypotheses which are inductively confirmed (cf. *EE* 1218^b37–1219^a5). Aristotle mentions perception here as an alternate basis for coming to know principles, in addition to the hypothetical procedure which he mentions in *Phys.* I. 2 and elsewhere emphasizes (*An. Pr.* I. 30, 46^a17–27, *An.* I. 1, 402^b21–5). But in the *Posterior Analytics* he makes it clear that induction is the process at work in both kinds of cases (II. 19, 100^b3–5 with II. 2 90^a24–30 and I. 31, 88^a11–17). Aristotle's general name for the processes by which we learn in science other than by demonstration is 'induction' (I. 18).[18]

It is sometimes suggested that induction for Aristotle, as he understands it in the *Analytics* and elsewhere, is not a procedure for proof or justification at all but rather a procedure for pointing to what is otherwise evident, e.g. by being self-evident.[19] However, Aristotle standardly uses the term 'induction' for a type of argument (*Top.* I. 12, 105^a10–16), and he explicitly says in the *Analytics* that inductive arguments '*establish* (δεικνύναι) the universal through the fact that the particular is evident' (I. 1, 71^a8–9, cf. II. 7, 92^a34–8).[20] This shows clearly that

[18] The same point is made in *EN* VI. 3, 1139^b26–31.

[19] A main source for this approach is Ross, *Aristotle's Analytics*, 47–51. For a review of the more recent literature see R. McKirahan, 'Aristotelian Epagoge in *Prior Analytics* II. 21 and *Posterior Analytics* I. 1', *Journal of the History of Philosophy* 21 (1983).

[20] Cf. *EN* I. 7 (1098^b1–4) where Aristotle also mentions induction as a procedure by which principles are established. Whether, in contexts which have to do with things

Aristotle does not conceive of inductive reasoning as a procedure for discovery only and not justification. This is why inductive reasoning can serve as a procedure for learning, as Aristotle supposes it will in *An. Post.* I. 1 and II. 19. For all learning through reasoning involves coming to know something by inferring it from what is previously known which, thereby, provides an adequate justification for what is inferred (I. 1, 71^{a}1–17 with I. 2, 72^{a}25–32; cf. *EN.* VI. 3, 1139^{b}26–31).

IX Why the Scientist's Justification is not Dialectical

Aristotle's account of principles as reached by induction, in *Met. E.* 1, corresponds in very close detail to what he says in *Phys.* I. 2 about how natural science reaches its fundamental existential principle—that a world of naturally changing things exists (185^{a}12–14, cf. 1025^{b}18–21). However, in the *Physics*, Aristotle contrasts this procedure, used by the inquiring scientist, with the non-scientific, dialectical procedure to be used for refuting the Eleatics. This supports the result that the scientist's own procedure for confirming his principles is not dialectic. But it is worth investigating in further detail just how and why Aristotle draws the contrast he does, between scientific and dialectical procedures for dealing with principles. If there is for the scientist a sound inductive proof that a natural world of changing things exists, why does Aristotle not simply draw on that proof to refute the Eleatics rather than assigning that proof to the scientist and using other, dialectical arguments to refute the Eleatics?

As the beginning of an answer to this question we may note that an inductive proof which starts from premisses such as 'Birds fly' and 'Trees grow in the summer' and concludes that, therefore, 'A natural world of changing things exists' will not be convincing to any Eleatic, since he will be no more ready to accept the premisses of such a proof than he was to accept the conclusion. This cannot be the whole answer by itself, however, since there is also no reason to think that any Eleatic will accept the dialectical premisses which Aristotle in fact goes on to use. In general, we need to remember, it is not a requirement for a good dialectical or peirastic argument that an opponent be convinced by it or accept the premisses. What is required is that the conclusion follows from premisses of the proper sort. The premisses must

other than the learning of principles, Aristotle thinks of induction differently is beyond the scope of this paper.

be appropriate ἔνδοξα which, in the case of peirastic, are matters of common knowledge. An opponent may balk and not accept these premisses but if he does he is simply refusing to do proper dialectic. There is nothing dialectically wrong with the argument itself (*Top.* VIII. 11, 161^a16–24, $^b1$–6).

To see the full answer to our question it is necessary to remember certain crucial technical details in Aristotle's theory of refutation, and in particular of dialectical refutation, since Aristotle clearly intends his refutation here to be dialectical. To begin with, an inductive argument simply does not count as a refutation for Aristotle. A refutation is by definition a type of syllogistic, and thus deductive, argument (*Soph. El.* 165^a2–3, 167^a23–6, 171^a2–5). The premisses for a dialectical syllogistic refutation can be established inductively (*Top.* VIII. 1, 155^b17–25, 34–156^a7). But the premisses in a syllogistic argument must, of course, be different from the conclusion; and in a refutation the premisses cannot include or even immediately imply the conclusion since that would beg the question (167^a23–6). So the inductively established claim that a natural world of changing things exists cannot be used as a premiss for a syllogism that refutes the (direct or immediately implied) denial of this, because the conclusion of such a purported refutation would already be (or be immediately implied by) one of the premisses. So no inductive argument for the principle that this natural world of changing things exists can be used to refute someone who denies this.

This result can be stated more generally since every dialectical proof, like every refutation, is a syllogistic proof (*Top.* I. 1, 100^a18–23, cf. *Soph. El.* 34, 183^a38 ff.).[21] So, quite generally, the inductive procedures which the scientist uses to establish his own proper principles cannot be used to prove dialectically those principles, since what these inductive procedures establish—namely the principles themselves—cannot be used to prove syllogistically or prove without begging the question, those very principles. If there is to be any syllogistic proof of those principles, and thereby refutation of those who deny them, it must be from premisses established by other procedures than the very ones which the scientist uses to establish inductively those very principles. Dialectic, as itself a method for syllogistic proof, can sometimes provide us with such procedures, Aristotle implies (185^a2–3). But this shows again, conclusively I think, that whatever procedures dialectic

[21] Dialectic uses inductive arguments, as Aristotle says in *Top.* I. 12, but as he makes clear in the passages in VIII. 1 referred to above, this use is to obtain premisses for a syllogistic argument.

itself may draw on to prove scientific first principles they are not and do not include the very inductive procedures for establishing the principles which Aristotle ascribes to the scientist. It would be a mistake to think that this result, even though conclusive, depends merely on certain arcane technical requirements of dialectic which *we* might not think it proper to invoke in describing refutation in general. As noted above, peirastic dialectic, which is Aristotle's instrument of skilled refutation, draws for its premisses on common knowledge. Among other things this means that skilled refutation aims, quite reasonably, to draw as much as possible on common ground to the disputants on a given point. To take the case relevant here, it aims to draw on common ground both to those who are right about a given scientific principle and to those who are wrong (*Soph. El.* 11, 172^{a}21–^{b}1, cf. *Rhet.* I. 1, 1355^{a}24–9 with *Top.* I. 2, 101^{a}30–4). But, of course, in any given case one cannot beg the question against an opponent, so one cannot use any common knowledge which would involve this. One must, rather, restrict oneself to other common knowledge. For the reasons just indicated, reliance on the items of common knowledge used in an inductive proof of the principle that a natural world of changing things exists *would* beg the question against the general denial of this. So it is quite reasonable that the technical requirements of dialectic should be geared to rule this out and to require non-inductive proof, if any, of such a principle. An opponent may, of course, in a given case, refuse to accept even the common ground provided by the relevant non-question begging common knowledge. But in this event the opponent eliminates the possibility of even the most minimal form of non-specialized rational discussion of disagreements over principles and it is, once again, reasonable for the technical requirements of dialectic to be geared to reflect this.[22]

X How are Principles 'Known Through Themselves'?

This puts us now in a position to understand a further passage in *Phys.* II. 1, in which Aristotle returns to the question of how the basic principle, that the world of naturally changing things exists, comes to be known by the scientist. He says there:

That the natural world [of changing things] exists it would be absurd to try to establish (δεικνύναι). For it is obvious (φανερόν) that there are many such

[22] I discuss more generally the issues involved here in 'Aristotle's Conception of Metaphysics as a Science', (forthcoming).

things. To establish what is obvious through what is not is one expression of the inability to distinguish what is knowable through itself from what is not knowable through itself. That it is possible to be in this state is quite clear. For someone blind from birth might reason syllogistically to conclusions about colours. So necessarily, in such cases, the proof (λόγος) is about the [colour] names only, and nothing [about colours] is understood [through the proof]. (193^a1–9)

This is a passage to which some have appealed to show that scientific first principles, for Aristotle, are self-evident. But the passage does not warrant such a conclusion. Aristotle is speaking here about the requirements of proof that are incumbent on the scientist. As earlier in *Phys.* I. 2 where he says there is no *proof* (λόγος) for the scientist that there are naturally changing things, when he says here that this cannot be established, he means that there is no strict scientific proof, no demonstrative syllogistic proof through a middle term for this principle. His further remarks make this clear. The reason why scientific principles cannot be established is that they are knowable 'through themselves'. This is a point which Aristotle emphasizes elsewhere, notably in the first chapter of the *Topics* (100^a27 ff.). He there explains why the principles are knowable through themselves: 'For in the case of the first principles of the sciences it is not necessary *to inquire why* (διὰ τί) they are so' (100^b19–20). In *An. Post.* II. 2 Aristotle explains what he means by 'inquiring why' (διὰ τί). This is, he says, inquiring for a 'middle term' which makes the thing in question to be so, and through which it can be syllogistically demonstrated that it is so (90^a1). Thus, Aristotle says here in the *Physics*, one would only try to establish, i.e. to demonstrate, principles if one was ignorant of this point and could not distinguish what is knowable through itself from what is not. That Aristotle means to be ruling out here syllogistic proof, through a middle term, of the principle that nature exists, is made further apparent by the comparison with the congenitally blind person. What this person mistakenly tries to do, is to 'reason syllogistically' to conclusions about colours which are, for sighted people, known through themselves (193^a7).

This passage further suggests that Aristotle thinks not only that there is no demonstration that nature exists but that there is no no adequate syllogistic proof of this, not even a dialectical proof. He could argue this on the ground that a dialectical proof has to have premisses which are more of the nature of ἔνδοξα and more intelligible to us, i.e. more obvious to us, than the conclusion,[23] and nothing from which it could be syllogistically derived that the world of naturally changing things exists is more obvious to us than that fact. But none of this would show *how* this fact came to be so obvious to us, or to the scientist, and in particular it would be perfectly compatible with Aristotle's claim in *Phys.* I. 2 and elsewhere that for the scientist this becomes obvious through inductive proof. As we have just seen, to say that a principle is known '*through* itself' and not through something else is not, for Aristotle, to say that it is known *in* itself or by epistemically direct apprehension. It is to say simply that it is known without use of any mediation of a middle term such as would be used in a syllogistic proof, whether dialectical or demonstrative. This does not rule out inductive proof, since inductive proof does not rely on the mediation of any such middle term. So propositions known inductively *are* known '*through* themselves'; the truth of the connection between the terms of such propositions is not grasped through the mediation of any other term. The basis which Aristotle offers in *Phys.* II. 1 for the scientist's knowledge that nature—the world of naturally changing things—exists is that 'it is obvious (φανερόν) that there are many such things' (193ᵃ3–4). This is just the sort of base which Aristotle requires for an inductive argument, since such arguments establish their conclusions by virtue of the fact that, in each of the cases which make up the base, 'the particular is evident' (δῆλον, *An. Post.* I. 1, 71ᵃ8–9).

That the scientist's proof of his principles is inductive non-dialectical proof does not mean, of course, that dialectic is of no use in scientific inquiry. Aristotle himself says here that the dialectical investigation of the views of the Eleatics is of some use for natural philosophy. This fits with his famous remark in *Top.* I. 2 that it is necessary (he does not say it is sufficient) to use dialectic in reaching the first principles of the sciences (101ᵇ1–2). There he assigns this role to dialectic, owing to its power as a method of testing or examining (ἐξεταστική 101ᵇ3). This is what he goes on to employ dialectic to do in his review of the Eleatics. He first, in chapter 2, tests and refutes their claim that what exists is one and unchanging and then, in chapter 3, tests and shows their argument for the claim to be fallacious (see I. 3, 186ᵃ4–6). It would be of interest to study these arguments in detail to see just how Aristotle's discussion is of use for the natural philosopher in his search for principles. That would require a good deal more

[23] See *Top.* VIII. 5 and Bolton, 'Basis of Dialectic', section 7.

space, however, than is available here. What is clear enough is that those dialectical procedures must somehow be in aid of the general inductive procedures which constitute the proper scientific method for reaching the first principles, since that is the way the scientist reaches them.

XI Aristotle's Non-Dialectical Treatment of the Principles

Of more direct interest to us here, in any case, is the question whether, in addition to his dialectical discussion, Aristotle produces, or at least indicates how he would produce, an inductive non-dialectical justification of the particular principles which he uncovers by the end of *Phys.* I. Those principles, of course, are form, privation, and substratum. Before we consider how Aristotle justifies the positing of these things as principles, it will be useful to say something about the sense in which they *are* principles.

Unlike the examples of principles which Aristotle uses in *An. Post.* II these principles are not, and are not introduced in, definitions. But, as we have seen from his remarks there and in *Met. E.* 1, Aristotle thinks that the same type of inductive procedure which is used to reach and to validate those principles that are definitions is also used in the case of those principles, called 'hypotheses' in the *Posterior Analytics*, that are existential in character. In the *Analytics* Aristotle uses the term πρῶτα, the primary things, to refer to the entities introduced in such principles (I. 10, 76^a32) and he uses the same term in *Phys.* I. 7 for the principles introduced in *Phys.* I (190^b18). An example from the *Posterior Analytics* is instructive here. The basic existential principle of geometry is that two-dimensional magnitude exists (I. 10, 76^a31–6). This corresponds to the basic principle of natural science that nature, i.e. the natural world of changing things, exists (cf. *Phys.* II. 1, 193^a1 ff.). But in the case of geometry it is possible to spell this out more fully. Aristotle does so when he indicates that the fundamental existential principle of geometry is that points and lines exist (76^b3–6). This fills out the content of the principle that two-dimensional magnitude exists because these are the basic objects which make all geometrical magnitude possible. All other geometrical objects are constructible out of points and lines, but the latter are not constructible out of, or otherwise reducible to, each other or to any more fundamental entities. Just so, in natural science the principle that nature, the realm of naturally changing things, exists can be more fully spelled out as the principle that form, privation, and substratum exist. This principle is formulated at a higher level of generality than the principle that points and lines exist because the type of form or substratum, for instance, which makes natural change possible in one natural science, say meteorology, will differ from the type which makes natural change possible in another, say biology (cf. I. 7, 189ᵇ30–2). Nevertheless, in each case there will be something which plays each of these roles and which is not reducible to anything more fundamental. In *Phys.* I. 5 Aristotle makes just this point when he says that 'the principles can come neither from each other nor from anything else, and everything else must come from them' (188ᵃ27–8). This is exactly what he would say about points and lines in geometry and it is a clear indication that Aristotle is thinking of the principles of natural science which he discusses in *Phys.* I in the way he would think of points and lines in geometry. How then does Aristotle establish the fundamental principle that form, privation, and substratum exist? Does he do so by inductive argument as we have been led to expect he will?

XII Aristotle's Proof that the Principles are Contraries: *Physics* I. 5

After the dialectical refutation of the Eleatics in *Phys.* I. 2–3, which Aristotle has told us in advance will not establish any principles for the scientist, he proceeds in I. 4 to review the opinions of his predecessors as to what the principles of natural change are. He finds that they 'all in some way make contraries principles' (I. 5, 188ᵃ26). The qualification that they do this only 'in some way' is perhaps important. But, in any case, Aristotle says he does not find in his predecessors any proof or argument (λόγος, 188ᵇ29) for this sort of conclusion, so he proposes to offer a proof or argument (λόγος, 188ᵃ31) himself. The specific conclusion which he argues for in his proof is that all natural change is change in which one thing comes to be from some contrary (either from the exact contrary, as white to black, or from some intermediate contrary, as white to grey 188ᵇ21–6). What argument does Aristotle use for this conclusion? It has been suggested by some that Aristotle's argument is deductive, or syllogistic. He starts from a general principle, which he does not justify, that random things do not act on and are not acted on by random things (188^a31–3). Then he assumes, without stating it, that every becoming is an action of one thing on another and concludes that random becoming, that is, becoming between randomly connected termini, is impossible. This, then, shows that the termini of becoming cannot be contradictories since there are no causal, non-random regularities between contradictories. So the termini must be contraries since there are non-random regular connections between contraries and changes can all be described with contraries as their termini. Those who offer this analysis of Aristotle's argument often then go on to criticize him both for supposing that the fact that change *can* be described in this way shows that it *should* be, and for the assumption that the termini of change are non-randomly, that is causally, connected. They argue, for example, that the fact that something lacks the shape of a statue is 'insufficient to explain' why it comes to be a statue, i.e. to 'explain why the change *happens*'.[24]

This objection, however, and the analysis of Aristotle's argument on which it is premissed, miss what it is that Aristotle is trying to show. He is not trying to establish the principle that the contrary state from which a change originates is *sufficient* to explain why the change occurs to the other *specific* contrary state. As we have seen, the contraries are existential principles required, so to speak, to construct change. They 'explain' why change occurs in that limited sense, and not because a given contrary state is sufficient to bring it about, by a causal regularity, that some resulting contrary occurs. The parallel with geometry is again instructive. A particular circle requires points and lines, at the least a point for the centre and a line for the radius, in order for the circle to be constructed. This does not mean, however, that a given point and line are sufficient to guarantee that a given circle or any circle is constructed.

So Aristotle is not guilty of the error with which he is charged. Nor does he in fact argue from the principle that things do not act randomly on other things to the conclusion that the termini of change are contraries. He uses the former claim, in the text, only to introduce the discussion of the latter (188^a31–4). When he comes to the latter as his reasoned conclusion, at 188^b21 ff., he prefaces this conclusion with the phrase: 'if then *this* is true . . .'. The 'this' in this case, must refer to the immediately preceding material (188^a35–$^b21$). What Aristotle offers us there, however, is simply a review of a full range of cases

[24] Irwin, *First Principles*, 70–1.

where natural change (and, in a parallel way, artificial change) occurs. What is white does not come to be from just anything, e.g. from what is musical, but from what is black or some intermediate colour, and only accidentally from what is musical in case it just happens that what is some opposite colour is musical, (188^a33 ff.). What is in tune comes to be from some opposite state of lack of attunement (188^b12–15); and so on. This is the material on which Aristotle explicitly bases his general conclusion and, given this, it seems clear that his own argument (λόγος, 188^a31) is an inductive argument.

Is it an inductive argument based on things known by experience, as Aristotle has indicated it should be? Some have wanted to argue that it is not. Rather, it has been suggested, Aristotle thinks that the claim that what is white always and only comes to be by nature from some opposite colour, and only accidentally from what is musical, is an a priori or a conceptual truth.[25] But there is no clear indication in the text of the argument itself that Aristotle believes this.[26] Moreover, it seems pretty clear that this is not in fact an a priori or a conceptual truth. If it *always* did happen that what is white came to be from what is musical, that is, if all and only musical things became white owing to something about musicality, which is not conceptually impossible, then Aristotle could not say that what is white comes to be by nature only from a contrary and otherwise accidentally. For what always happens, at least in this sort of way, for Aristotle, happens not accidentally but by nature (*Phys.* II. 8 198^b35 ff.). So even if in this case what is white *also* always came to be from some opposite colour it would be an additional natural fact, and not an accidental fact, as Aristotle claims, that it comes to be from what is musical.

In addition to this, there are conceivable situations in which white does not come to be from some opposite colour in any way. For the atomists, for instance, as Aristotle understands them, white comes to be when certain atoms are arranged in a certain way. But previous to this those atoms might have been so dispersed that they gave rise to

[25] See e.g. Charlton, *Physics I–II*, 66. The contemporary notion of a 'logical' or 'conceptual' truth is notoriously problematic. Since Aristotle uses no such terminology, this is itself a reason for avoiding its use in interpreting him. I follow some minimal accepted rules for its use in rejecting its application here, however. (The notion of a priori knowledge is not as problematic.)

[26] He says: 'How *could* what is white come to be from what is musical except where it is an accident that what is not white, or black, is musical?' (188^a35–6). As a rhetorical question this expresses the ridiculousness of what is rejected but not Aristotle's basis for so regarding it.

no colour at all, or not dispersed but so arranged that they gave rise to no colour at all (perhaps by neither reflecting nor absorbing any light at all).[27] In this case white could come to be from (what is) no colour at all. A physical theory of this type is not conceptually impossible or ruled out a priori. One might say that Aristotle is rather cavalier here in ignoring this type of view which opposes his own, given that he is aware of it. But we should remember that he is confident that he has adequate arguments, which he produces elsewhere, against this type of theory.

There is, then, no good reason to suppose that Aristotle thinks that the premisses of his inductive argument are either a priori or conceptual truths. Rather, since he indicates in *Phys.* I. 1, and in parallel passages in the *Analytics*, that the scientist's starting premisses are items of experience, or of what is 'more knowable to perception' (184ᵃ24–5), we should conclude that this is the way he views the premisses of his inductive argument here. This does not mean, of course, that Aristotle's inductive argument is a good one. It may well be that something enough like the atomist's story mentioned above is not merely conceivable but really holds, so that Aristotle's argument is undermined. But if it is undermined then we will also know that fact not by any a priori reasoning but by empirical reasoning of which Aristotle himself would approve.

XIII Aristotle's Proof of the Third Principle: *Physics* I. 7

Aristotle's conclusion at the end of *Phys.* I. 5, that 'it is evident that the principles must be contraries' (189^a9–10), is, of course, not the end of the story. In fact, this 'evident' conclusion will turn out to be true only in a way. In the language of *Phys.* I. 1 it turns out that the notion of a contrary as it has been developed so far is a καθόλου in the sense of a συγκεχυμένον, which contains so far undistinguished things as parts. When, for instance, 'the musical' comes from 'the unmusical', the unmusical can and should be understood in two ways. In one of these ways it is a state which is strictly opposite to the musical state which

[27] In response to the first example it might be said that Aristotle doesn't intend his principle to cover the case of a new (white) thing coming into existence. But it is not clear that he would not have to count this example as involving a case of white (or a white thing) coming to be, i.e. of λευκὸν γίγνεται (188^a37). In any case, the second example avoids this problem.

comes to be. In another way it is the thing, for example, the person, in the unmusical state who is acted on when the musical thing comes to be.

Hence, in one way the principles are no more numerous than the contraries but are in a manner of speaking two; but they are not two without qualification, because of the difference in being which belongs to them, but rather three. For being a man and being unmusical are different. (190^b35–191^a2)

So change requires three principles, two contraries—form and privation—and a substratum which is acted on when the change occurs. Uncovering these three things serves, again in the language of *Phys.* I. 1, to distinguish the things which lie undistinguished in the καθόλου notion of a contrary.

What kind of argument does Aristotle use to justify his final conclusion that there are these three principles? As he did in working his way to his initial result that the principles are contraries, Aristotle begins in I. 6 with a discussion of the views of his predecessors. His discussion here is, again, perhaps largely dialectical, but the most that he draws from it is that his own view 'has some argument' in its favour but still 'is very problematic' (189^b17–18, 29). In I. 7, following his earlier pattern, Aristotle moves from this inconclusive discussion of his predecessors to produce a conclusive argument of his own (189^b30 ff.). It is important to look at the way in which he proceeds in the crucial part of this argument (I. 7, 190^a31 ff.).

What comes to be is so-called in many ways. In some cases there is no coming into being but a particular thing comes to be *something*. There is coming into being simply only in the case of substances. In the other cases it is obvious that the *thing* which comes to be something must serve as subject. For a quantity, or a quality, or a relation, or a time or a place comes to be [a quantity or a quality, etc.] of some subject; since only a substance is ascribed to nothing else as its subject, while everything else is ascribed to the appropriate substance. But that substances too, and whatever else exists without qualification, come to be from some subject is obvious on review. For in every case there is something which serves as subject, from which the thing which comes to be does come to be—as for instance plants and animals come from seed. Some of the things which come to be without qualification come to be by change of shape, for instance a statue; some by addition, for instance things which grow; some by taking away, for instance the Hermes from the piece of stone; some by composition, for instance a house; some by alteration, for instance things which change their matter. It is obvious that all things which come to be in this manner come to be from subjects. Thus it is evident, from what has been said, that *everything* which comes to be is always composite: there is something which comes to be and something which comes to be that—the latter in two ways, either the subject or what is opposed. (190^{a}31–^{b}13)

As with Aristotle's earlier proof that the principles are contraries, here also his proof that every case of coming into being involves three things is based on a review of particular cases. It is the fact that those cases are obvious (φανερά) cases of the general conclusion that leads Aristotle to draw that conclusion as something which is evident. So Aristotle's argument is, again, inductive in form. There has been no temptation to claim here that Aristotle thinks our knowledge of the particulars on which the generalization is based is conceptual or a priori. That plants come from seed, that things require the addition of material to grow, that houses are produced by putting building materials together; these are not conceptual or a priori truths.[28] Nevertheless, some have wanted to argue this cannot be Aristotle's real argument for his conclusion since such an empirical argument could not explain why Aristotle is so confident of his conclusion. Rather, it is suggested, he must have in mind the following a priori argument:

> Considering now the mere concept of becoming we may argue that if one thing is properly said to become another then obviously ... there must be something which does persist throughout the change, for otherwise the change would merely consist in one thing coming to be where another had ceased to be and there would be no reason to say that the one *became* the other. This argument appears to have an *a priori* certainty.[29]

The first thing to be said about this argument is that it is not the argument which Aristotle uses. The argument which he uses is the one quoted earlier. Secondly, it is not in fact clear that the 'mere concept of becoming' does require that something persists throughout any case of becoming. It is conceivable that cases of becoming could involve no single persisting object but rather only sequences of non-persisting objects (call them, say, space-time slices) which are ordered in a certain way. It is the proper ordering of the non-persisting objects

[28] Aristotle prefaces this argument by making certain distinctions, based in part on things 'we say' (189^{b}32 ff.); e.g., 'A man becomes musical' or 'An unmusical man becomes a musical man'. (Here and above I follow the traditional rendering 'musical' for the word which is perhaps better translated 'cultured'.) Aristotle does clearly think that various things of this sort, which we say, are obvious truths but he can hardly think that these particular things are known a priori or that it is inconceivable that they should be false.

[29] D. Bostock, 'Aristotle on the Principles of Change in *Physics* I', in M. Schofield and M. Nussbaum (eds.), *Language and Logos* (Cambridge, 1982), 186.

rather than the presence of any single persisting object which, on this proposal, distinguishes cases of becoming from cases where things merely succeed other things. Nothing in our ordinary understanding of becoming rules it out that we might discover, or that some already have discovered, that the actual cases of becoming which we are all aware of are like this. Since Aristotle does not himself argue that this is ruled out on conceptual or a priori grounds but only that it is ruled out by a proper review of actual cases we should not attribute such an argument to him. It is easy to feel, of course, that Aristotle's review of actual cases here is rather fast and loose. But the explanation for his confidence could be, not that he has an a priori argument up his sleeve to use if challenged, but that he can call on extensive acquaintance with the phenomena with which the natural scientist deals to rebut any proposed counter-example to his principles. This is, in any case, the appropriate response for him given that as a natural scientist he reaches his own knowledge of his principles by induction.

XIV Conclusion

Aristotle's method for the natural scientist, then, as he describes and uses it in the opening book of the *Physics*, is not an a priori method or a method which is appropriate for philosophy as opposed to science. However crude his description or use of the method may be his endeavours belong, in spirit, with those which we now think of as clearly scientific.[30]

[30] An earlier version of this paper was presented at a special session of the Los Angeles Area Colloquium in Ancient Philosophy on Aristotle's *Physics*. I am indebted for very helpful comments, on that occasion and subsequently, to James Bogen, David Charles, Cynthia Freeland, Lindsay Judson, Frank Lewis, Richard McKirahan, Michael White, and Charles Young.

2

What Makes Reality Intelligible? Reflections on Aristotle's Theory of *Aitia*

JULIUS M. MORAVCSIK

To say that some parts of reality are intelligible is to say that we can understand these and can offer explanations for them. Understanding and explaining are what we do; reality provides the objects for these states or activities. This suggests that theories of reality are one thing, while accounts of explanations and understanding constitute something else. The former embraces science and metaphysics, the latter a part of epistemology and the philosophy of mind. From the point of view of a conception adopting this bifurcation Aristotle's theory of *aitia* seems puzzling. For on the one hand it is a theory about various configurations of elements of reality, and on the other hand it seems to be a theory about explanations. We can, however, remove the puzzlement if instead of asking, 'How do we explain various parts of reality?' we ask, 'What features of parts of reality make these intelligible?' In this essay Aristotle's theory of *aitia* will be interpreted as an answer to this question. The answer is, roughly, that just as certain configurations of elements of reality make corresponding statements true, so certain configurations of elements of reality make some explanations adequate, true, and insightful. Aristotle's theory of *aitia* is a correspondence theory of explanations, thus dealing at once both with the world and with the ways in which we explain the world. Furthermore, for Aristotle the intelligibility of the world depends partly on fundamental changes and processes in the world being the manifestations of powers and potentialities that have an intelligible nature of their own.

Some will object immediately to any attempt to uncover the ontological configurations underlying adequate explanations by saying that explanation is a many-splendoured thing, and thus there exists a wide variety of configurations that correspond to explanatory patterns. Is there any non-vacuous characterization of the basic relations structuring the relevant configurations? Aristotle's answer is the affirmative, and his examples show that he is casting a wide net indeed for the range of phenomena his theory is to cover. For example, his account analyses such diverse configurations as propositions arranged in the pattern of premisses and conclusion, the constituents of biological specimens and the developing specimens themselves, and the relation between human actions and their initiating antecedents.

One might still feel that explanations are in the head while parts of elements of nature and their developments are in the world. Which is the proper subject-matter for Aristotle's theory? Or is it the case that his theory does not quite fit either mind or matter?[1] Asking, however, 'What are the constituents, aspects, and antecedents of a human action?', for example, is not only to raise a sensible question about ontological configurations, but at the same time—given an assumption about correlations between structures in reality and structures in human understanding—it is also to ask a searching question about explanation and understanding. Answering the question in the 'material mode' gives us data for giving an account of what we regard as adequate explanation in this kind of case.

The subsequent sections will show how Aristotle can give a non-vacuous general account of fundamental types of explanations and their ontological correlates.[2]

I What is in the Head and what is in the World?

Before we explicate the Aristotelian position, we need to lay aside two issues. First one might raise the Kantian claim that there is a dichotomy between the noumenal and phenomenal world, and that the configurations mentioned apply only to the latter. This issue needs to be laid aside, for it never surfaces within the realist conception of the world and the human mind that is adopted by Aristotle and his predecessors. Similar considerations apply to raising Humean doubts about the reality of causal powers in contrast with mere temporal conjunctions of events. As we shall see, Aristotle interprets natural

[1] This question is raised in J. Annas, 'Aristotle on Inefficient Causes', *Philosophical Quarterly* 32 (1982), 311–26.

[2] For an extended account see J. Moravcsik, '*Aitia* as Generative Factor in Aristotle's Philosophy', *Dialogue* 14 (1975), 622–38.

changes as the realization of potentialities. His analysis of change leaves room for something like the modern notion of an event, but these entities do not have an independent status in his ontology. A realist conception of potentiality is a presupposition rather than a consequence of Aristotle's views on what makes the world intelligible. This presupposition was not challenged by his contemporaries, though some of his predecessors tried to avoid the notion of potentiality altogether in their ontologies.

Given, then, this much 'realism', what is an Aristotelian *aitia*? One of his characterizations looks suspiciously epistemological (194^b18–20). For it says that we do not really understand what something is until we know or grasp what its primary *aitia* is. The 'something' in this formulation is supposed to include also genesis and destruction; we need to understand the *aitiai* of these items as well (194^b20–2). However further evidence shows that Aristotle has in mind the grasp of ontological entities and their configurations, and not mental or linguistic items. For in 198^a14–16 he says that the different kinds of *aitiai* correspond to different kinds of items that answer *dia ti* questions. The lines following give us some examples (16–21). These items are entities in the world. For example one of the *aitiai* is a raid that prompted people to go to war. Another is health as a goal or pursuit of people engaged in certain activities. Again, the essence, or nature, of something can bring about ways in which the object in question develops or grows. For essence or nature makes an element of nature into a self-determining entity with its own schedule of growth and development. *Dia ti* questions are, then, asking on account of what entity and relationship events take place, actions are performed, or substances are the way they are. The answer singles out the appropriate entity and relationship. This shows that it is misleading to characterize the *aitiai* as 'answers to why-questions' or 'the content of because-clauses'. For both of those items are mental or linguistic, whereas the examples show us that the *aitiai* are not.

There is nothing strange about Aristotle's way of looking at these things. We often point to elements of reality that are in some ways responsible for other elements. Here are a few simple examples. In our effort to understand in some context why people are dying we single out a drought, and say, 'That explains it'. In our attempt to understand why the car overheated, we point to a leak in the radiator and say, 'That accounts for it'; and the plumber pulls out something from the pipe and says, 'That accounts for the blockage'. In a medical school the research team looks not for linguistic items but for virus or bacteria that will account for certain illnesses.

In all of these cases our search is prima facie for elements of reality that stand in certain relations to other elements of reality. We may be able to analyse these efforts in what is called in contemporary philosophy the 'formal mode'. But even if such attempts were successful, they would not destroy the thesis that in many of our endeavours to understand nature we are in search for elements of reality and their interrelations rather than for explanations interpreted as mental or linguistic items. Just as some philosophers assume that corresponding to a simple true assertion of subject-predicate form such as 'Socrates is wise' there is a particular that has the attribute of wisdom, so Aristotle thinks that certain configurations of elements of reality underlie and correspond to certain explanations.

In proposing what might be called a correspondence theory of explanation Aristotle's motives are not linguistic. Quine once suggested that just as we do not posit 'sakes' in order to explain the meaning and truth of sentences involving the 'for the sake of' construction, we need not posit other—from his point of view suspicious—entities in our semantic analysis. Aristotle posits configurations of elements of reality corresponding to adequate explanations not in order to assign reference to certain linguistic expressions, but to account in a realist way for that in virtue of which some explanations are adequate and yield insight.

Intelligibility contrasts not only with unintelligibility, but also with mere describability. For example, a world consisting of independent singular facts of the '*a* is *F*' type would be describable but not subject to explanations.

Let us suppose that one could sum up all of the truths about the world in a book containing only a series of singular propositions of subject-predicate form. (We shall bypass the issue of identity, and take that for granted.) We could then also form, on the basis of the book, generalizations of the form 'all *A*s are *B*s'. Both science and common sense are not satisfied with such generalizations. They seek generalizations that have law-like qualities, i.e. that admit of counterfactual formulation as well. For example, we are not satisfied with noting that all leaves at a certain time are green. We claim that even if the leaf population were not identical with the one that happens to exist in our world, the statement would still be true. The reason for this is that we think that the colour of leaves is related to important processes that take place in plants such as photosynthesis. But if we treat this generalization as law-like, it makes sense to ask what in the world makes it a true law-like generalization? In response to this question we posit dispositions.

Some dispositions may be acquired as a result of interaction with the environment, while others are in the nature of the entity in question. The latter form the core of what Aristotle calls potentialities. We can explain changes as the realization of potentialities, and then account for the potentialities in terms of what initiates these, what their completion amounts to, and what structures they require. These layers of entities and structures renders the world intelligible in Aristotle's view. Adequate explanations must reflect the nature and effect of potentialities.

Sections 194^b23–5 and 195^a15–26 present us with an account of the basic *aitiai*. One kind of *aitia* is constituency. As we shall see, this notion as used by Aristotle is much wider than the modern notion of material. Another *aitia* is the motion- or process-initiating factor. This has been identified at times with the modern notion of efficient cause. I have argued elsewhere that this identification is misleading at best and false at its worst.[3] Within Aristotle's non-mechanistic teleological conception of reality there are many factors that initiate change and process that cannot be described by the modern term 'cause'. We see causal links primarily as links between events. But Aristotelian process-initiating factors can include the skill of a craftsman, a substance, as well as a realization of a potentiality. Another *aitia* is the 'form' of an element. We shall explore this difficult notion subsequently, and show that its proper interpretation is in line with the overall interpretation of *aitia* presented here. Finally, function and goal also serve as *aitia*, since these too contribute to bringing about processes and activities.

Aristotle does not say that every part of reality must have in its nature all four of these *aitiai* in order to be intelligible. Numbers do not have motion-initiating factors or goals, for example. But to the extent that things are intelligible, the conditions of such intelligibility will involve some of these four factors.

This brief review of the *aitiai* shows that Aristotle's theory of these is about intelligibility and hence both about the structure of the objects as well as the structure of accounts that humans need to attain.

[3] For a discussion see J. Moravcsik, 'Aristotle on Adequate Explanations', *Synthese* 28 (1974) 9.

II Processes and their Descriptions

We saw that according to our interpretation Aristotle's candidates for adequate explanations should not be description-dependent. They should simply single out the element responsible for another element in the way specified by the explanation. In view of this we need to pay attention to one of Aristotle's examples of *aitia*-giving in which a man called Polyclitus is said to be responsible for a certain statue coming into being (195^a33-^b3). On Aristotle's account, to say that Polyclitus was the process-initiating factor is less illuminating than to point out that Polyclitus the sculptor brought the statue into being. Does this not show that *aitia*-giving is description-dependent, since presumably Polyclitus = Polyclitus the sculptor? This objection assumes that Aristotle would agree to the modern versions of the semantic theory of identity.[4]

To understand Aristotle's approach let us look at a substance like Polyclitus from a metaphysical point of view. Within that perspective the deceptively simple proper name 'hides' a highly complex entity with many parts and potentialities and attributes. Different parts and different attributes account for different abilities and activities that Polyclitus performs or possesses. Modern common sense, science, and much of philosophy would surely agree with Aristotle.

Presumably moderns and Aristotle would also agree that referring to an entity is not the same as referring to it and all or some of its parts. In referring to a person Jones *simpliciter* I am not also referring to his arm. This accounts for the contrast between

(1) Jones was responsible for the cricket victory.

and

(2) Jones and his strong arm were responsible for the cricket victory.

The two statements are clearly not equivalent. Sentence (2) entails (1) but not the other way around. We can see this also by reflecting on

(3) Jones and his keen eye were responsible for the cricket victory.

[4] For further discussion of this issue, see Cynthia Freeland's essay, 'Accidental Causes and Real Explanations', Chapter 3 in this volume, section I; for some of the recent literature on the subject, see nn. 15 and 16 of her essay.

Sentences (2) and (3) are not equivalent. They point to different factors leading to a victory. Aristotle would say that the two statements propose different *aitiai* for the same outcome.

Aristotle's approach to examples like the one involving Polyclitus can be understood if we think of him as treating attributes and powers in the same way as we would treat parts. Indeed such a move is not unnatural. Let us consider

(4) Jones and his shrewdness were responsible for the cricket victory.

It seems entirely natural to treat the reference to an attribute here in the same way as the references to parts in (2) and (3).

Thus I suggest that we read:

(5) Polyclitus the sculptor was responsible for the statue.

as equivalent to:

(6) Polyclitus and his sculpting were responsible for the statue.

These reflections show that Aristotle's *aitia*-giving is not description-dependent, though it can be described as 'feature-dependent' in the sense that the *aitia* will single out not only an element but in some cases also a part or feature of the element that played a vital role in the generation or destruction to be accounted for.

One might ask why we should not simply say that Polyclitus' art was responsible for the statue? Such a view would admit of two interpretations. According to one of these the art referred to is something general that can be shared among many practitioners. According to the other reading the art in question is something unique to Polyclitus; a specific non-substantial particular that only he possesses. (On non-substantial particulars see Strawson.[5]) While the second reading would please a typical modern audience obsessed with the importance of individual talent, genius, etc., the first is much more likely to be what Aristotle, or earlier Plato, would have had in mind.

Fortunately we need not decide the issue. For on either reading we would miss an essential component of Aristotle's conception of this kind of *aitia*. Arts—universal or particular—cannot by themselves initiate processes. They can do that only in so far as they are embodied in substances. Thus a complete characterization of the 'source' ('*archē*') of the statue involves both the art of sculpting and the agent practising it. Thus 'Polyclitus the sculptor'—and also the more direct 'a sculptor'—

[5] P. F. Strawson, *Individuals* (London, 1959).

must be interpreted so as to introduce both the relevant agent and his art.

These considerations show that *aitia*-giving even in this case remains ontological and is feature-dependent. How Aristotle's insight might or might not be reconcilable with modern semantic notions of reference and intensionality is beyond the scope of this paper.

What was said here about the creating agent applies also to the object produced. The statue produced may be the ugliest object in the room, but for the purposes of explicating the creative process and generation in question, it is the object being a statue that is relevant. Under normal circumstances the sculpting ability is exercised to produce a sculpture. This is the part of the nature of the entity that the art in the artist produces. Being the ugliest entity in the room is merely an accidental feature of the object and is not directly related to the exercise of the sculpting ability.

Distinguishing the object as a whole from its parts, properties, potentialities, etc., does not introduce an ontological jungle into Aristotle's philosophy. If we analyse an opera into acts, arias, chorus, etc., and realize that to refer to the opera is not *eo ipso* a reference to its parts, aspects, etc., this is not to lapse into ontological promiscuity, but simply to realize that operas are complex entities, with different constituents playing different roles. Aristotle does not wish to sweep this fact under the semantic rug of simply naming things.

III A Role for Events in a Metaphysics of Power and Potentiality?

As noted above, in modern philosophy accounts of changes are given typically as relations between events. Does Aristotle share this view? If not, what is his 'realist' account of changes? Does he have a notion similar to causal relations between events? Julia Annas suggests a negative answer to this question.[6] The closest item among *aitiai* to efficient cause is the motion-initiating factor. We saw, however, how problematic any identification between these notions is.

We shall leave efficient causes out of the picture, and concentrate on the question of what the closest analogue to an event is in Aristotle's ontology. Let us lay aside conceptions of events as mere passings of

[6] J. Annas, unpublished paper on Aristotle and causation.

time, or as the instantiations of attributes over time even in an unchanging world, and consider for our purposes events as changes of location or quality over time. In this sense Aristotle does acknowledge events, since he acknowledges the existence of changes and aims at an account of these. However, there are no 'mere' changes in Aristotle's ontology and the analogues to changes do not have an independent ontological status.

Changes for Aristotle are parts of processes. A process, in turn—such as the development of an animal, the building of a house, or the returning of foliage in the spring—is given a partly teleological or functional interpretation. Shall we say, then, that Aristotle has events in his ontology, but that these have necessarily teleological orientations? This would still not be satisfactory. One could have a conception of a world in which events cause events, and the whole historical process tends towards realizing a cosmic purpose, e.g. the will of God. Such a conception would be markedly different from that of Aristotle. For he interprets the process as constituted by sequences of changes in which potentialities that are unique to species or their genera are realized or brought to fulfilment. Substances constitute—ultimately—these kinds. Hence changes in the normal course of history are the result of self-determined schedules of potentiality realizations of substances and interactions between substances. Thus it is misleading to think of Aristotelian changes as causing each other. The changes have as one of their sources potentialities, and these in turn are ingredients of other ontological kinds.

Still, there is a link between modern causal explanations and Aristotelian explanations of change and locomotion. For if we add up in any given case the potentialities that are to be realized in that context and the conditions that result from the potentialities not being hindered, then this set of conditions is equivalent to what in at least some modern schemes would be as causally necessitating some changes. Thus one might say that there is 'extensional equivalence' between some versions of modern causal theories and the Aristotelian account of substance, activity, and potentiality that bring about a change, but in terms of analysis the two conceptions differ.

These reflections shed light also on why for Aristotle there should be non-eventlike entities as motion-initiating factors. For though his conception of nature posits chains of changes, there are additional factors that contribute to the coming into being of any one member of such a chain. This is so because in the normal case a change is a partial realization of a potentiality, and different potentialities have different kinds of sources. In the case of procreation, which is a part of essential human functioning, the key motion-initiating factor is the father, since—according to Aristotle—he is the substance whose active power is being realized. In the case of processes that are not essential to human functioning, such as sculpting, the motion-initiating factor is an art, since that is the unique element needed to combine with human agency—a condition for producing artefacts in general—to bring about that particular kind of artefact.

Some passages suggest, however, that there is no room in Aristotle's ontology even for processes. In *Phys.* 200^b32–201^a2 Aristotle says that there are no processes 'in addition' to the things (*pragmata*) that undergo these. This, however, does not mean that Aristotle denies the existence of changes and *kineseis*. Let us return to the point made earlier about a substance being a complex entity, made up of parts, attributes, and potentialities. The potentialities of a natural substance are realized according to a self-determined schedule. Within this conception of nature, once we obtained a complete description of every active substance there would be nothing left over, waiting to be accounted for. The full account of any substance would include its life history and its interactions with other substances. Thus Aristotle does not mean to adopt what in modern parlance would be a reductionistic programme for changes. His view is, rather, that once we have a full account of substances, giving an additional account of all of the natural *kineseis* would be to count some elements of reality twice. Ross suggests that the *pragmata* in this passage are not the things undergoing change, but respects in which change occurs, thus referring forward to the reference to four categories in the following lines. But on that reading—as Ross admits—it is difficult to find a motivation for the inclusion of this statement. The reading proposed in this paper motivates the statement, since it places it within the framework of Aristotle's substance-oriented ontology. Since all changes are related to various elements in some of the categories, there are no processes over and above the things undergoing change; all changes and processes are accounted for in a 'full' description of other elements in Aristotle's ontology.[7]

[7] See W. D. Ross, *Aristotle's Physics: A Revised Text with Introduction and Commentary* (Oxford, 1936), 536. My interpretation assumes the admittedly controversial thesis that Aristotle regards artefacts not as genuine substances but only analogues of these. They are like real substances except that their 'soul' is the artist or in the artist, and thus

Further evidence for this interpretation is provided by Aristotle's locating changes and processes in the categories of substance, quantity, quality, and place.[8] Processes too have essences and are redescribable in ways analogous to the ways in which items in other categories are—although, to be sure, they do not have a full essence, since their nature and being depends on the nature and being of substances.

These considerations show that there is a place for event-like entities in Aristotle's ontology. For at least some events are changes, and Aristotle not only acknowledges these as having an ontological status, but also has an account of them in terms of sequences of potentiality realizations. These changes and processes can be accounted for—in the sense explained in this paper—by other elements of reality. They, in turn, have also an explanatory role, helping to account for the nature of the potentialities that are—in the normal case—their sources.

IV The Ontology of Explanatory Chains

According to the interpretation presented one entity can account for, or be responsible for, another entity. But if entity *B* explains entity *A*, do we not need entity *C* to explain the nature of entity *B*? Aristotle's conception seems to lead to explanatory chains within which a given

outside the entity. This deprives artefacts of having a self-determining nature and a developmental schedule. But it is still true of them that they are brought about by the four familiar factors, and that these are also responsible for their maintenance. They are brought about by an agent—the instantiated artistry of the artist—and have a form and function, though this is designed by the artist and not by forces and structures within the object. The matter of these objects consists of the realizations of capacities to endure. These must include their weight, hardness, etc. Without the constant realization of potentiality resulting in constancy in stuff, weight, etc., the artefact would not survive. But since the 'soul' is outside, maintenance and persistence are not the dynamic processes that they are in the case of biological substances. In the biological case constant motion and activity are required for persistence; in the case of artefacts stability is required for persistence, and this is achieved not by constant self-determining activity within the object, but by the static realizations of the capacity to survive, resulting in at most minor changes in hardness, weight, etc.

The analogy with artefacts is used by Aristotle in order to give intuitively simple and persuasive introductions to the four *aitiai*. It is unlikely to be a matter of different developmental stages, since all the basic Aristotelian ontological constructions already have their ancestors in Plato's thought. On this point see J. Driscoll, 'The Platonic Ancestry of Primary Substance', *Phronesis* 24 (1979) 253–69.

[8] *Cat.* ch. 14.

explanandum has another entity as its explanans and this in turn another one. Where do these chains end? Sooner or later some entity needs to be taken as ultimate or self-explanatory. Within modern non-entitative conceptions of explanations this does not cause a problem. Explanations that humans give can end for a variety of reasons. These can be conventional delineations of departmental boundaries, lack of interest in further pursuits, or historically and culturally limited notions of what is fundamental. There may be more to it than that, but there need not be. However, any such termination of Aristotelian entitative explanatory chains is unacceptable: entities in the world cannot cease to exist or become ultimate and self-explanatory on account of barriers between disciplines invented by humans. We need metaphysical grounds for terminating entitative explanatory chains.

Plato's account of what makes the world intelligible (e.g. *Phaedo* 96–106) is a striking example of entitative explanations. He interprets the Forms as underlying whatever order and stability the world has, in virtue of the unique natures the Forms have and the relationships that some elements of reality can have with them. The unique nature of the Forms is discussed in recent literature under the heading of self-predication. This is misleading, as it suggests semantic considerations. A better interpretation is suggested by reflection on Plato's characterization of these natures as something complete that can be at most be reflected partly by other elements of reality. Our grasp of these natures enables us to see how and why a Form is ultimate and self-explanatory. The question 'What makes a Form ultimate and self-explanatory?' admits of possible answers varying greatly in terms of subtlety. We need not assume that whatever accounts for the equality of two things must have in a quantitative sense 'a lot of equality', or be 'very equal'. We do owe, however, some metaphysical account of why 'the buck stops here', and Plato attempts several times to address this issue. A detailed examination of his efforts is beyond the scope of this paper.

Plato regards the soul as self-moving (see, e.g. *Phaedrus* 245c–e). Why should the soul not be another end of an explanatory chain? If Plato had agreed, he would have had more than one kind of explanatory chain. But Plato construed the soul as dependent on the Forms, both for the objects of cognition and for the harmony and order that its functioning requires. Thus his remained a unitarian theory of explanatory chains.

Aristotle differs from Plato in two main respects. First, his theory of *aitia* is a pluralistic theory of explanatory chains. Second, he separated the question of what is ontologically fundamental from the question of what is ultimate in terms of explanatory chains. This required that he reconsider the notion of self-sufficiency, and not share Plato's simple and austere view on that. However, this topic too would take us too far from our main subject.

The different *aitiai* constitute for Aristotle different entitative explanatory chains. For example, flesh with its dynamic structure stands in the constitutive relation to the human body, helping to account for its functioning. A male human stands in the motion-initiating relation to his offspring, contributing to the full account of generating a new substance. Health stands in the function or aim relation to humans and some of their activities. Yet none of these entities need be for Aristotle ontologically fundamental.

In this way Aristotle can have four different explanatory chains, and yet insist that substances are what is ontologically fundamental. All else depends on substances in a way in which substances do not depend on other things. Spelling out this asymmetry, and listing the various proposals as to what Aristotle really thought substances are, cannot be done here. But this interpretation of *aitia* is compatible with the main lines of interpretation concerning these other topics.[9]

Aristotle's scheme also admits hierarchies of, for example, constitutive relations. The constituency of an element *E* has in turn other constituents, and these have constituents as well. But being the final item in such a chain is not what endows an entity with being ontologically fundamental.

Someone might ask for explanations of notions like constituency or goal and function. For Aristotle, however, these remain basic, non-definable notions. He explains explanations in terms of these and not the other way around. A Kantian might posit these as basic innate human ideas in terms of which we organize experience. But given Aristotle's 'in-put' theory of concept formation he would locate these elements not in the mind but in the world.

V Form and Matter as *Aitiai*

In this section we will see how form and matter fit into the roles assigned by this interpretation to *aitiai*. This might seem puzzling at

[9] For more discussion see J. Moravcsik, 'Essences, Meanings, and Generic Propositions', forthcoming.

first. How could bits of stuff do any explaining? Why should attributes account for natural processes and functioning? Finally, how do attributes and material fit into an ontology that according to our interpretation centres on power and potentiality?

We shall begin with matter. Is Aristotelian matter something like material in the modern sense?[10] Some examples might support such a view. In *Phys.* I. 7 bronze is given as an example of matter. On the other hand, in *Met.* 1045^a20–5 matter is identified with potentiality. How are we to reconcile this evidence? Assuming as we have all along that artefacts are not real substances, the way out is not difficult. The second passage applies to natural entities and their aspects while the first only to artefacts that are brought into the picture by Aristotle only to illuminate some aspects of his concept of a substance. With respect to the notion of matter the analogy is infelicitous. For a bronze statue does not contain a self-determined schedule of development and maintenance as the biological elements do. The analogy is good only for pointing out that just as not anything can develop into this or that animal, so not anything can serve as matter for a statue. The dynamic aspect of existence and persistence, so vital to living things, is absent in the cases of artefacts.

Others have noted already that Aristotle's notion of matter is much wider than our notion of material.[11] For it includes also abstract elements, such as premisses in an argument, and letters in a syllable (*Phys.* 195^a15–20). Even in these abstract cases we see the work that potentiality does. Letters have no independent status in spoken language. Their role is simply that of potential contributors to syllables.

In the core cases, namely those of biology, the identification of matter and potentiality is easy to see as long as we keep in mind that biological entities can be and remain what they are only by constant activity, and that these activities constitute the sequential—or gradual—realizations of potentialities whose source is in these entities themselves. Frozen at a moment we can see a substance as informed potentiality, i.e. an actual so-and-so. But over stretches of time the substance is a series of potentiality-realizations, hierarchically arranged.

[10] For a defence of a negative answer to this question see A. Code, 'Soul as Efficient Cause in Aristotle's Embryology', *Philosophical Topics* 15 (1987) 51–9.

[11] Ross, *Aristotle's Physics*, 293.

Matter for animals, such as flesh, seems to be just what we call stuff or material. But underlying this appearance there is constant activity needed to keep flesh a living tissue. The flesh as informed matter and the potentiality-realizations coincide, as a modern semanticist might put it.

Given this interpretation of matter as potentiality we can see how the entitative explanatory theory applies to this *aitia*. We understand what a specimen of a biological species is by understanding the powers and potentialities that are realized in its existence and life-span.

Needless to say, a full account of Aristotelian matter would require much more to be said. These sketchy remarks, and the ones following on 'form', are meant to show only how adequate interpretations of these notions make it possible to fit these elements of Aristotle's metaphysics into the entitative account of *aitia* presented here.

The conception of Aristotelian matter sketched helps in the understanding of the Stagirite's notion of 'form'. Again, some considerations pose initial difficulties. Shape is given as one of the examples, and if one thinks of 'form' as attributes in general, it is difficult to see how these are necessarily related to potentialities and how they could fit into explanatory roles related to these. But shape is given as an example only in connection with artefacts like statues. As we saw, their 'matter' contains potentialities only in a very weak sense. Furthermore, as we shall see, being an attribute is not what is essential to an Aristotelian form.

In *Met.* 1041^{b}7–8 and 27–8 form is characterized as that by reason of which this kind of matter becomes this definite thing. This is hardly a characterization of attributes in general. Rather, it links certain items closely to potentialities. Forms are invoked not to answer a nonsensical question like 'What makes this actual thing actual?', but rather to answer the question 'What is responsible for the realization of this potentiality as a part of something coming into being or persisting?'

In the light of this passage, we can understand the description we find in *Phys.* II. 7, where form is described as whatever accounts for an entity being what it is or for its functioning. If forms are invoked to answer the question 'On account of what can this thing be what it is and do what it does?', then surely items from a variety of modern ontological categories could fill the bill. Different kinds of items will be required in different contexts.

For example, in *An.* 412^{a}19–22 the soul is described as the form of a natural body potentially having life. This is, then, what makes the potential living thing into an actual one. Thus it is not surprising to find in 413ª25–8 that the soul is partly a power, partly activity, and partly a structure. These are the items that make something potential into an actual living thing. (The same point is made at 424ᵇ1–3.)

Structures, activities, and powers all satisfy the formula for 'form'. Furthermore, there is no good reason to suppose that Aristotle thought of all of these as universals or attributes. In the case of powers and activities it is just as easy to think of these as 'mass-terms', i.e. the sum of what is denoted by 'activity', rather than what can be divided into universal and particular.

Form, then, is whatever actualizes a potentiality, thus leading to the development and flourishing of an element in nature in the primary cases, and of items that are related by virtue of various dependency relations to parts of nature. So understood, Aristotelian form can easily play the role required of it within an entitative explanatory theory. One can point to an activity or structure and say, 'This is what makes this a flourishing human, beaver, tree, etc.'.

VI Conclusion

Once we understand potentiality, actualization, and the priority of what is self-determined in Aristotle's scheme of things, we can make sense of his theory of *aitia* as an entitative explanatory theory, differing from Plato's in its pluralism, and its separation of what is ontologically prior from what is prior in terms of explanatory role.

If we try the entitative explanatory theory on a collection of notions including thing, event, and causality in the modern sense of these expressions, we run into trouble. Why should events account for each other? How do events account for things? Causal links presuppose events and things rather than serve as a constituent helping to account for them.

If we exchange things and events for powers, potentialities, and their realizations, then we can understand Aristotle's *aitiai* as the basic ingredients in potentiality realizations. There must be a source for the potentiality, there must be something that brings it to actuality, and there must be a 'goal', or some final state which is the culmination of the realizations. Within this dynamic picture, the different elements of reality discussed in this paper do serve in explanatory and accounting roles. Aristotle's dynamic conception of nature turns out to be more like Heraclitus' dynamic ontology than the 'thing-metaphysics' of Descartes, Locke, or in our times Strawson.[12]

[12] I wish to thank Nancy Cartwright and Susan Levin for useful suggestions.

3

Accidental Causes and Real Explanations

CYNTHIA A. FREELAND

Introduction

Among Aristotle scholars current opinion is virtually unanimous that the theory of the four causes is really a theory of explanation. Julius Moravcsik, for example, writes that the theory, 'properly understood', 'is the Stagirite's account of what constitutes an adequate explanation'.[1] Julia Annas remarks in a similar vein, 'It is a great improvement to cease thinking of an *aitia* as a cause and to treat it instead as an explanation, a "because".'[2] A host of other Aristotelian scholars have joined in, as has the philosopher of science Bas van Fraassen, who sees in Aristotle's texts a theory of causes emphasizing, like his own, 'the pragmatics of explanation'.[3]

* An earlier version of this paper was read at the Central Division meetings of the APA in 1986; I am grateful to David Charles, my commentator on that occasion, for his criticisms, and to others present, in particular Dan Graham, for their comments. A later version was read at a conference on Aristotle's *Physics* co-sponsored by Rutgers University and USC in 1989; I would also like to thank the participants at that conference for their questions and suggestions. Finally, I am indebted to Lindsay Judson for a number of helpful suggestions about improvements in the structure, style, and substance of this essay.

[1] Julius Moravcsik, 'Aristotle on Adequate Explanations', *Synthese* 28 (1974), 3. Moravcsik's view is presented in a somewhat modified form in 'What Makes Reality Intelligible? Reflections on Aristotle's Theory of *Aitia*', in this volume. I will refer to this more recent paper in more detail below.

[2] Julia Annas, 'Aristotle on Inefficient Causes', *Philosophical Quarterly* 32 (1982), 319.

[3] Jonathan Barnes, *Aristotle's Posterior Analytics*, Clarendon Aristotle Series (Oxford, 1975), esp. p. 96; William Charlton, *Aristotle's Physics I, II* (Clarendon Aristotle Series; Oxford, 1970); Martha Nussbaum, *Aristotle's De Motu Animalium* (Princeton, 1978); Richard Sorabji, *Necessity, Cause and Blame: Perspectives on Aristotle's Theory* (London and Ithaca, NY, 1980), esp. chs. 1, 2; and Bas C. van Fraassen, *The Scientific Image* (Oxford, 1980), esp. ch. 5; and also 'A Re-examination of Aristotle's Philosophy of Science', *Dialogue* 19 (1980), 20–45. I should note that these writers (Moravcsik in particular) do not assume it is at all evident what is meant by saying that the four αἰτίαι are explanations, or what form explanation takes in Aristotle. For an especially clear answer to these

The motivation for the explanation approach seems mainly negative. Advocates of this approach deny that Aristotelian αἰτίαι are causes because they take it as obvious that (at least) three of the Aristotelian αἰτίαι differ quite radically from causes as 'we' conceive them (necessary or sufficient conditions, events constantly conjoined to effects, events probabilistically conjoined to effects, or others). On the explanation approach we can preserve some parallels between Aristotle's and 'our' accounts of causes, provided we reserve use of the term 'cause' for what he calls moving or efficient causes. Then presumably the other αἰτίαι (formal, final, and material) fall into place in this picture by being interpreted as explanatory factors relative to this one 'real' (i.e. efficient) sort of causal relation that seems most to resemble our own.[4]

I believe that several serious problems for the explanation approach have not been well thought out in the literature. Three puzzles in particular concerning the positive commitments of this approach deserve further critical attention. The first problem for the explanation approach concerns its understanding of Aristotle's views on the metaphysics of causal relations. There is no clear, well-defended consensus about whether Aristotle should be understood to be a realist about any or all of the four types of causal relations. Further questions can be raised about the sorts of entities he takes to be involved in causal relations, and about whether he understands causal contexts as intensional or extensional.

The second problem with interpreting the canonical presentation of the doctrine of the four cases in *Phys.* II as a theory of explanation is that Aristotle elsewhere offers his own theory of explanation: I mean his account of scientific explanation, framed in terms of the syllogism, in the *Posterior Analytics*. On this account, scientific explanation requires the provision of a definition that links the explanandum to pre-eminently knowable first principles via a proper deductive framework.[5] One would suppose that this theory of scientific explanation is

questions see Gail Fine, 'Forms as Causes: Plato and Aristotle', in Andreas Graeser (ed.), *Mathematics and Metaphysics in Aristotle* (Bern, 1986). For a dissenting, or at least more tentative, opinion about the merits of the explanation approach see David Charles, *Aristotle's Philosophy of Action* (London and Ithaca, 1984), esp. p. 47.

[4] This motivation seems fairly close to the surface in Fine's article.

[5] For an account of the relevant sort of explanation as co-ordinate with 'understanding', see Myles Burnyeat, 'Aristotle on Understanding Knowledge', in E. Berti (ed.), *Aristotle on Science: The Posterior Analytics*, Proceedings of the Eighth Symposium Aristotelicum (Padua, 1981), 97–139. Also see Barnes, *Posterior Analytics*, pp. ix–xvii;

distinct from the *Physics*' account of causes or of 'causal explanations', though of course somehow related to it. Advocates of the explanation approach ought to supply a clear account of this relationship.

The third, most pressing problem for the explanation approach is that explanation is not itself a transparently clear notion. There is a great deal of controversy about the nature of explanation within discussions of contemporary metaphysics and philosophy of science. Settling questions about whether Aristotle is a causal realist will not suffice to settle questions about whether he is an explanatory realist: these positions map on to one another in complex ways.[6] Scholars discussing Aristotelian explanation have concentrated on the formal model of scientific explanation in the *Posterior Analytics*, but this could well differ from his theory of the four causes in the *Physics*. Perhaps in the latter theory, though not in the former, Aristotle emphasizes that explanation is a 'pragmatic rather than syntactic or semantic concept.'[7] Explanations are relative to inquirers, situations, contexts, presuppositions, and interests; van Fraassen has argued that precisely these sorts of factors are the basis for Aristotle's distinguishing the four types of causes in the *Physics*. That is, he sorted out the four kinds of αἰτίαι he did as the result of classifying different sorts of why-questions that arise in different contexts for inquirers with distinct sorts of interests.[8] This pragmatic variation on the more standard-explanation approach raises further questions: To what extent *does* Aristotle recognize the role of pragmatic factors in explanation? Is this a fruitful line to follow in describing differences between the *Physics* and the *Posterior Analytics*?

In what follows I will discuss all these problems more fully. In section I, which provides a reading of Aristotle's account of causation in *Phys.* II, I will defend the claim that Aristotle is a causal realist of an

and Barnes, 'Aristotle's Theory of Demonstration', in Jonathan Barnes, Malcolm Schofield, and Richard Sorabji (eds.) *Articles on Aristotle*, I, *Science*, (London, 1975), 65–87).

[6] As explained by Jaegwon Kim in 'Explanatory Realism, Causal Realism, and Explanatory Exclusion', in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (eds.), *Realism and Antirealism*, (Midwest Studies in Philosophy 12) (Minneapolis, 1988), 225–39, esp. 231–2). (I am grateful to Christopher Shields for drawing my attention to Kim's article and its relevance for my project in this essay.)

[7] Michael Scriven, 'Causation as Explanation', *Noûs* 9 (1975), 4.

[8] Van Fraassen, 'A Re-examination', 32–4. This is in contrast to a version of the explanation approach proposed by other scholars who argue instead that Aristotle distinguishes the four causes on the basis of a kind of linguistic analysis of different types of why-questions. Such a view is defended by Dan Graham in *Aristotle's Two Systems* (Oxford, 1987).

unusual sort who believed that there are four types of objective causal relations in the world. Next, in section II, I will discuss Aristotle's views on the nature of explanation; here I will argue that he is an explanatory realist. I will consider differences between his accounts of explanatory knowledge in the *Posterior Analytics* and the *Physics*, paying particular attention to the pragmatist version of the standard-explanation approach. Though in the end I deny that the pragmatist approach affords a satisfactory interpretation of Aristotle's doctrine of the four causes, I show that it improves upon the more standard-explanation account by prompting us to recognize and reflect upon differences between Aristotle's discussions of *αἰτίαι* in the *Physics* and in the *Posterior Analytics*.

In the final section of this essay I will turn to an odd but intriguing aspect of Aristotle's picture of causes, namely, his thesis that there are, in addition to real or essential causes of various types, also accidental causes. I will explore the particular use Aristotle makes of the notion of accidental causation in his treatment of chance. Because such pragmatic factors as contexts, presuppositions, and interests seem particularly prominent in Aristotle's discussions of this sort of causation, my claims about Aristotle's causal and explanatory realism might well seem to break down for chance and other accidental causes. Still, I want to show that Aristotle's causal and explanatory realism are at play here also.

I Aristotle's Causal Realism in *Physics* II

If causation is 'really' explanation, according to Aristotle, then presumably causal statements involve certain descriptions of events or states of affairs. Although there may be some temptation on the 'explanation' interpretation to regard explanations as linguistic entities, and so to ascribe to Aristotle a Davidsonian-style concern with the logical form of causal statements,[9] most of the interpreters mentioned earlier have resisted this temptation, recognizing that Aristotle's attention is on things in the world (or that he writes in the material not the formal mode). So they call his theory a classification of 'explanatory factors', or, in Sorabji's case, of 'what provides an explanation'.[10]

[9] See Donald Davidson, 'Causal Relations', in *Essays on Actions and Events* (Oxford, 1980), opening sentence.

[10] Sorabji, *Necessity, Cause and Blame*, 40.

Gail Fine has offered an especially clear and detailed interpretation along these lines. She attempts to regiment Aristotle's theory of the four causes by suggesting that each type of cause could figure as an explanatory factor within an event cited as a cause:

[E]ven if we restrict causes to events productive of change, still, each of the four *aitiai* can be—if not a cause, at least—causally relevant, by being a factor one cites in explaining change; they will still be referred to in, be parts of, causal explanations.[11]

Several questions arise for any proposal which interprets Aristotelian αἰτίαι as explanatory factors. One concerns whether causal contexts, understood thus in terms of explanatory factors, are intensional. Notoriously, both events and states of affairs may be described in varying ways, some more or less explanatory than others. Did Aristotle regard causal contexts as intensional—i.e. is something an explanatory factor only under a certain description? Moravcsik, after some discussion, seems to answer this question 'yes'; Charles, similarly after discussion, 'no'.[12] Annas writes that Aristotle 'does not envisage that a statement might truly pick out a causal relation, but in a quite unexplanatory way'.[13] Fine comments in a criticism of Annas, 'I take it . . . that explanation, for Aristotle, can be construed extensionally'.[14]

But in my view, both the question just raised and these varying responses to it assume too much about Aristotle's theory of reference—and seeing this will show up a problem about the explanatory-factor interpretation. Specifically, what is assumed is that Aristotle could have formulated and considered the question of whether co-referential expressions are intersubstitutable in causal contexts. Such a question presupposes that Aristotle roughly shares our views on identity and reference. But a number of articles published in the last fifteen years have challenged the assumption that Aristotle grasped the modern (post-Leibnizian) concept of identity.[15] Furthermore, it has been argued that Aristotle would not have regarded such phrases

[11] Fine, 'Forms as Causes', 72.

[12] Moravcsik, 'Aristotle on Adequate Explanations', 12–13; Charles, *Philosophy of Action*, 47.

[13] Annas, 'In efficient Causes', 320.

[14] Fine, 'Forms as Causes', 71, n. 7.

[15] For the discussion that first raised doubts, see Nicholas White, 'Aristotle on Sameness and Oneness', *Philosophical Review* 80 (1971), 177–97. See also F. J. Pelletier, 'Sameness and Referential Opacity in Aristotle', *Noûs* 13 (1979), 283–311. Both White and Pelletier think that Aristotle had, but abandoned or lost his grip on, the concept of identity as we know it. For related discussions see also Alan Code, 'Aristotle's Response

as 'Polyclitus' and 'the sculptor' as co-referential, because his ontology includes, in addition to substances and accidents, certain entities (dubbed 'kooky objects' by Gareth Matthews) which are conjunctions of the two.[16] These are accidental unities, such as Socrates seated or the pale man. It turns out, on this interpretation, that Aristotle holds that the entity identified as an intrinsic cause of a statue—a sculptor—would be a kooky object dependent upon, but not identical with, the true substance—a man in this case. Thus the terms 'Polyclitus', 'the man', and 'the sculptor' are not, strictly speaking, co-referential. Still, Aristotle could well say that to call Polyclitus the cause of a statue is (*pace* Annas) to pick out the cause in an unexplanatory way, for it is to pick out something bearing a certain relationship—namely, accidental unity—with the cause of the statue, properly speaking. But because of his distinct views about identity and the ontology of accidents, this is different from saying that Aristotle envisaged either causal or explanatory contexts as intensional.

I have raised doubts about the appropriateness of directly comparing Aristotle's views of identity and reference to contemporary ones. Further doubts can be raised about whether this is a well-grounded approach to Aristotle's picture of the nature and ontology of causal relations. The explanatory-factor interpretation maps Aristotle's ontology on to a contemporary view in which causal relations obtain between events.[17] But Aristotle does not seem especially concerned to specify a uniform category of entities as causes. Sometimes he speaks of a property, like the art of building, as a cause; at other times he mentions a substance, like a man. Another difficulty is that Aristotle says that all the four causes can be either actual or potential. This means that he is perfectly willing to describe even the moving cause—the one that seems most similar to 'our' notion of causes as events—as a substance (i.e. when it is potential, rather than actual).[18]

to Quine's Objections to Modal Logic', *Journal of Philosophical Logic* 5 (1976), 159–86, and Fred Miller, Jr., 'Did Aristotle Have the Concept of Identity?' *Philosophical Review* 82 (1973), 483–90.

[16] Frank Lewis, 'Accidental Sameness in Aristotle', *Philosophical Studies* 42 (1982), 1–36; Gareth Matthews, 'Accidental Unities', in Malcolm Schofield and Martha Nussbaum (eds.), *Language and Logos* (Cambridge, 1982), 223–40. I should stress that, while their topics and views are related, these two authors reached their conclusions independently.

[17] This is especially true in Fine, 'Forms as Causes', 71–2; she suggests that this is more plausible for Aristotelian actual, rather than potential, causes.

[18] On this point, see Fine, 'Forms as Causes', 71.

Let us look more closely at Aristotle's discussion to pursue this question about the ontology of causal relations. It will come as no surprise to the reader familiar with Aristotle that he writes, in his treatment of causation in *Phys.* II. 3, that cause is said in many ways: 'Since many different things are called causes, it follows that many different things can all be causes . . . of the same thing' (195^a3–5). Like a number of parallel claims in the corpus—about being, goodness, health, etc.—this statement need not be interpreted as a claim about the meaning of the word 'cause' (αἰτία or αἴτιον), but can be read more straightforwardly as an assertion about how a variety of things are said to *be* causes.[19] Aristotle lists these things at 195^b12–16:

All this comes to six things, which may each be spoken of in two ways. There is the particular and the general, the accidental and the genus to which the accidental belongs, and these may be given singly or in combination. (Charlton's translation, altered)

Now presumably the first-mentioned 'particular and general' causes are meant to be proper, as opposed to accidental, causes. Aristotle does not himself supply a designation here for this contrasting concept; I shall proceed to describe such causes as 'intrinsic'.[20] We may review examples of what he has in mind, as follows:

Intrinsic causes of a statue, taken alone:

particular: this sculptor
generic: an artist

Accidental causes of a statue, taken alone:

particular: a man (Polyclitus)
generic: an animal

Combined causes:

particular: Polyclitus the sculptor
generic: an artistic animal

[19] I have in mind G. E. L. Owen's influential discussion of Aristotle on 'focal meaning'; his view is criticized by T. H. Irwin in 'Aristotle's Concept of Signification', in Schofield and Nussbaum (eds.), *Language and Logos*.

[20] I borrow this terminology from Susan Sauvé, 'Self-Movement, Character, and Moral Responsibility in Aristotle's Ethics', unpublished paper, 1988; 'intrinsic' seems preferable to 'essential' here, since the latter has misleading connotations (that there's just one cause of something; that being a cause is essential to something). Annas does use the term 'essential' here; see Annas, 'Inefficient Causes', 322, and also her discussion of 195^b13, 322 n. 33. Charles also uses this way of speaking, *Philosophy of Action*, 46 ff.

Aristotle's notion of an accidental cause in this chapter appears to be grounded in two other notions: (1) some concept of essential or intrinsic causes, and (2) the metaphysical relation of accidental unity. Properly speaking, he tells us, a sculptor is the cause of a statue. Suppose that this sculptor is pale, musical, or named 'Polyclitus'; in Aristotle's metaphysics, the sculptor, pale man, and Polyclitus are related in a relation called 'accidental unity'. This permits us to describe the pale man, and so on, as causes accidentally (κατὰ συμβεβηκός) of the statue.[21] He tells us at 195^b6–10 that there may be accidental effects as well. There is an illustration in *Met. E* 2:

> Again a confectioner, aiming at giving pleasure, may make something wholesome, but not in virtue of the confectioner's art; and therefore we say 'it was an accident', and while there is a sense in which he makes it, in the unqualified sense he does not. (1027^a3–5)

Suppose that the tasty treat made by the confectioner's art happens to be something wholesome, say, an oat-bran muffin with raisins. What has the confectioner's art actually made?—something tasty and tempting. It just so happens that this is (1) a bran muffin and (2) something good for you. The relation of accidental causation devolves upon first, an intrinsic causal tie between the confectioner and the tasty treat, and second, the accidental unity on the effect side, between the tasty treat and the wholesome bran muffin.

Aristotle's account of intrinsic or essential causal relations does not require that either the cause or the effect side of a relationship be described by referring to what we would call the essence of the thing itself. To be a sculptor is not the essence of the man named 'Polyclitus'; nevertheless, it is as a sculptor that this man is in an intrinsic causal relation to the statue he makes. Neither is being a tasty treat the essence of an oat-bran muffin. Further, what Aristotle's metaphysics counts as an accidental conjunction may not be an 'accident' in our ordinary sense of the term; there may be a perfectly good explanation for Polyclitus' being a sculptor, as for the tasty treat's being a bran muffin. The relation between the accidental and the random or unpredictable will occupy me in section III below.

The key feature of the accidental is that it is not regular or predictable. Aristotle characterizes it in *Met. E*, 2 as 'what is neither always

[21] In what follows I will interchange various ways of describing this relationship, assuming they can be unproblematically transformed: '*x* is the accidental effect of *y*'; '*y* causes *x* accidentally'; and '*x* and *y* stand in a relation of accidental causation'.

nor for the most part' (1026^b31–3). This remark provides a valuable clue about the nature of intrinsic or καθ' αὑτό causal relations; they *do* obtain either always or for the most part. Such regular and orderly connections, even merely usual, not exceptionless ones, depend upon the natures of the connected substances. For example, trees are said 'usually' to shed their leaves if their sap coagulates at the leaf-stems (*An. Post.* II. 16). Aristotle seems to believe that even exceptions themselves must admit of some explanation.[22] We would say that such καθ' αὑτό relations, by contrast to accidental relations, are natural and necessary relations. They hold in the real world, and statements about them are entailments—they support counterfactuals. These relations in the world obtain between universals of some sort, probably properties or dispositions, rather than between individuals (although of course within Aristotle's ontology such properties do not exist apart from individuals).[23] That Aristotle himself seems to think along these lines is suggested by the fact that when he does identify an individual (e.g. Polyclitus) as the intrinsic cause of something, it is *as* the possessor of a certain sort of power or skill. He writes,

> As elsewhere, so here, we should always look for the first [ἀκρότατον] cause of each thing. Thus a man builds because he is a builder, and a builder builds in accordance with the art of building; the art of building is, then, the prior cause, and similarly in all cases. (195^b21–5; Charlton's translation, altered)[24]

This advice appears to direct us to look at objective relations among properties in the world in identifying the most precise cause of something. It would apply concerning various of the four causes, for example, my frying pan conducts heat because it's made of iron, and iron conducts heat.

We can attempt to find some clarification of Aristotle's views by studying what he says elsewhere about καθ' αὑτό connections. However, this leads into a significant circularity. In his explication of the notion of καθ' αὑτό connections in the *Posterior Analytics*, Aristotle

[22] This is what I take to be the point of his example at 1027^a21–6; but the example is obscure, and so I may not be right. There are important difficulties about the relevant nomic connections that obtain within the Aristotelian universe in cases where a causal relation obtains usually ('for the most part') but not necessarily; I cannot pursue these here.

[23] For arguments in support of this presumption, see D. M. Armstrong, *What Is a Law of Nature?* (Cambridge, 1983), II.

[24] Presumably he means here that the builder builds in virtue of his possession of the art of building; he possesses an instance of this property. The universal art of building builds nothing, nor does it exist independently, according to Aristotle.

actually refers to causal relations. Thus, he says, while on the one hand a triangle is καθ' αὑτό a plane figure, and white is καθ' αὑτό a colour,

Again, in another way what belongs to something *because of itself* belongs to it in itself [καθ' αὑτό]; and what does not [belong] because of itself is accidental—e.g. if it lightened when he was walking, that was accidental; for it was not because of his walking that it lightened, but that, we say, was accidental. But if because of itself, then in itself—e.g. if something dies while being sacrificed, it died in the sacrifice since it died because of being sacrificed, and it was not accidental that it died while being sacrificed. (73^b11–16; Barnes's translation, altered; emphasis mine)[25]

That is, in the *Posterior Analytics*, Aristotle explicates the καθ' αὑτό relation in terms of causation, whereas in the *Physics* he does the reverse, explicating causation in terms of the καθ' αὑτό relation. I conclude that these two notions must be jointly illuminated.

The καθ' αὑτό relations Aristotle points to in analysing intrinsic causation must be various types of objective nomic connections. To see what these are like it is useful to consider more examples. Recall that accidentalness may apply to any of the four causes; for instance, if the iron in my frying pan is from West Virginia, then it is accidentally the case that my frying pan conducts heat *because* it's made of material from West Virginia. Accidental causation involving the final cause strikes us as more strange, but it clearly is possible within Aristotle's framework.[26] For example, Aristotle thinks that camels have four stomachs *for the sake of* digesting thorny food.[27] If I have a greenhouse with a desert room including a variety of thorny plants, then I might say that camels have four stomachs 'accidentally' *in order to* digest the sort of plants I have in my greenhouse.

The natural entailments or καθ' αὑτό relations missing in these cases of accidental causation are present in cases of intrinsic causation. There are natural and necessary connections between being a sculptor and producing statues, being made of iron and conducting heat, being round and being able to roll, or having four stomachs and digesting

[25] A substantive issue can be raised concerning the translation of σφαττόμεν in this passage: is Aristotle describing an intrinsic connection between being sacrificed and dying that should be understood as definitional, or is it (what we would call) a causal link between having one's throat cut and dying? I interpret it as the latter.

[26] Some of the most interesting examples involve human actions in which the agent is said not to know the act's real purpose. In these cases, what the agent knowingly does is not the same as what actually—and accidentally—happens. See my discussion in 'Aristotelian Actions', *Noûs* 19 (1985), 400–1.

[27] See *PA* III. 14.

rough food. In each case, of course, Aristotle's own analysis of the relevant καθ' αὑτό connection in the world is in terms of natures or powers. Iron has a nature comprised of some particular combination of the four fundamental powers, the hot, cold, moist, and dry. Hence, whatever is made of material having this particular nature is such that it will necessarily behave in certain ways given certain circumstances: it will conduct heat, it will sink in water, it will be attracted to magnets, etc. Similarly, even though it is not in the nature of a person to be a sculptor, sculpting is a sort of quasi-nature in virtue of which a person has the power to make statues given appropriate conditions, including desires and aims. Aristotle even speaks in this way of doctors or builders acting 'by nature' (πέφυκε), in *Met. E.* 2 (1027^{a}1). And again, it is in the nature of a stomach that it serves the purpose of digesting food, as it is in the nature of a horn that it is a means of self-defence.[28]

This is not the place to launch a full-scale investigation of Aristotle's concepts of powers and natures.[29] What is important to keep in mind is that these concepts play a role in his thinking about causation by grounding objective entailment relations of quite distinct sorts—not just ones that we think of as causal (i.e. efficient-causal) relations. He describes these grounding relations as various types of καθ' αὑτό connections. We on the other hand might express this by saying that in Aristotle's view the world contains four broadly different types of causal laws. Some natural laws concern how things with certain shapes will behave in particular circumstances (circular wounds will heal more slowly).[30] Other natural laws concern how things with certain material compositions and constituents will behave. Perhaps most strangely to our minds, Aristotle also must believe that there are natural laws articulating the necessary links between specific features of substances and certain purposes. For example, camels have four stomachs for the sake of concoction or good digestion. This is an instance of a more general law: natural creatures with digestion as an end will possess some means to this end. An even more general connection underlies this one: nature dispenses to living beings the means to achieve their ends.

[28] Aristotle discusses horns at some length in *PA* III. 2. Some animals have appendages resembling horns, but these are not horns because they do not serve the function of self-defence.

[29] I have discussed these notions in 'Aristotle on Possibilities and Capacities', *Ancient Philosophy* 6 (1986), 69–89.

[30] This is an example I borrow from van Fraassen; see 'A Re-examination', 40.

I believe that the evidence I have just been considering is enough to show that Aristotle is a causal realist, that is, he holds that causality is an objective feature of the world. Actually, I doubt that defenders of the explanation approach ever meant to deny this point—that they ever meant to argue that, understood in terms of explanations, Aristotelian *αἰτίαι* are merely subjective factors of our experience. Where my interpretation differs significantly from what has become the standard view is in my insistence that Aristotle is a causal realist about four distinct kinds of causal relations. This is an outcome that is fudged or resisted by adherents of the explanation approach, even those who speak of explanatory factors as real features of the world.[31] Moravscik, for example, after explicitly recognizing that there are (at least) four distinct types of explanatory factors, further inquires about what, in the Aristotelian ontology, grounds this belief. His answer is that this typology reflects aspects of Aristotelian substances.[32] This does not seem helpful, for it simply displaces the problem on to substances. My own view is that it is not especially illuminating to refer to explanatory factors, even if they are aspects of substances; such factors are explanatory because they are involved in explanatory relations that are themselves real. As I will argue in more detail immediately below, Aristotle is a realist about explanations as well as about causes. That is, differences in types of explanation are objectively grounded in different types of relation in the world. Specifically, I take it that they are grounded in the four kinds of causal relations identified in *Phys.* II. I have further suggested that these determinate objective relations are distinct types of *καθ' αὑτό* relations between properties.

II Explanatory Realism and Explanatory Knowledge

I have just tried to show that Aristotle believes there are different kinds of intrinsic causation precisely because there are different types of *καθ' αὑτό* connections in the real world. My interpretation imputes

[31] Fine writes that '[Aristotle's] explanatory factors are not (usually) sentences describing things, but real features of the world', 'Forms as Causes', 71.

[32] Moravcsik, 'Aristotle on Adequate Explanations', 5. This position is developed in complex ways in Moravcsik's 'What Makes Reality Intelligible?', this volume; more specifically Moravcsik now talks about the role played in causation by attributes, powers, and activities.

causal realism to Aristotle in the specific sense of saying that there are in reality four distinct kinds of causal ties. It is a short step to showing that Aristotle is also an explanatory realist. Explanatory realism is a position that holds that

> *C* is an explanans for *E in virtue of the fact* that *C* bears to *E* some determinate objective relation *R*. Let us call *R*, whatever it is, an 'explanatory relation'.[33]

As Jaegwon Kim has pointed out, 'the obvious first thought' about the identity of the relevant objective relation is that it is the causal relation.[34] Since Aristotle holds that there are four varieties of determinate objective causal relations, it is natural to interpret him as an explanatory realist who also holds that explanation statements refer to these objective relations. What is significant is that explanations cite real features of the world rather than psychological or doxastic states. The four kinds of καθ' αὑτό relations in virtue of which causal explanations are true are themselves 'out there' in the world. I do not think that proponents of the explanation approach (whether in the explanatory factors or the pragmatist version) have explicitly ascribed this form of explanatory realism to Aristotle; and, given my account of their negative motivation, this would be unlikely, for they are intent instead upon showing that the four kinds of explanation ultimately all refer to diverse aspects of events related by one real ('determinate objective') relation in the world, that is, efficient causation, understood however 'we' understand causation now.[35]

Two major problems arise for my interpretation of Aristotle as an explanatory realist. First (as I noted in my introduction) it will be important to offer some view of how the theory of causal explanation in *Phys.* II fits with Aristotle's theory of scientific explanation in the *Posterior Analytics*. And second, it is necessary to get clearer about Aristotle's views on the nature of explanatory knowledge (what form

[33] Kim, 'Explanatory Realism', 226 (author's emphasis). Kim thinks it likely that there are different kinds of explanation, and hence that an explanatory realist is committed to maintaining that there are different kinds of objective correlates to the explanation relations. (He mentions 'the relation of supervenience' and 'the micro-reductive relation' as examples; 238 n. 2) I think Kim, like most of us today, would find it extremely odd to hold that there could be different objective correlates to explanation relations that all count as causal relations.

[34] Kim, 'Explanatory Realism' 226.

[35] Moravcsik's version of the explanation view is different, and closer to my own (he speaks of four 'different entitative explanatory chains'; see 'What Makes Reality Intelligible?', this volume, p. 43).

does it take? do pragmatic factors enter in? etc.). The pragmatist version of the explanation approach offers attractive responses to both questions. As van Fraassen in particular has argued, Aristotle's reasons for distinguishing the four causes may reflect the significance of human interests: the distinction among four kinds of causal explanation in the *Physics* is 'pragmatic' in the sense that diversity among these sorts of causal explanation reflects facts about human contexts and inquiries. On this version of the explanation approach, Aristotle could be described as an explanatory realist in the *Posterior Analytics* without the commitment I ascribe to him—to saying that there are four various kinds of objective explanatory relations in the world. Instead, diverse explanations which arise from the human point of view in various situations all ultimately may be mapped on to one sort of objective correlate in the real world. Thus the pragmatic line seems to have some prima-facie plausibility. I want to examine it further as I seek to articulate Aristotle's views on the nature of explanatory knowledge. I think it will become clear that the pragmatic approach presupposes an account of scientific explanation that is significantly different from Aristotle's.

It is helpful to begin by locating the pragmatic approach within a framework of competing contemporary approaches to the nature of scientific explanation:

> Philosophical theories of scientific explanation can be roughly divided into three types: purely logical theories, which analyze explanation solely in terms of logical relations and truth conditions; theories with extra objective structure, which impose objective conditions on explanation beyond those of truth and logical structure; and theories with extra subjective structure, which impose on explanations psychological conditions of belief, interest, and so forth.[36]

The pragmatic approach belongs to the third, or subjectivist, category. Leading scholarly accounts of Aristotle's notion of scientific explanation have tended to interpret it as representative instead of the second, 'objectivist' group.[37] However, these accounts have focused primarily on the *Posterior Analytics* rather than the *Physics* or the

[36] Clark Glymour, 'Explanation and Realism', in Paul M. Churchland and Clifford A. Hooker (eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen* (Chicago, 1985), 104.

[37] Burnyeat's interpretation of Aristotelian explanation was explicitly based on a view articulated in Michael Friedman's article 'Explanation and Scientific Understanding', *Journal of Philosophy* 71 (1974), 5–19. Friedman is cited as a representative of Glymour's second category.

scientific treatises. Notoriously, Aristotle's scientific treatises do not exhibit the syllogisms we would expect of ideal sciences. Instead, in these treatises causal explanations are typically offered and clearly marked according to category or type. It might well be true, then, that in scientific practice, as in *Phys.* II, where Aristotle emphasizes the plurality of causes, he is utilizing a notion of explanation distinct from that set forth in the *Posterior Analytics*.

Aristotle's formal model in the *Posterior Analytics* provides a unified account of scientific explanation within which all four types of causal relation are to be represented by one specific deductive format. There is some debate about whether his formal machinery actually permits such reductive simplification; but for my purposes here what is important is that this was indeed his intention.[38] This formal model of explanation permits us to gloss over differences in types of causal or καθ' αὑτό connections. That is, though syllogisms are the proper means to represent any of the four different types of causal explanation, all types will be cited as middle terms in syllogisms having identical formal features. It will not be obvious on the face of a syllogism, so to speak, which kind of causation it describes. If scientific syllogisms differ at all, it will be in the grounds for asserting that the connections involved in each premiss are indeed necessary, i.e. in an epistemic (or perhaps pragmatic), not a syntactic feature.

What are the implications, then, for Aristotle's theory of explanation? On the one hand, it seems right to assert that the scientist who possesses a syllogistic demonstration of some proposition of the form 'All *C* is *A*' has in hand an explanation of *C*'s being *A*—because *B* is *A* and *C* is *B*. The scientist knows these propositions, and knows in addition that they are appropriately linked in an organized body of knowledge to evident first principles. We need not refer to any extra propositions known, in addition to these, to explain what is meant in claiming that the scientist has an explanation of why *C* is *A*. Yet on the other hand, this account seems to leave out something very important: what *kind* of explanation does one have in hand, in having a scientific syllogism? That is, the scientist will know that *C* is *A because* of some middle term *B*; but the syllogism alone will not reveal what sort of 'because' this is.

The pragmatist line proposes an account of the difference between science in its idealized form and science put to use in explanation by

[38] Barnes points out his difficulties in trying to accomplish this end; see Barnes, *Posterior Analytics*, 214–23.

stressing a sharp distinction between the descriptive and explanatory roles played by scientific statements. Whether a given statement is part of science or explanatory depends not upon its form or content but upon its context. A statement belonging to science proper can be cited in a certain context where it suffices in order to give an explanation. Van Fraassen gives the following example:

> [I]f you ask a scientist to explain something to you, the information he gives you is not different in kind (and does not sound or look different) from the information he gives you when you ask for a description. Similarly in 'ordinary' explanations: the information I adduce to explain the rise in oil prices, is information I would have given you to a battery of requests for description of oil supplies, oil producers, and oil consumption. To call an explanation scientific, is to say nothing about its form or the sort of information adduced, but only that the explanation draws on science to get this information (at least to some extent) and, more importantly, that the criteria of evaluation of how good an explanation it is, are being applied using a scientific theory.[39]

Explanation is always extra-scientific, in the sense that it involves the application of a part of science to answering specific questions that have arisen in a particular context outside of science.[40]

Van Fraassen's pragmatic line offers a plausible suggestion about why Aristotle distinguished the four causes, namely, at least in part to provide a solution to certain notorious problems about the asymmetries of explanation.[41] A statement, for example, that 'walking after dinner contributes to non-regurgitation of food' might well be the middle term of two distinct scientific syllogisms. This proposition is a bit of scientific knowledge in either case, but it can only be cited as an explanation when the context clarifies the nature of the inquiry, and hence the nature of the causation revealed by this relevant middle term. In one context this statement could be cited as the explanation which answers a person's question about why Socrates has good digestion (because he takes a walk after dinner), whereas in another context it could instead be the answer to the question about why Socrates is taking a walk after dinner (because he wants good digestion). In the first context this premiss supplies the efficient cause, the second, the final cause—but in either case the proposition functions in the syllogism in the same way. Instances of each of the four types of

[39] van Fraasen, *The Scientific Image*, 155–6.

[40] In discussing this particular topic I am indebted to my colleague William Austin.

[41] van Fraassen, 'A Re-Examination', 28–31.

scientific causal explanations may be cited to respond to diverse questions having particular presuppositions, contexts, contrast-classes, etc.

On the pragmatist interpretation, Aristotle argues that explanations that answer the four types of causal why-questions involve the application of bits of science (perhaps in the form of scientific demonstration) to answer specific questions in particular contexts: 'To sum up, explanation is an application of scientific knowledge; and it is the "telling" application, for, Aristotle maintains, you do not have scientific knowledge of something unless you know the "why" of it'.[42]

But van Fraassen's way of distinguishing between the intra- and extra-scientific is plainly not Aristotle's. This is an extremely important point. I think it is clear that in his *Phys.* II discussion Aristotle aims to highlight certain differences among the four causes, while in the *Posterior Analytics* he attends to common features of their formal (i.e. syllogistic) representability. But even though in his formal model of science Aristotle does not develop adequate machinery to emphasize and clarify differences among the four kinds of causal explanation, in the *Posterior Analytics* he does clearly take it to be a part of science to reveal causes—and he must mean by this to reveal the right *sort* of cause. Otherwise one would not have a 'telling' explanation, to use van Fraassen's term, even when one could supply the relevant syllogism. Further, in his actual scientific practice Aristotle insists very clearly on separating and categorizing explanations citing various from among the four causes, as recommended in the *Physics*. This is especially evident in the biological works, where he emphasizes differences between purposive accounts and efficient or material causal explanations. There is no reason to believe that he regards recognition of such differentiations among types of causal statements as 'extra-scientific'.

Part of the difficulty here, I take it, involves the difference between the *Posterior Analytics* syllogistic model of science and the actual Aristotelian sciences we have. The *Posterior Analytics* envisages explanation as, in effect, identical with intra-scientific citation of scientific demonstrations. Explanation requires knowing that certain first principles are appropriately linked with syllogistic deductions. This account obscures differences among the four types of causal explanation, envisaging an ideal possibility of formulating accounts of all four types of causes by citing them indifferently as middle terms. To know which type of

[42] Ibid. 28.

explanation a given syllogism affords might then seem to be something 'extra', a subjective not an objective feature. But it is doubtful that Aristotle would see it this way. I suspect instead he imagines that when scientists grasp middle terms they can tell which sort of causes they thereby capture. So although van Fraassen's reading is helpful to some extent, in that it fastens on to significant differences between Aristotelian sciences and their formal model, it does not get things right about either. For Aristotle, our diverse interests are generated by the world itself, and diverse kinds of causal statements are about relations that obtain in the world itself. Neither the interests nor the relations can be explained as an artefact of a specifically or idiosyncratically human curiosity, as van Fraassen's pragmatic line would propose. Furthermore, even if these different interests are not readily visible in the actual structure of the body of fully formulated scientific knowledge itself (that is, in the set of syllogisms that would constitute a completed science, if there were such a thing), they must be understood to be built into its truth-conditions—and by this I intend something semantic, not pragmatic.

III Accidental Causes

So far I have been discussing Aristotle's views about causation and causal explanation involving genuine or intrinsic causes. I wish now to focus instead on accidental causes. I have argued that Aristotle is both a causal and an explanatory realist, but this interpretation might seem to be stretched beyond the limits of plausibility when we come to accidental causes. The range of the accidental is indefinite, disorderly, seemingly chaotic; could Aristotle nevertheless really have believed that accidental causal relations are an objective feature of the world?

The most important role played by accidental causes in Aristotle's philosophy is in grounding his metaphysical account of chance and related phenomena. In *Phys.* II Aristotle moves directly from discussing the four causes in chapter 3 to a discussion of spontaneity and chance, in chapter 4. This is not surprising, for he regards chance as itself a kind of cause. (You can, after all, explain something's happening as due to chance: 'Why did the stone urn fall off the roof just now? By chance or because someone pushed it?') More specifically, chance is a certain kind of accidental cause:

> Things do, in a way, occur by chance, for they occur accidentally and chance is an accidental cause. But strictly it is not the cause—without qualification—of anything (197^a12–14).

This last remark is misleading in two ways. That chance doesn't 'strictly speaking' cause anything doesn't mean that there aren't things that happen by chance:[43] a horse that runs away happens thereby to escape a fire; Oedipus, in trying to defend himself, kills his own father. Aristotle's remark also misleadingly suggests that chance events are uncaused. If chance is a special type of cause that accounts for events happening when there is otherwise no underlying (i.e. no essential) causation at all, then a defence of chance would be related to a defence of indeterminism. But I want to argue that chance does not involve causal indeterminism any more than does accidental causation generally. Recall, for example, that if a doctor builds a house, there is a cause for the house's being built, though the doctor builds it, as it were, by chance.

However, Aristotle himself seems to link chance to indeterminism in an argument in *Met. E*. 3, where he supplies an example involving chance to argue that it is not the case that all the consequences of our actions are determined in advance. His example involves a chain of events leading up to a man's accidental death: the man eats spicy food, gets thirsty, goes out to a well to drink, encounters a band of ruffians there, and is murdered. The man is murdered by chance; eating spicy food was the accidental cause of his being murdered. (There is no intrinsic causal link between eating spicy food and being murdered.) Aristotle appears to argue that chance events like this one involving accidental causes are undetermined: it is (fortunately) not a foregone conclusion that if one eats spicy food one will be murdered.

This example and its implications are important to consider, for they seem recalcitrant to analysis either in terms of the explanation approach to causation or on my own favoured approach, which insists on Aristotle's explanatory and causal realism. I will offer an interpretation of this passage; but first I want to examine an alternative interpretation of this text that proceeds on the assumption that for Aristotle causes are explanations.

[43] Strictly speaking, chance (τύχη) applies only to certain accidents in the realm of human endeavours, and the spontaneous (τὸ αὐτόματον) is a broader notion not confined to human agents. I use the former term more broadly here for expository convenience. See Lindsay Judson's paper in this volume for further clarification of these notions in relation to Aristotle's account of causes.

Richard Sorabji offers such an interpretation in *Necessity, Cause, and Blame*, reconstructing Aristotle's argument as follows: Aristotle reasoned that, since causation is explanation, then all events that have causes (αἰτίαι) can be explained. Since coincidences lack explanations, they also lack causes. Hence, Aristotle concluded, coincidences are not determined. Sorabji faults this argument, pointing out that if causation is really determination rather than explanation, to say that an event lacks a bona fide explanation does not suffice to show it is undetermined:

> ... [A] determinist could rightly protest that necessitating, i.e. sufficient, conditions are unlike explanations ... [If there are sufficient conditions for each arrival at the well] then it must be conceded that there are sufficient conditions for what is entailed by this, namely, their being at the well simultaneously.[44]

Clearly Sorabji's critique of this argument depends on at least two points: (1) that it is intended as an argument against causal determism, and (2) that it relies on an understanding of causes as explanations. I have argued earlier in this paper against (2), but I have also said things that render (1) problematic. Sorabji assumes that Aristotle is primarily worried about threats to human freedom posed by a determinism of mechanical (or efficient) causes. But if I am right that he is a causal realist about four distinct types of causal relation, then his worry might just as well be about a forward-looking intrinsic causality, or teleological determinism. Indeed, this makes more sense in relation to his example: he's saying it would be absurd to think one ate food or went to a well *in order to be* murdered.[45]

If chance is like other instances of accidental causation—there are often intrinsic causes for events Aristotle describes as due to chance—then it does not figure into the world by intervening in the causal order; chance is not a source of causal indeterminism.[46] But what is it

[44] Sorabji, *Necessity, Cause and Blame*, 13.

[45] My suggestion is in accord with a rather different account of this chapter that was proposed by Dorothea Frede, who also faults Sorabji for focusing on accidents' interruptions of chains of efficient or moving causes. On her view, Aristotle's worry is not about 'the hegemony of necessitating efficient causes but rather the totalitarian regime of the τέλος'. See Dorothea Frede, 'Aristotle on the Limits of Determinism: Accidental Causes in *Metaphysics E* 3', in Allan Gotthelf (ed.), *Aristotle on Nature and Living Things* (Pittsburgh and Bristol, 1985) 207–25.

[46] Cf. Frede's comment: 'It would be a strange oversight on Aristotle's side had it escaped his notice that, according to his own example in the *reductio*, it is not just nature and the natural, as one would expect, that is "saved" from strict universal determinism by the existence of accidental causes but the contingency of human actions too! How does

exactly? Chance outcomes are accidents or exceptions that don't exemplify regular types of final causation; they involve 'as-if' purposes, processes which reach a bad or good end without really aiming at it.[47] Why not follow the pragmatic line, saying that there are certain of these ends that we attend to because they appear especially interesting, poignant, fateful, etc.? Getting murdered when you go to a well is one such case; similarly, we pity poor Oedipus, who chanced to kill his father when he killed a man. This question is especially acute given my interpretation of Aristotle's explanatory realism: in saying something occurred by chance we cite an explanation that is somehow odd or funny, and so it would seem especially forced to insist that there is a corresponding objective, determinate relationship out there in the world.

At this point we need to think more generally about accidental causes. When I explicated the Aristotelian notion of accidental causation earlier, I said that such a relation always depends upon two things; a proper or intrinsic causal relation, and an accidental unity. Aristotle emphasizes the former feature in *Phys.* II. 6, when he writes that 'Since nothing which is accidental is prior to what is καθ' αὑτό, it is clear that no accidental cause can be prior to a cause καθ' αὑτό' (198ª6–8). In the *Metaphysics* Aristotle describes the case of a builder who accidentally produces health. Imagine this sort of case in more detail, modernized. At the scene of an emergency near a construction site, a man dressed in hard-hat and work clothes rushes out to perform cardio-pulmonary resuscitation (CPR) on a heart-attack victim. A reporter might cover the story, and headlines would read 'Construction worker saves life!' Now, the pragmatist emphasizes that this accidental relation in effect comes into existence when we pick it out. It simply seems noteworthy (has 'human interest') in a way others might not. The newspaper reporter regards the accident of a hard-hatted labourer's knowing CPR and being able to save a life as more newsworthy than numerous other accidental effects which he ignores—e.g. the heart-attack victim's squashed lunch, the labourer's mislaid hammer, etc.

Still, accidental causes are really 'out there' in the world, since

this square with his tenet that choice and the ensuing actions are up to us?' (Frede, 'Limits of Determinism', 207).

[47] '[I]f it were not for such interferences [viz.: by accidental causes] one could not maintain a distinction between the *intended* end and the end *actually* achieved'. Frede, 'Limits of Determinism', 221 (author's emphasis).

intrinsic causal relations and accidental unities in which they are grounded are objective features of reality. There must be some basis in reality for the assertion of accidental causation in a case like this: it was after all the construction worker and not someone else who saved the person's life. Presumably he did so because he happened to have learned CPR; it was in virtue of this that he saved a life, not because he was a construction worker. That is, there is an intrinsic causal relation here between an agent who knows a medical procedure and an act of healing, and there is also an accidental unity, between the construction worker and the possessor of medical knowledge.

My rejection of the pragmatist's interpretation of accidental causes has a parallel in contemporary disputes among philosophers of science. In discussing a variety of people's interests in and views about the causes of a carriage accident, N. R. Hanson commented:

> There are as many causes of *x* as there are explanations of *x*. Consider how the cause of death might have been set out by a physician as 'multiple haemorrhage', by the barrister as 'negligence on the part of the driver', by a carriage-builder as 'a defect in the brakeblock construction', by a civic planner as 'the presence of tall shrubbery at that turning'.[48]

In this example, extra-scientific interests determine the part of the causal network someone is concerned about. The civic planner worries about obstructions to views on the highway, whereas the manufacturer worries about brake failure. Not surprisingly, van Fraassen finds here grist for his pragmatic mill. The pragmatist maintains that explanations of the accident are relative to human presuppositions and inquiries.

On the pragmatist line, there are a variety of explanations of the carriage accident that do not necessarily refer to determinate relations in reality. And this might similarly seem to hold for Aristotle's treatment of accidental, as opposed to intrinsic, causation. But, as should be clear from the carriage accident example, the physician or the manufacturer *are* looking for real explanations in terms of genuine or intrinsic causes.[49] Although their interests in diverse aspects of the

[48] N. R. Hanson, quoted by van Fraassen in *The Scientific Image*, 125.

[49] In criticizing van Fraassen's account of this example, Wesley Salmon has argued as follows: 'We can see readily how contextual factors are involved in the correct identification of the why-question that is used to call forth a scientific explanation. It obviously does not follow ... that once we have ascertained by reference to pragmatic and contextual factors what explanation is being sought, the explanation itself must embody pragmatic features.' Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the*

accident may seem, like the newspaper reporter's, to reflect pragmatic considerations, the causal stories they point to are genuine.[50]

I want to say that when we focus on accidental effects and causes such as these, we are simply selecting from an indefinite number of accidental causes and effects the ones which we find especially striking or which answer to particular interests of ours. But the ones we do not pick out are equally objective. When confectioners make tasty treats, they also produce items that may be wholesome, an accident we notice because it is important to us. But they also produce things having a certain colour, weight, size, and shape, a particular price and weight, and a specific number of raisins. They make things that my father dislikes, though he knows they're good for him.

In the world at any time an indefinite number of such accidental causal relations really exist; perhaps this is just what Aristotle means when he speaks of accidental causes as 'indefinite' ($197^{a}9$). Hence there can be no Aristotelian science of either chance or accidental causes. Still, Aristotle recognizes the important role of the accidental in his account of causation, as in his metaphysics in general, and in so doing he allows for the more quirky aspects of human interests in the world. What is strangest about this doctrine, I believe, is that these are features of the world that are only noticed and significant for us humans, creatures who make our way about in the world by observing, classifying, and discoursing on accidental unities (kooky objects). Aristotle's God, of course, thinks only of itself; but even the lesser intellects of the star-souls would surely not conceive of the world or of its connective fabric via the accidental. They would have no use for a notion of accidental causation. In a way, humans, who pay more attention to chance and the accidental, know more about the universe than God.[51] But what we know is not fit to be deemed science; in other

World (Princeton, 1984), 130. (I am grateful to William Austin for drawing this criticism to my attention.)

[50] This has an important consequence for the practice of science. As scientists develop better accounts of the causes of a phenomenon, they acquire better explanations. For further discussion of the possibility of launching a scientific inquiry from a merely accidental grasp of the explanandum, see Robert Bolton, 'Essentialism and Semantic Theory in Aristotle: *Posterior Analytics* II. 7–10', *Philosophical Review* 85, 514–44. Problems parallel to those I have been discussing here about Aristotle's notion of explanation have been raised in modern discussions of evidential relations in science. See Helen Longino, 'Evidence and Hypothesis: An Analysis of Evidential Relations', *Philosophy of Science* 46 (1979), 35–56.

[51] Cf. van Fraassen: 'It is sometimes said that an Omniscient Being would have a complete explanation, whereas these contextual factors only bespeak our limitations due

words, it is not really knowledge. The interests of people, in so far as we strive to be truly rational animals, will presumably direct our inquiries to proper causal connections, supplanting the accidental ones which remain, after all, 'inscrutable to man' (197^a9–10).[52]

to which we can grasp only one part or aspect of the complete explanation at any given time. But this is a mistake. If the Omniscient Being has no specific interests . . . and does not abstract . . . then no why-questions ever arise for him in any way at all—and he does not have an explanation in the sense that we have explanations' (*The Scientific Image*, 130).

[52] This quotation is from a passage in which Aristotle is only discussing 'what men think', but he seems to agree.

4

Chance and 'Always or For the Most Part' in Aristotle

LINDSAY JUDSON

This paper has two aims. One is to examine Aristotle's analysis of chance in *Phys.* II. 4–6; the other is to try to elucidate a concept which has been very much neglected by commentators despite its central importance to a number of Aristotle's doctrines and theories—his concept of 'happening always or for the most part'. These two aims go hand in hand, because the analysis of chance invokes the concept of happening always or for the most part, and also provides important evidence as to what this concept is. I shall begin with the analysis of chance (sections I–II), then turn to 'happening always or for the most part' (sections III–VI), considering along the way what light the former sheds on the latter. In pursuit of these broad questions a number of points of detail raised by the analysis in II. 4–6 will have to be ignored.

I

'Chance' is intended as a translation of τὸ αὐτόματον, the main *analysandum* of the *Physics* discussion. It does not capture the idea of τὸ αὐτόματον perfectly, since for us 'chance' applies to a much wider range of cases: events could be ascribed to chance when they satisfy Aristotle's other conditions for being ἀπὸ τοῦ αὐτομάτου, but fail to be ἕνεκά του in the sense in which the products of Aristotelian chance must be—that is, when they do not benefit anyone (see section II below).[1] None the less 'chance' comes closer to Aristotle's concept than any other translation.[2] Although elsewhere Aristotle frequently

[1] Although he does not discuss the point, Aristotle of course has a classification at his disposal which would cover these cases, namely that of accidents (συμβεβηκότα).

[2] Recommending the translation 'spontaneous', Guthrie writes: '*Tyche* was the common word for chance, luck or fortune . . . *Automaton* had much in common with its

uses τύχη in the same sense as τὸ αὐτόματον, in *Phys.* II. 4–6 he generally gives it a much narrower sense, so that 'things due to τύχη' form a subset of 'things due to τὸ αὐτόματον'; I shall translate τύχη as 'luck'.[3]

Aristotle has a twofold purpose in introducing chance at this point in *Phys.* II. One purpose is forward-looking: Aristotle wants to go on to discuss the claims of other physicists that certain things are due to chance—the atomists' claim that the ordered cosmos is due to chance (II. 6, 198ᵃ5–13), and Empedocles' (supposed) claim that the arrangements of the parts of animals are due to chance (II. 8). Commentators tend to see this forward-looking purpose as paramount—and sometimes berate Aristotle for tailoring his analysis to suit this 'hidden agenda'.[4] But although the forward-looking aim is clearly important to Aristotle, these complaints are unwarranted. Aristotle's analysis is

modern derivative 'automatic', as used e.g. in Homer of the tripods invented by Hephaestus, which ran of themselves' (*A History of Greek Philosophy* vi (Cambridge, 1981), 236 n. 1). Now things which happen 'of themselves' do share a feature with τὸ αὐτόματον as Aristotle analyses it in the *Physics*: they imitate the workings of some natural or deliberative subject (see sections II and IV, below). But there the resemblance ends, for they do this in virtue of some mechanism (usually internal to their subject), which operates in a regular, repeatable, and predictable way—this is the case for Hephaestus' tripods and the αὐτόματα described by Aristotle at *MA* 7, 701ᵇ1–10. Events which happen ἀπὸ τοῦ αὐτομάτου, on the other hand, are on Aristotle's analysis anomalous, irregular, and opaque to science. So I take αὐτόματος to be ambiguous: it simply does not mean the same in *MA* 7 and in *Phys.* II. This point is worth stressing for another reason. It is a familiar complaint that Aristotle's account of spontaneous generation in the biological works is inconsistent with his account of spontaneity in the *Physics* (see D. M. Balme, 'Development of Biology in Aristotle and Theophrastus: Theory of Spontaneous Generation', *Phronesis* 7 (1962) 91–104; cf. Augustin Mansion, *Introduction à la physique aristotélicienne*, 2nd edn. (Louvain and Paris, 1946), 310 and n. 54). If αὐτόματος is ambiguous, as I claim, the problem of wholesale inconsistency simply does not arise; indeed, the arguments for the inconsistency are just further arguments for the view that αὐτόματος is ambiguous. (Note, however, that generation ἀπὸ τοῦ αὐτομάτου (καὶ ἀπο τύχης) in *Met.* Z (7, 1032ᵃ28–32 and 9, 1034ᵇ4–6) seems to be *chance* generation—exceptional cases in which an animal normally generated from seed is generated without seed—and not the regular spontaneous generation discussed in the biological works.)

3 I shall concentrate on Aristotle's analysis of τὸ αὐτόματον in general, and ignore the interesting question of his distinction between τὸ αὐτόματον and τύχη. Like τύχη, 'luck' has a narrower extension than 'chance', and is most appropriate when used in connection with a special class of subjects—persons—which is reasonably close to the class to which Aristotle's τύχη is confined (subjects capable of προαίρεσις). Within II. 4–6, Aristotle uses τὸ αὐτόματον in two ways: in a wider sense which includes luck, and in a narrow sense which covers chance things which are not due to luck. Throughout this paper I use 'chance' for τὸ αὐτόματον in the wide sense; in nn. 50 and 52 below I use 'chance$_N$' for the narrower sense.

4 See D. M. Balme, 'Greek Science and Mechanism I: Aristotle on Nature and Chance', *Classical Quarterly* 33 (1939), 129–38; Guthrie, *History of Greek Philosophy* vi. 241–2.

actually guided by his other purpose, and quite naturally arises from the discussion of the four causes in chapter 3: Aristotle wishes to explain how chance fits into his account of causation.

That this is his main concern in chapters 4-6 is evident from the *endoxa* he selects. He does mention the physicists' claim about the cosmos, the refutation of which forms the final section of the discussion; but most of the *endoxa* relate to the question of the *causal role* of chance:

> Both luck and chance are said to be among the causes, and many things are said both to be and to come to be because of luck and because of chance. So we must consider in what way luck and chance are among these causes, and whether luck and chance are the same thing or different, and the whole question as to what luck and chance are. (195^b31-6)

The earlier physicists, Aristotle says, adopt two very different attitudes towards chance, each of which is problematic. Some say that nothing happens by chance; but this is absurd, as many things clearly do. But those who agree that some things happen by chance give no account of what it is; it is clear that most of them do not regard it as a type of cause on a par with 'such things as love, strife, mind, fire, or anything else of that kind' (196^a18–19), but if it is not, how can things happen as a result of chance? We seem driven back to the first arm of the dilemma.

Aristotle does not attempt to detect truth in all the *endoxa* which he describes: as I have mentioned, he thinks that those who hold that the cosmos is due to chance are simply wrong. Nevertheless his main purpose is the characteristic one of seeking a middle way which will show how most of the *endoxa* are partially true. In effect, his aim is to do justice to the idea that chance is *something* and to the idea that it is *nothing*. He does this by denying that chance is a causal force in its own right (or a fifth type of cause co-ordinate with the other four), while maintaining that there is none the less a sense in which events do happen by chance. In the process he demonstrates what is right about other *endoxa*, especially the sense in which chance is obscure: on Aristotle's analysis, chance *is* obscure, but not in virtue of being a mysterious or supernatural force (196^b5–7, 197^a8–21).

II

Aristotle begins his positive account of luck and chance in chapter 5 by pointing to the existence of two overlapping classes. The first is the class of things which come to be neither always nor for the most part. I shall speak of members of this class as happening *rarely*;[5] and for convenience, I shall use the term 'event' to cover the whole range of things which might be said to come to be either rarely or regularly—events, processes, states of affairs, activities. It is the existence of rare events which demonstrates the existence of chance:

> First, then, since we see some things always coming to be in the same way, and some for the most part, it is evident that neither luck nor what is due to luck[6] is said to be the cause of either of these—either of what is of necessity and always, or of what is for the most part. But since there are also things which happen in addition to these, and all say that these happen from luck, it is evident that luck and chance are something. For we know that such things are from luck and that the things due to luck are of this kind. (196^b10–17)[7]

The second class is that of events which happen 'for the sake of something'. We might be forgiven for taking this to be the class of events which have final causes; but as we shall see in a moment, Aristotle has a much wider class in mind.

Aristotle is now ready to give what we may call his first definition of chance:

> Now things of this sort [i.e. events which happen for the sake of something], whenever they come to be incidentally, we say are from luck. (For just as in the case of being also, there is being *per se* and being incidentally, so it is possible for something to be a cause.) . . . As has been said, therefore, whenever this comes to be [i.e. when something comes to be incidentally] among the things which come to be for the sake of something, then it is said ⟨to be⟩ from chance or from luck. (196^b23–31)

This gives two conditions which an event must satisfy to be the product of luck or chance: (1) it is 'for the sake of something', and (2) it 'comes to be incidentally'.

[5] Strictly speaking this class includes things which happen exactly half the time as well as those which are less frequent, and I intend such events to be covered by my term 'rare'.

[6] Simplicius' remark on 196^b10–17 applies to most of ch. 5: 'It is clear that, having not yet distinguished luck and chance, his argument concerns both, even though he only mentions luck' (*in Phys.* 334.35–335.1).

[7] For rarity as a necessary condition for chance, see also (e.g.) *An. Post.* I. 30, 87^b19–27, *Cael.* I. 12, 283^a32–$^b1$, *GC* II. 6, 333^b4–7, *EE* VIII. 2, 1247^a31–3, *Rhet.* I. 10, 1369^a32–$^b5$.

The first condition raises an obvious difficulty, since Aristotle standardly *opposes* 'by chance' and 'for the sake of something'.[8] It would be possible to interpret ἐν τοῖς ἕνεκά του γιγνομένοις (196ᵇ29–30) merely as 'in the general class of things which are normally ⟨but need not be⟩ for the sake of something'; but this solution will not do. At 196ᵇ34–6 he clearly refers to a chance event as being for the sake of something: ἦλθε δ' οὐ τούτου ἕνεκα, ἀλλὰ συνέβη αὐτῷ ἐλθεῖν καὶ ποιῆσαι τοῦτο τὸ τοῦ κομίσασθαι ἕνεκα;[9] similarly at 197ᵃ1 he describes the benefit produced by the chance event as 'the τέλος'. Ἂν πραχθείη at 196ᵇ22 also suggests that here ἕνεκά του applies to more events than it does in the standard sense, while in one and the same sentence at 197ᵇ18–21, Aristotle says both that chance events are ἕνεκά του and that they do not come to be for the sake of what results. It is hard to resist the conclusion that Aristotle does wish to say that chance events are ἕνεκά του, and that he is using the phrase not in the familiar, explanatory sense, but in an unfamiliar, non-explanatory one.[10]

If this is so, what is the non-explanatory sense? Simplicius' interpretation is generally accepted: he takes '*E* is ἕνεκά του in the non-explanatory sense' to mean '*E* is such that it might have been ἕνεκά του in the explanatory sense' (*in Phys.* 335.33–336.5).[11] I prefer to take it to mean '*E* actually confers some benefit', i.e. '*E* is the *per se* cause of some benefit'.[12] There is little direct evidence bearing on the question,

[8] See e.g. *An. Post.* II. 11, 95ᵃ8–9 (ἀπὸ τύχης δ' οὐδὲν ἕνεκά του γίνεται), *Phys.* II. 8, *PA* I. 5, 645ᵃ24–5, *Rhet.* I. 10, 1369ᵃ32–ᵇ5.

[9] I follow William Charlton in reading τοῦτο τὸ τοῦ κομίσασθαι, since this makes the sentence rather less awkward by attaching the non-explanatory sense of ἕνεκά του to a description of the event, and not to the agent's performance (*Aristotle's Physics, Books I and II*, Oxford, 1970, 48–9). but the sentence would still associate chance with ἕνεκά του even if we retained the reading of the MSS. Deleting τοῦ κομίσασθαι ἕνεκα to avoid this (so Bonitz and the Oxford translators) is a counsel of desperation, as emendation would be needed in many other places in chs. 5–6 as well.

[10] This is convincingly argued by Jim Lennox in 'Aristotle on Chance', *Archiv für Geschichte der Philosophie* 66 (1984), at 52–6.

[11] W. D. Ross (*Aristotle's Physics*, (Oxford, 1936), 517–19), Charlton (*Physics I and II*, 106), and Lennox ('Aristotle on Chance', 56–60) all endorse some version of this view.

[12] There are three points to note here. (1) 'Benefit' is intended in the most general sense: *E* produces a benefit if it serves someone's interests, or satisfies a desire or need. (2) *E* will normally be an *efficient per se* cause of the benefit; but Cynthia Freeland has pointed out to me that there are cases in which it would be a formal cause. For simplicity's sake I ignore these cases in what follows. (3) Apart from the remarks at 197ᵃ25–30, Aristotle's discussion ignores the idea of *bad* luck entirely. Mansion claims that Aristotle's analysis of luck does not leave room for bad luck, and that this is a sign of a substantial change of doctrine which he thinks occurred between an earlier discussion of chance and a later revision (*Introduction à la physique*, 307–8; cf. Balme, 'Greek Science and Mechanism I', 129). The first of these claims is formally correct, but the second does

but my reason for preferring this interpretation to Simplicius' is that in the only other context in which Aristotle appears to use ἕνεκά του in a non-explanatory sense, he does mean 'actually produces an end' and not 'is suitable for producing an end'. This context is the list of things ignorance of which prevents an action from being ἑκούσιον—a list which includes ignorance of the ἕνεκά του.[13] Ignorance of what end the action is *suitable* for producing is irrelevant to whether or not it is ἑκούσιον; what is relevant is ignorance of what end it will actually produce, and the *EN* v. 8 and *EE* passages confirm that this is Aristotle's meaning:

> [Injuries] done in ignorance are *errors* when the person acted on, the act, the instrument or the ἕνεκά του is not what the agent understood; he thought . . . that he was not hitting . . . with this end, but ⟨the end⟩ which resulted was not what he thought the ἕνεκά του was—e.g. he threw not in order to wound but only to graze. (*EN* v. 8, 1135^{b}12–16)[14]

What does condition (2) in the definition of chance involve? Aristotle's remark about incidental causation at 196^{b}24–9 indicates that coming to be incidentally is a matter of being incidentally caused in some way, and that being incidentally caused is at least a necessary condition of coming to be incidentally. This idea is supported by the explanation Aristotle gives of the sense in which chance is ἀόριστον (see p. 79 below), and is confirmed both by the example of the lucky creditor at 196^{b}33 ff. and by the subsequent descriptions of luck and chance as incidental causes (197^{a}12–13 and 32–5; cf. 198^{a}5–10).

The concept of incidental causation raises many questions which I shall not attempt to pursue here.[15] Briefly, if we use 'E_i' as a place-

not follow. Clearly what Aristotle needs in order to be able to include bad luck is the idea of 'producing a benefit *or harm*'; and while the association of ἕνεκά του even in its non-explanatory sense with the notion of a τέλος certainly does not suggest that idea, it is easy enough to extend Aristotle's analysis in this way so as to cover bad luck as well as good. (Ross, *Aristotle's Physics*, 41, offers a different account of how bad luck fits into the Aristotelian analysis.)

[13] *EN* III. 1, 1111^{a}2–6 and 18–19; v. 8, 1135^{b}12–16; and *EE* II. 9, 1225^{b}1–5.

[14] For a discussion of these passages in slightly different terms, see Cynthia A. Freeland, 'Aristotelian Actions', *Noûs* 19 (1985), at 400–1. Note that producing an actual benefit seems to be a condition of being a lucky event in *our* sense: a man who stumbles upon treasure is neither lucky nor unlucky if he is indifferent to riches. But of course this is not true of our notion of chance, as I said at the beginning; and in any case it does not prove that it is Aristotle's condition also.

[15] For a discussion of some of the issues see section III of Freeland's paper 'Accidental Causes and Real Explanations' in this volume. Ascriptions of *per se* and incidental causation can be treated as highly intentional, so that E_1 may be a *per se* cause of E_2 under one set of descriptions of E_1 and E_2, but an incidental cause of E_2 under another; or as extensional but presupposing a very fine-grained ontology, so that these different descriptions

holder for events, '+' to indicate accidental concurrence in a subject, or 'accidental unity', and '⇒' to signify *per se* causation, then E_1 can be an incidental cause of E_3 in two different ways:

Schema I	Schema II
E_1	$E_1 \Rightarrow E_2$
$+$	$+$
$E_2 \Rightarrow E_3$	E_3

Aristotle's illustrations at 196ᵇ26–7 and 197ᵃ14–15 are cases of type I; in *Met. E.* 2 he gives examples of both types—a cook (aiming at giving pleasure) produces health in someone, a housebuilder (who happens to be a doctor) cures someone (1026ᵇ37–1027ᵃ5).

We can now see why chance is ἄδηλον and ἀόριστον (196ᵇ27–9 and 197ᵃ8–14): its causes are ἀόριστα (cf. 197ᵃ9–10). Take any given event E_1 in schema I; since an unlimited number of types of event can attach to the subject of this event, an unlimited number of types of event can stand in the position of E_2, and hence an unlimited number of types of event can stand in the position of E_3 (and a similar argument holds for E_3, E_2, and E_1 respectively in schema II). Thus there is no finite specification of the possible incidental causes which a given type of event may have, and there is equally no finite specification of the possible incidental 'effects' which a given type of event may have. For this reason any putative 'laws' which connect types of events under incidental causation would simply be constituted by infinitely long disjunctions, and these could not in Aristotle's view be objects of knowledge.[16]

The *per se*/incidental distinction applies to all four causes (*Phys.* II. 3, 195ᵃ3 ff.); does Aristotle have a particular mode of causation in mind when he speaks of chance events coming to be incidentally? It is sometimes held that Aristotle takes luck and chance to be incidental *final* causes, but this is not so. Aristotle himself raises the question at 195ᵇ33–4 and 196ᵇ8–9, and says that it is *efficient* causation which is

in fact pick out different events. (And, of course, there are intermediate positions too.) How fine-grained Aristotle's ontology of events, etc. actually is makes no difference to the substance of his account of chance events as incidentally caused. Simply for convenience of presentation, I shall use the language of a very fine-grained ontology (in which, for example, 'the man's going to the agora' and 'the man's going to where his debtor is' in the story of the lucky creditor pick out different events which unite accidentally in the man).

[16] Note that nothing in this account rules out our being able to know *in a particular case* that a concurrence of such-and-such a sort is likely, and hence being able to predict the chance event itself. See further pp. 91–2 below.

involved: 'of the modes of cause, each of these [luck and chance] is in the sources of change' (198ᵃ2–3). Since luck and chance are elsewhere described as incidental causes, as we noted above, it is clear that the mode of incidental causation in question is incidental efficient causation. This is not to say that one cannot identify incidental causes of the other modes in cases of chance: in the example of the lucky creditor, for instance, the *per se* final cause of the man's going to the agora—seeing a spectacle, or whatever it might be—will stand as an incidental final cause of his meeting his debtor. But Aristotle makes it plain that 'coming to be incidentally' in the analysis of chance involves—and only involves—incidental efficient causation.[17]

It is an interesting feature of the way we talk about luck and chance that events which lead to a chance result—and the benefit which this chance result produces—can also be ascribed, in a derivative way, to chance or luck, and can themselves be called chance events (cf. Ross, *Aristotle's Physics*, 516–17). For this reason, let us distinguish the events which satisfy Aristotle's definition—i.e. which are chance events in the non-derivative sense—by calling them 'chance outcomes' (or 'lucky outcomes'). As we have seen, it is a necessary condition of an event's being a chance outcome that it have an incidental efficient cause. Employing the two schemata for incidental causation which I gave above, we can derive two causal structures in which chance outcomes can be embedded; in both, the chance outcome has E_1 as an incidental cause:

Schema I*	Schema II*
E_1	$E_1 \Rightarrow E_2$
$+$	$+$
$E_2 \Rightarrow \mathrm{CO} \Rightarrow$ *telos*	$\mathrm{CO} \Rightarrow$ *telos*[18]

[17] Holders of the view that chance is an incidental *final* cause include Lennox, 'Teleology, Chance, and Aristotle's Theory of Spontaneous Generation', *Journal of the History of Philosophy* 20 (1982) 219–38, e.g. at p. 233. Aristotle's explicit statement directly contradicts this. The only passage which might seem to favour the final-cause interpretation is Aristotle's remark at 197ᵃ1–2: ἔστι δὲ τὸ τέλος, ἡ κομιδή, οὐ τῶν ἐν αὑτῷ αἰτίων. I agree with Lennox that Aristotle is explaining why the lucky event satisfies the 'coming to be incidentally' condition. Clearly the word τέλος is suggestive of final causation, but τῶν ἐν αὑτῷ αἰτίων is equally suggestive of efficient causation; thus Aristotle could just as well mean that the *desire* for the end was not a *per se* efficient cause as that the end was not a *per se* final cause. What is more, at 197ᵃ17–18 Aristotle describes the possible incidental causes of the lucky event as desires for various ends.

[18] 'CO' is a place-holder for the chance outcome, and '*telos*' for the benefit; as before, '+' signifies concurrence in a subject, or accidental unity, while '⇒' now signifies *per se efficient* causation.

The analysis of chance as incidentally caused enables Aristotle to avoid denying the existence of chance or having to take chance events to be altogether random (i.e. with only forwards connections with the causal order), and yet also enables him to avoid taking chance to be a causal force in its own right.

Now that we have a sketch of incidental causation, we can ask, what is it for an event *E* to come to be incidentally? Given the way in which Aristotle links this notion with that of incidental causation, three possible answers suggest themselves:

(1) *E* has at least one incidental cause.
(2) *E* is incidentally caused, but has no *per se* cause.
(3) *E* is incidentally caused by at least one, and is *per se* caused by none, of some favoured set of events.

In section IV I shall defend the view that (3) is Aristotle's answer.

The final question to be asked in connection with the first definition of chance is: what has happened to the idea that chance events happen neither always nor for the most part? Aristotle began the chapter by outlining two conditions: being rare and being 'for the sake of something'. His obvious intention is to define chance events in terms of the events which satisfy both of these conditions. In the account of chance and luck which follows, and which culminates in the first definition, the second of these conditions is straightforwardly carried over, as we have seen. The rarity condition, on the other hand, disappears, and is replaced by 'coming to be incidentally'.[19] Yet that he intends rarity to be part of his analysis of chance is clear both from his strategy at the start of chapter 5,[20] and by the fact that the rarity condition reappears later in the chapter (197ᵃ3–5, 18–20, 30–2).

Clearly Aristotle intends the condition that the chance event 'comes to be incidentally' to be the *same* as the condition that it happens rarely. How can this be? Coming to be incidentally is a matter of being incidentally caused in some way, and hence of being incidentally connected in some way; so these incidental connections must be *rare* connections. That what is incidental is rare is, of course, confirmed by *Met. Δ*. 30 (quoted below, pp. 86–7) and *E*. 2.[21] The most natural way

[19] At 197ᵃ5–6 Aristotle gives a definition of τύχη in which the rarity condition is likewise replaced by the condition that luck is an incidental cause.

[20] Note also that his remarks there must imply that rarity is a necessary condition of being a chance event (196ᵇ10–17, quoted at the beginning of section II above). For rarity as a necessary condition for chance, see also the references listed in n. 7.

[21] I do not of course include in this claim τὰ καθ' αὑτὰ συμβεβηκότα (see *An. Post.* I. 4–6; *Top.* V. 3–5, VII. 5; *Met. Δ*. 18 and 30).

to cash out the idea of rarity in the case of incidental causation is to take Aristotle to hold that tokens of a given type of incidental 'effect' only occur rarely within the range of occurrences of tokens of the incidental cause. I shall return to this idea in the next section.

III

I shall leave the analysis of chance to one side for the moment and ask in this section and the next what Aristotle means by saying that an event happens always, or for the most part, or rarely. These expressions are judgements of frequency—judgements as to how often a given event happens or a state of affairs holds.[22] This point may seem too obvious to mention, but I do so for two reasons. First, it is a familiar fact that Aristotle very closely associates judgements of these three types with others whose reduction to judgements of frequency is highly problematic if not downright impossible: 'always' is associated with '(by nature and) of necessity', 'for the most part' with 'by nature (but falling short of necessity)', and 'rarely' with 'incidentally' (*κατὰ συμβεβηκός*).[23] How to give a good account of these various associations has long been a vexed question; but clearly the right place to start is by asking precisely what sort of judgements of frequency figure in them.

Second, while commentators have tried to give accounts of one or other of these problematic associations, and have discussed the frequency-judgement relevant to the particular association in ques-

[22] 'Frequency' is not meant in any technical sense, of course; nor do I mean to suggest that Aristotle developed any sort of probability theory.

[23] 'Always' and 'of necessity': *Int.* 9, 19^{a}9–18; *An. Pr.* I. 13, 32^{b}4–10; *Top.* II. 6, 112^{b}1–20; *Phys.* II. 5, 196^{b}10–20; *Cael.* I. 12; *GC* II. 9, 335^{a}33–^{b}3 and II. 11; *PA* I. 1, 639^{b}23–4; *Met. E.* 2, 1026^{b}27–30. 'For the most part' and 'by nature': *An. Pr.* I. 13, 32^{b}4–13; *Cael.* I. 12, 283^{a}2–3 and III. 2, 301^{a}7–9; *Phys.* II. 8–9, especially 198^{b}34–6; *PA* III. 2, 663^{b}28–9; *GA* IV. 8, 777^{a}17–21; *EE* VIII. 2, 1247^{a}31–3; *Rhet.* I. 10, 1369^{a}32–^{b}2. 'Rarely' and 'incidentally': *Met. Δ*. 30, 1025^{a}14–19 (quoted on pp. 86–7), *E*. 2, 1026^{b}27–1027^{a}28, as well as a number of the passages just cited. The difficulty of giving a good account of these associations might lead one to think that Aristotle uses the expression 'always or for the most part' simply to *mean* 'by nature', and 'rarely' to mean 'not by nature' (Henry Mandell tried to convince me of this idea at the conference mentioned in n. 60 below). But although it is beyond doubt that it is for Aristotle an a priori truth of some sort, when he says that what happens always or for the most part happens by nature, he clearly does not intend it as a banal analytic truth or as a stipulative definition. *Mutatis mutandis*, the same points hold for the association of 'rarely' with 'incidentally'.

tion,[24] none, to my knowledge, has considered all three together. Yet it is clear that the three notions of 'always', 'for the most part', and 'rarely' must be examined and understood together, not piecemeal; and if this is so, the problematic associations should not be examined piecemeal either.

There are two types of frequency-judgement to which Aristotle's use of 'always', etc. might refer; we may call these judgements of 'absolute' and 'conditional' frequency. If the judgements are of absolute frequency, then to say that an event happens always, for the most part, or rarely means that events of this type[25] happen *all the time*, or *most of the time*, or neither. The judgement of frequency is made without reference to the occurrences of other events (except by way of comparison: 'events of this type happen less often than ones of that type'). It is this sort of frequency which is involved in saying that visits by the Queen to Christ Church are rare occurrences—she visits the college perhaps once in ten years—but diligent academic study is going on in Christ Church most of the time (there are more hours when someone or other in the college is studying than there are hours when no one is), and breathing is going on there all the time. Conditional frequency, by contrast, concerns *ordered pairs* of events: thus E_1 is usual/rare relative to E_2 iff E_2 is usually/rarely accompanied by E_1 *when* E_2 *occurs*. We make judgements of conditional frequency if we say that the Queen usually comes to Christ Church *when she is invited to open a new building*, that academic study rarely takes place in the college on Christmas Day, or that an eclipse always takes place when the sun and moon are situated thus-and-so relative to the earth. Whereas on the absolute frequency view 'always', etc. can be represented as quantifiers ranging over periods or instants of time, on the conditional frequency view they range over certain *cases*—that is, over cases in which E_2 occurs. The terminology of 'always/for the most part' is in itself neutral

[24] The association of 'always' and 'of necessity' is discussed in Jaakko Hintikka, *Time and Necessity: Studies in Aristotle's Theory of Modality* (Oxford, 1973), and Sarah Waterlow, *Passage and Possibility: A Study of Aristotle's Modal Concepts* (Oxford, 1982). 'For the most part' and 'by nature' are discussed in Gisela Striker, 'Notwendigkeit mit Lücken: Aristoteles über die Kontingenz der Naturvorgänge', *Neu Hefte für Philosophie* 24–25 (1985), 146–64, and the nature of 'for the most part' judgements in Mario Mignucci, ''Ὡς ἐπὶ τὸ πολύ et Nécessaire dans la Conception Aristotélicienne de la Science', in E. Berti (ed.), *Aristotle on Science: The Posterior Analytics* (Padua, 1981). I discuss Mignucci's paper in n. 26 and on p. 87.

[25] Obviously this categorization of events must be primarily a categorization of event-types, since individual event-tokens happen only once; event-tokens can none the less be classified in a derivative way, as being tokens of a type which happens always/etc.

between these two interpretations, because even 'temporal' words such as ἀεί can be understood (as can 'often'—and indeed 'sometimes', 'more times', etc.—in English) either temporally, or non-temporally in terms of numbers of cases.[26]

It is not hard to see which of these Aristotle must have in mind. If we classify events in terms of absolute frequency, the results are quite at variance with those of Aristotle's classification. Many of his standard examples of events which happen 'for the most part'—such as hot weather in summer, or confectionery-makers producing sweets—do not happen most of the time in this sense: there are fewer times when confectioners are making sweets or when there is hot summer weather than when no sweet-making is taking place or when it is not hot summer weather. Again, there are many celestial events, such as the conjunctions of planets, the completion of a Great Year, and so forth, which happen only very rarely in terms of absolute frequency; but Aristotle would classify them with the perpetual motion of the celestial spheres (and regard them as happening with the same necessity), rather than with royal visits to Christ Church.

Moreover, a number of Aristotelian doctrines involving the threefold classification have little prospect of making sense on the absolute-frequency interpretation. (1) Aristotle holds that there is no λόγος or ἐπιστήμη of rare events;[27] yet how could the timing of solar eclipses, or of the appearance of Halley's comet, be studied and calculated by astronomers if that were the case?[28] (2) As I have said, he associates

[26] As Mignucci points out in 'Ὡς ἐπὶ τὸ πολύ', when Aristotle does add some qualifying phrase to his talk of always or for the most part, it is sometimes a 'temporal' phrase, and sometimes a 'cases' phrase. Mignucci himself views the temporal usage as primary, and the 'cases' usage as merely reflecting the flexibility of ordinary language (pp. 185–7). But if anything the reverse seems to be true. He cites three passages in which 'temporal' words appear in connection with ὡς ἐπὶ τὸ πολύ (ποτε at *HA* VI. 4, 562ᵇ3–9, ὀλιγάκις at *GA* I. 19, 727ᵇ26–30, and ἀεί at *EN* IX. 2, 1164ᵇ30–4); in all three it is obvious that Aristotle is concerned with the frequency of cases, and not with the frequency of times. On the other hand, the passages which Mignucci sees as referring to cases (*Cat.* 7, 7ᵇ24–7; *HA* V. 20, 553ᵃ4–6 and VI. 29, 578ᵇ15–16; *PA* III. 2, 663ᵇ28–9; *Rhet.* II. 15, 1390ᵇ23–4) cannot easily be construed as referring to times.

[27] *Phys.* II. 5, 197ᵃ19–20; *Met. E.* 2, 1027ᵃ19–23; cf. *An. Post.* I. 30, 87ᵇ19–27.

[28] As Richard Sorabji has pointed out, there are some things which could come under the heading 'the timing of eclipses' but which are incidental events (*Necessity, Cause, and Blame: Perspectives on Aristotle's Theory* (London, 1980), 9–11 and n. 16); Aristotelian examples include an eclipse occurring just when the Athenians are about to withdraw, and, perhaps, an eclipse occurring *tomorrow* (although what Aristotle means at *Met. K.* 8, 1065ᵃ16 is unclear). In these cases the incidental event is the concomitance of the eclipse with some other event (if the eclipse occurring tomorrow is incidental, 'tomorrow' must be construed as an indexical: what is incidental is the connection between the timing of

'for the most part' with 'by nature', and opposes 'rare' and 'by nature';[29] but a proposition's being true at more times than it is false is simply irrelevant to the issue of whether the event which it describes occurs in accordance with nature. If the number of goats in the world were sufficiently large, for instance, 'Some goats are hermaphrodites' could be true at most or all times (cf. *GA* IV. 4, 770ᵃ35–6 and ᵇ33–6), yet this condition could still be, in Aristotelian terms, an unnatural and anomalous one for goats. Likewise, there are fewer times when a birth of a normal goat is taking place than when one is not; this hardly makes such births contrary to nature. (3) The same point holds for the doctrine that rarity is part of the analysis of chance. Thus, for instance, at 197ᵃ2–5 Aristotle says:

It is in these circumstances that he is said to go [to the market place] from luck, but if he went having chosen and for the sake of this, or frequenting the place either always or for the most part when collecting subscriptions,[30] he would not be said to go from luck.

Aristotle can hardly be saying, 'this type of case is due to luck—unless the man spends more than half his whole life collecting subscriptions in the market'.

(196ᵇ34–197ᵃ1 might seem to support the absolute frequency interpretation: '[The lucky creditor] went not for the sake of this [getting the money], but it came about for him that he went and did what was for the sake of collecting [in the non-explanatory sense: i.e. 'did what produced the beneficial result that he collected his money']—and he did this not frequenting the place either for the most part or of necessity.' It might seem as if Aristotle is saying that the creditor's getting the money would not be due to luck if his going to the agora were not itself unusual. But this would be an absurd claim for Aristotle to make: how usual it is for the creditor to go to the agora has nothing to do with whether his meeting the debtor on this occasion is a coincidence or a piece of good planning. Ross rightly suggests that we are meant to understand τούτου ἕνεκα with φοιτῶν (this is made easier by the fact that Aristotle several times glosses ἐλθεῖν with '(for the sake of) getting the money'): if the creditor usually goes to the

the eclipse and what our position in the time series happens to be). Such incidental conjunctions are not the subjects of astronomical study, whereas the interval between two particular eclipses, for instance, is.

[29] For references, see n. 23 above.

[30] There are no good reasons for deleting κομιζόμενος here, as Ross does; but my point does not depend on its retention (see below).

agora *to get his money*, we would not have a case of luck at all.[31] Now 'going there to get his money always or for the most part' still admits of interpretation in terms of conditional frequencies; but what is relevant to whether the case is one of luck is whether the man regularly goes to the agora *when collecting*—a point about conditional frequencies. The precise way in which conditional regularities are relevant to luck is discussed further in section V below.)

If Aristotle has in mind judgements of conditional frequency, all these difficulties disappear. First, the right events turn out to happen rarely, for the most part, and always; second, the various doctrines involving the threefold classification which were so problematic on the absolute-frequency interpretation promise to make sense.[32] Thus the association of 'always or for the most part' and 'by nature', and the claim that there is no ἐπιστήμη of what happens rarely, have some plausibility if the frequencies in question are conditional: the workings of nature are causal workings (operative in all four Aristotelian causes), and hence the obvious regularities to see as 'natural' are regular *connections* between things. As we have seen, the same is true for Aristotle's account of chance as rare.[33]

Further evidence supports the conditional-frequency interpretation. (1) As we saw in section II, Aristotle treats the condition that chance events are rare as equivalent to the condition that they come to be incidentally. Since coming to be incidentally is a matter of being incidentally caused in some way or other, it would be natural to construe the rarity of chance events as rarity relative to the occurrences of the relevant incidental cause(s). I shall return to this point in the next section. (2) In two important passages Aristotle applies the threefold classification in a way which is very suggestive of conditional frequency and which would be hard to square with absolute frequency. In *Met.* Δ. 30, he connects rarity with the incidental:

We call *incidental* that which attaches to something and can be truly asserted, but neither of necessity nor for the most part; for instance, if someone in digging a hole for a plant found treasure. This—the finding of treasure—is incidental for the man who digs the hole; for neither does the one come of

[31] Ross, *Aristotle's Physics*, 520; Ross presumably thinks that we should take 197^{a}2–5 in the same way.

[32] Though of course much more needs to be said to explain what Aristotle means by these claims and why he thinks that they are true. I return to these issues in sections IV–V below.

[33] On the doctrine that there is no ἐπιστήμη of rare events, see pp. 91–2 below.

necessity from the other or after the other, nor, if a man plants, does he for the most part find treasure. (1025^a14–19)

This is echoed in the discussion of teleology in *Phys.* II. 8:

For these things, and everything which is by nature, happen in a given way either always or for the most part, whereas none of the products of luck or chance do. For we do not take the frequent rains in winter to be due to luck or to coincidence, but frequent rains in summer we do; nor heat in the summer but only if it happens in winter. (198^b34–199^a3; cf. *Met.* *E.* 2, 1026^b30–6)[34]

I shall end this section by mentioning two dissenting views. In '*Ὡς ἐπὶ τὸ πολύ*', Mignucci defends a view of 'for the most part' which is essentially the same as the absolute-frequency interpretation, though he does not formulate the distinction between absolute and conditional frequency. He rejects the idea that 'for the most part' is a quantifier ranging over cases, and argues that it ranges over instants of time. His main argument[35] is that Aristotle twice associates being for the most part with being *εἰκός* (*An. Pr.* II. 27, 70^a2–6, *Rhet.* I. 2, 1357^a34–$^b1$), and that this association can only be understood if 'for the most part' ranges over times. But there is nothing in this association which favours Mignucci's times-interpretation over the cases-interpretation; in fact, it favours cases over times for reasons analogous to those given above. Mignucci thinks that, if 'for the most part' refers to cases rather than to times, then for-the-most-part propositions would be capable of being always true, and hence, via the association of 'always' and 'of necessity', of being necessary and certain rather than merely probable. But this line of argument in effect begs the question, since it assumes that the relevant sense of 'always' is 'at all *times*', rather than 'in all *cases*'.[36]

John Cooper argues that 'always' and 'for the most part' carry a much stronger sense for Aristotle than the one which I wish to ascribe

[34] A further passage, *An. Post.* II. 12, 96^a8–11, might appear to support my interpretation, since Aristotle there associates the 'always/for the most part' classification with propositions of the form 'All/most *A*s are *B*s': 'some things are not always, but for the most part; e.g. not every male human has hair on his chin, but for the most part they do.' But since Aristotle's interest here is in *syllogistic propositions*, the use of propositions of the form 'All/most *A*s are *B*s' is only to be expected, and does not tell us whether 'always' and 'for the most part' propositions are essentially of this form.

[35] See also n. 26 above.

[36] Mignucci also appeals to *Rhet.* II. 25, 1402^b34–1403^a2 in support of his view of the link with *εἰκός* ('*Ὡς ἐπὶ τὸ πολύ*', 197 and n. 51); but, as he agrees, 'le passage est difficile', and it can be interpreted so as to support either view.

to him.[37] Cooper seems to agree that the terms involve what I have called conditional frequency; but he insists that they involve the idea of *eternal instantiation* as well:

> When Aristotle speaks . . . of some arrangement as being so 'always or for the most part' . . . he means to say that that arrangement is found existing eternally or recurring regularly throughout all time, with only the occasional exceptions implied in the 'for the most part' rider. . . . [N]or would he count animal generation as happening 'always or mostly' in some particular way if animals were found in the universe only in a certain finite period of its existence, even though *when* they exist they are always or mostly generated in that way. (p. 203, n. 4)

The basis for Cooper's view is his interpretation of two arguments against 'materialism' in *Phys.* II. 8–9 which, he thinks, have as their starting points premisses concerning what happens always or for the most part. He claims that these arguments make best sense if we construe their respective initial premisses as saying that the formation of animal parts in ways which benefit the animal, and that the ordered growth of complex organisms from a given kind of seed, are *eternal* features of the world. It is this eternality which Aristotle thinks cannot be explained by his materialist opponents.

But on this interpretation Aristotle would be very blatantly begging the question: precisely because of the features of their theories which are in question, none of his opponents will agree that these processes are eternally occurring ones.[38] If Aristotle's premisses that these things happen 'always or for the most part' have this loaded sense, then he merely assumes at the outset that the theories of Empedocles and the atomists are inadequate, and presents no argument which they are obliged to consider. So we should take the initial premisses about happening always or for the most part to be making the weaker claims, neutral between the materialists' theories and Aristotle's, that observable regularities of these sorts exist (though not necessarily eternally).[39]

Furthermore, although Aristotle does believe that most if not all natural phenomena which 'happen always or for the most part' *are*

[37] J. M. Cooper, 'Aristotle on Natural Teleology', in Malcolm Schofield and Martha Nussbaum (eds.), *Language and Logos* (Cambridge, 1982).

[38] This is true not just for Empedocles, the opponent Aristotle mentions by name, but also for his other obvious target, the atomists.

[39] I wish to thank David Charles for drawing my attention to a similar criticism of Cooper by Charlotte Witt: see n. 17 of Charles's paper in this volume. Cooper's interpretation deserves a fuller treatment than I can give it here; I hope to discuss it elsewhere.

eternally instantiated, there is good evidence that he does not intend the phrase to mean that. When he uses it in II. 5 in connection with the creditor going to the agora,[40] Aristotle is clearly not considering the supposition that the man goes to the market throughout all time![41] Again, at *Met. E.* 2, 1027ᵃ22–6 Aristotle speaks of honey water curing fever for the most part: there is no suggestion that honey water's effect on fever need be regularly instantiated throughout the whole of time, or even that honey water itself regularly exists throughout eternity. Aristotle simply points to a natural regularity without regard to the question of its eternal instantiation.

IV

There is good reason, then, to take Aristotle's threefold classification as based on judgements of conditional frequency—that is, on judgements of the frequency of tokens of E_1 given the occurrence of tokens of E_2. The analysis of 'always', 'for the most part', and 'rarely' is not yet over, however, as we can see by considering an objection to this interpretation.

Let us call the event-type which occupies the 'E_2' position in the conditional schema the 'event-condition' for E_1. There is a striking difference between some 'always/for the most part/rarely' judgements and others. In some judgements of this sort, the event-condition is explicitly stated (rain *in winter*, finding treasure *when planting*), or it is obvious from the immediate context (honey water cures fever—sc. when it is given to fever victims). But in many cases, Aristotle's judgements contain no overt or contextual reference to any event-condition: are these judgements of conditional frequency?—and if so, what is the intended event-condition? These questions are especially urgent because when Aristotle speaks of his threefold classification of the totality of events—those which happen always, those which happen for the most part, and those which happen rarely—no event-condition is specified.

[40] 196ᵇ36–197ᵃ5; see pp. 85–6 above.

[41] It might be thought that Aristotle's usage is simply different when it comes to regularities in human affairs; but the passage is intended to illustrate the condition in Aristotle's general analysis of chance events that chance events happen neither always nor for the most part—the analysis to which the argument in II. 8 adverts—and there is no hint that this condition is radically different when it applies to πράξεις and when it applies to natural phenomena.

To answer these questions we must return to the analysis of chance. As I have said, the condition in that analysis that E is rare turns out to be equivalent to the condition that E 'comes to be incidentally', and three possible meanings for this phrase suggest themselves: E is incidentally caused by something or other; E has an incidental cause but lacks any *per se* efficient cause; or E is incidentally caused by at least one, but is *per se* caused by none, of some favoured set of events.

Aristotle's meaning cannot be the first of these. The condition that E comes to be incidentally is meant to distinguish chance events from other beneficial events; but every beneficial event will be rare relative to *something* and will be incidentally caused by something. Coming to be incidentally must be a stronger notion, or every beneficial event would be a chance event. The same conclusion can be drawn directly from the fact that when Aristotle applies the threefold classification to the totality of πράγματα, the resulting classes are clearly meant to be exclusive: but they would not be exclusive if being rare were simply a matter of happening infrequently relative to something or other, since events which happen always or for the most part also satisfy that condition. Likewise one and the same event could happen 'always' relative to one thing, and 'for the most part' relative to another.

The second of the three possible senses of 'coming to be incidentally' is indeed stronger in the requisite way. But it would be very surprising if Aristotle's intention here were that what distinguishes chance events from other beneficial events is that the former lack any *per se* efficient causes. The examples of incidental causation which Aristotle gives when explaining the notion of 'coming to be incidentally' (196^b26–7 and 197^a14–15) are all instances of my schema I (p. 79 above); and instances of schema I actually *require* a *per se* cause.[42] If Aristotle intended 'comes to be incidentally' to mean 'has an incidental cause but no *per se* one', he could hardly have introduced the notion in this way.

Now it is true that being an incidentally caused event of the type represented in the other schema (schema II) is compatible with having no *per se* cause; and it is also true that many chance events which fit schema I also contain somewhere a causal structure which fits schema II, and arise because of some coincidence—perhaps even all do so. Hence it would not be impossible for someone to hold that events

[42] This point is made by Charlton (*Physics I and II*, 108). He believes that it is true that chance events have no *per se* causes, and regards it as an error on Aristotle's part that he fails to recognize this truth.

which fit schema I owe their status as chance events to their connection with some event(s) which lacks *per se* causes. But again there is no hint of any of this in Aristotle's account in *Phys.* II.[43] Moreover, being incidentally caused in the schema II mode is, of course, also compatible with *having* a *per se* cause; and in the cases we are considering—coincidences—there is an obvious candidate for the *per se* cause, namely the conjunction of the *per se* cause of each of the event's co-inciding 'elements'. Thus the *per se* cause of the coincidental encounter in the agora is the conjunction of the *per se* cause of the creditor's going to the agora and that of the debtor's going there. This can be represented by the following schema:

$$\begin{array}{ccc} A & \Rightarrow & B \\ & & + \\ (A \text{ and } C) & \Rightarrow & (B \text{ and } D) \\ & & + \\ C & \Rightarrow & D \end{array}$$ [44]

While decisions on the part of the creditor to see a spectacle are in this case only rarely (and hence incidentally) connected with meeting his debtor, instances of this conjunction of causes would regularly result in their meeting—and would do so in a perfectly understandable and predictable way. Someone might object that the existence of conjunctive *per se* causes of coincidental events is incompatible with Aristotle's claim that there is no ἐπιστήμη of the incidental. But in fact Aristotle can consistently admit the existence of such causes while denying that they are objects of knowledge. Although they will produce their effects in a law-like fashion, they cannot be the objects of theoretical, scientific knowledge, since scientific understanding as Aristotle conceives it is restricted to the natures of species taken individually, and so excludes the conjunctive operations of more than one subject. And as for practical knowledge, the indefiniteness of incidental connections explains why there are no τέχναι of incidental causes (*Met.* E. 2,

[43] Nor can I see any evidence elsewhere in Aristotle's writings that he held the view that chance events lack *per se* causes. The best attempt to argue the case (for co-incidences, at least) is Sorabji's interpretation of *Met.* E. 3 (*Necessity, Cause, and Blame*, ch. 1); but I am not convinced by it (for some criticisms, see Gail Fine, 'Aristotle on Determinism', *Philosophical Review* 90 (1981), at 563–4 and 576).

[44] '(*B* and *D*)' denotes the co-presence in the agora of creditor and debtor, while '(*A* and *C*)' denotes the conjunctive event which, on this view, is the *per se* cause of their co-presence. When, as in this example, *B* and *D* have different subjects, the relation denoted by '+' will have to be weaker than 'accidental unity'; but it will still be a relation of concomitance.

1027^a19–23). Getting from incidental cause to incidental effect will involve bringing together two chains of causation; but there may be any number of ways of doing this, and no systematic way of ordering these as more and less efficient, for which means are the most efficient may depend entirely on the circumstances of the particular case. Keeping a man in a dark cellar will make him pale: but how does one get the man into the dark cellar? Violence may be appropriate in one case, the mention of Amontillado in another.[45]

How else could we interpret coming to be incidentally so as to distinguish chance events from other beneficial ones? I suggest that Aristotle has in mind the third of the possible senses of 'coming to be incidentally', and that his doctrine is this: a chance event is incidental *relative to all the natural and deliberative processes taking place in the subject*;[46] in other words, *E* comes to be incidentally if neither the φύσις nor the προαίρεσις of its subject is the *per se* cause of *E*. What marks out chance outcomes is that they are beneficial events which are not reliably connected with the natural workings or the choices of the subject who benefits from them.

It is easy to see how an account of chance along these lines could be motivated. Our aim of understanding the world about us—of making sense of the operations of nature and the strategies of rational agents—requires us to distinguish pieces of behaviour which are reliably connected with those operations and strategies from those which, even if they appear to be, are not. And Aristotle is certainly committed to something like this line of thought. For he thinks that chance events lack *per se* final causes: 'whenever things come to be not for the sake of what results . . . we say that they come to be from chance' (197^b19–20, to be discussed below).[47] It is central to his account of chance that chance outcomes involve the appearance of final causation (since they

[45] Note that one cannot maintain that conjunctive *per se* causes *as such* are impossible; for without them we could not give a sensible account of *planned* conjunctions—as when I arrange things so that the creditor and the debtor unwittingly arrive at the agora at the same time.

[46] By this I mean relative to exercises of these natural and deliberative processes by any (relevantly similar) subject, and not merely relative to *this* subject's exercise of them. In *Phys.* II. 5, it is true, Aristotle illustrates the rarity condition in terms of the same subject's activities (197^a2–5); but this illustrative procedure is really compatible with either view, and if these frequency judgements are to be straightforwardly useful in explanatory theories—as they are clearly meant to be—they should be relativized to the activities of any relevantly similar subject.

[47] See n. 8 above for references to Aristotle's opposition of 'by chance' and 'for the sake of something'.

are beneficial outcomes) while having no actual *per se* final cause;[48] and the distinction between merely apparent and actual final causation is fundamental to his natural science. Why is it that chance involves merely the appearance of final causation? Precisely because chance outcomes are only rarely connected with the natural capacities and choices being exercised in the subject, and hence the ends towards which those processes are directed cannot explain the occurrence of these outcomes. So Aristotle is committed to the view that a chance event is rare not merely relative to something or other, but relative to all the then-active workings of 'mind and nature' in the subject.

Of course, to say that Aristotle is committed to a certain view is not to say that he realizes this, still less that it lies behind a particular text. But there is in fact good reason to suppose that Aristotle is aware of the line of thought which I have sketched, and that he utilizes it in his treatment of chance; this is to be found in his *second* definition of chance.

This definition occurs at 197^b18–20, part of which I have already quoted:

> Consequently, it is clear that among the things which come to be for the sake of something without qualification, whenever things come to be not for the sake of what results, things whose cause is external, we say that they come to be from chance.

Thus *E* is the outcome of chance iff (1) *E* is among the things which come to be for the sake of something 'without qualification', (2) *E* does not come to be for the sake of what results, and (3) the cause of *E* is external.

'Without qualification' (ἁπλῶς) in clause (1) seems to mean 'in the most general sense',[49] and so indicates the broad, 'non-explanatory' sense of ἕνεκά του which characterizes chance outcomes. If this is right, clause (1) is the same as the first clause of the first definition. The first definition's second clause (the condition that *E* comes to be incidentally) has been replaced in the second definition by clauses (2) and (3). As with the replacement in the earlier definition of the rarity

[48] I ignore here the question of bad luck, which does not materially affect the present point. (On bad luck, see n. 12 above.)

[49] So Ross, *Aristotle's Physics*, 522. Lennox takes Ross to be saying that the phrase 'things which come to be for the sake of something without qualification' means 'things which generally (i.e. normally) come to be for the sake of something' (Lennox, 'Teleology, Chance', 233 and n. 30). But this is a doubtful reading of Ross, and a virtually impossible reading of ἁπλῶς.

condition by coming to be incidentally, the natural inference is that these new clauses are meant to be equivalent to the condition that E come to be incidentally—and hence equivalent to the original condition that E happen rarely.[50] Yet clause (2) by itself appears to be equivalent to 'coming to be incidentally' construed as I have suggested: for there will be a teleological explanation of the occurrence of a beneficial outcome E if and only if E has a *per se* cause in one or more of the natural and deliberative processes going on in the subject.[51] What about clause (3), the 'external cause condition'? Unless Aristotle is introducing out of the blue a wholly new condition on chance events (see n. 50), the best interpretation is to take it as spelling out the clause (2) condition in a different way; that is, to take 'things whose cause is external' to mean 'things whose cause is incidental to the operations of the subject's φύσις and νοῦς'.[52] Aristotle uses two descriptions of the same condition because they focus on different aspects of it. The first is concerned with (*per se*) final causation in chance events—or rather with the lack of it—while the second focuses on the structure of *per se* efficient causation which explains the absence of any final cause: the cause of the chance outcome is *external* precisely in the sense of being external to the natural and rational processes occurring in its subject.

If this is right, Aristotle is prepared to treat as equivalent the condition that E is rare and the condition that E is only incidentally related to its subject's natural and deliberative processes, and not susceptible of final-cause explanation—in terms of the τέλη of those processes. We have good reason, then, to think that Aristotle defines a

[50] It would be all the more surprising if clauses (2) and (3) were meant as *new* conditions, given that the definition of chance in general was the business of ch. 5, and ch. 6 is largely concerned with distinguishing between luck and chance$_N$ (chance in the narrow sense: see n. 3 above). Nor can we suppose that these conditions only apply to chance$_N$; for 197^b^20–2, and in particular τούτων at ^b^21, makes it clear that the definition at ^b^18–20 applies to chance in the wide sense—the sense which includes luck. Aristotle says here that lucky events are those which satisfy the definition at ^b^18–20 and which also satisfy two further conditions: so ^b^20–2 describes luck as a subset of the 'definiendum' of ^b^18–20. Conditions (2) and (3), then, apply both to chance$_N$ and to luck.

[51] Note that E must have at least an incidental connection with the natural and/or deliberative processes of its subject S, or S would not be its subject.

[52] This interpretation of ὧν ἔξω τὸ αἴτιον is suggested by Charlton (*Physics I and II*, 110). Note the comparable expression at 197^a^1–2: 'the τέλος—collecting the money—is not one of the causes *in him*'. Aristotle uses 'external cause' again in a puzzling passage later in the chapter, 197^b^32–7. I take the relevant sentence (^b^35–7) to mean 'but (between chance$_N$ events and events *which happen by nature*) there is this difference: of the former the cause is external, while of the latter it is internal'. Although this squares with my interpretation of ἔξω at 197^b^20, the later passage is too obscure to rest a case on it.

chance outcome as a beneficial event which is rare relative to the natural and deliberative processes taking place in its subject.[53]

In his analysis of chance Aristotle never adverts to—let alone discusses—this relativization of rarity to the subject's natural and deliberative processes. On the contrary: he introduces in a very formal way the two classifications 'beneficial/non-beneficial' and 'always/for the most part/rarely', and appears to think that it is enough to define the class of chance outcomes as the intersection of the classes of beneficial and rare events. This suggests that when no event-condition is given, Aristotle's 'always/for the most part/rarely' classification is meant to *incorporate* this relativization to the nature and choices of the relevant subject(s): to say that E happens neither always nor for the most part just is to say that it occurs only rarely on the occasions when the same natural capacities are being exercised or the same choices have been made—and likewise, *mutatis mutandis*, with 'always' and 'for the most part'.[54] Indeed, it is only by taking the threefold classification this way that we can understand it as involving conditional rather than absolute frequency, and as generating the classes of events which Aristotle takes it to generate.

V

In this section I wish to distinguish two possible views which Aristotle might hold of the relationship of 'rarely' to 'incidentally', and of 'always or for the most part' to 'by nature'.[55] This distinction is pertinent here because these associations are involved in the analysis of chance. The obvious question is, which of these views does Aristotle actually hold?; but answering this question involves more issues and more controversial texts than could be dealt with here, and is a task for another paper.

[53] This will be explored further in section V below.

[54] The case of judgements that something is or happens *incidentally* presents an analogous feature. These too are essentially relational judgements. As such they can be made relative to any condition, providing that this is specified or made obvious by the context: cold in the dog-days happens incidentally, a stumble may be incidentally worse than pleurisy (*EN* v. 11. 1138^{b}2–5). But these judgements can also have the particular function of signifying what is incidental relative to the φύσις of the subject (as they do when Aristotle employs the distinction between essential and accidental predication). In the nature of the case, however—since 'incidental' here usually modifies a predicate—the analogue of the 'event-condition' is specified via the designation of the subject of the predication.

[55] For simplicity of exposition I shall for the most part ignore *mind* here, and speak solely in terms of nature.

(1) Aristotle might take *E*'s happening rarely as by itself both necessary and sufficient for its happening incidentally to its subject, and likewise *E*'s happening always or for the most part as necessary and sufficient for its happening in accordance with the nature of that subject. On the interpretation outlined in sections III–IV, this would amount to the view that events which normally or always accompany the operations of the nature of a thing are due to its nature, and events which only rarely do so are not.

These straightforward 'equations'[56] leave no metaphysical room for regularities which are incidental to their subjects, or for rare happenings which are none the less in accordance with φύσις. We can distinguish two types of 'incidental regularity'. (i) Suppose a given species of animal, *S*, never or only rarely interbreeds with another species—that is, members of *S* refrain always or for the most part from interbreeding in this way. It seems that we ought to be able to distinguish at least in principle between cases in which this failure to interbreed is due (in part) to the φύσις of *S* in some way, and cases in which interbreeding is perfectly possible physically, and the failure is due to something incidental to *S*'s nature, such as geographical distribution.[57] (ii) An incidental regularity can also be *completely accidental*, that is, although each instance of the event has an explanation, there is no overarching explanation of why it occurs regularly. If Aristotle takes *E*'s regularity to be sufficient for *E* to belong *per se* to its subjects, he cannot distinguish either of these types from natural regularities; nor can he distinguish the corresponding types of 'non-incidental rarity' from the genuinely incidental.

This is relevant to Aristotle's analysis of chance because if he does hold this 'simple view' (as we might call it) of the connection between regularity/rarity and nature/the incidental, his analysis faces counterexamples involving all these types of case. Thus suppose the creditor knows that his debtor is always in the agora at a particular time, and regularly intends to go the agora on subscription-day to meet him; but suppose also that the creditor has a fairly bad memory, and so usually forgets to go. When he does go, his meeting the debtor, though rare

[56] That is, equations of the extensions of the relevant terms, and not equations of their meaning: see n. 23 above.

[57] Note the explanation which Aristotle reports of why interbreeding is more common in Libya than elsewhere, at *GA* II. 7, 746^{b}9–11: 'the scarcity of water means that they all meet together at the few places where there are springs, and in consequence mating takes place even between animals of different species'.

relative to the creditor's intentions to meet, is not a piece of luck (the creditor's memory is not so bad that remembering the trip to the agora is a fluke). This counter-example involves a non-incidental rarity of type (i). Again, the creditor might have a run of very good luck, and, without any overarching explanation, just happen to bump into his debtor on the right day more often than not—an incidental regularity of type (ii).

(2) Alternatively, Aristotle might regard the equation of 'always or for the most part' with 'by nature' as reflecting a fundamental truth about the connection between regularity and nature, but as being itself only very nearly true: in exceptional cases the inference from regularity to happening by nature is defeasible by considerations of the nature of the relevant subject (and perhaps by other parts of our scientific theory). On this view an event's happening always or for the most part constitutes highly privileged but defeasible evidence for the existence of a *per se* connection with its subject's nature: it is sufficient for ascribing such a connection *providing that* there are no countervailing considerations concerning the nature of the subject. This position would leave conditional frequency relative to the φύσις of the subject at the centre of the picture, while recognizing that the *relatum*—φύσις—has a role to play in the relevant explanatory theory in its own right.[58] The idea of defeasibility by considerations of the nature of the subject would make it possible to draw the distinction between the two cases of regular failure to interbreed described above; and extending the defeasibility condition to include defeasibility by other parts of the theory (i.e. parts which concerned other natures) would make it possible to acknowledge incidental regularities of type (ii) as well.[59]

The considerations adduced above show that what I called the simple view—view (1)—is false, and that an analysis of chance which relies on

[58] It is also to these considerations of the subject's φύσις that we must appeal when the actual occurrences of the 'event-condition' are too few to make a scientifically respectable judgement of an event's conditional frequency. Of course, the roles of frequency judgements and of judgements about the subject's φύσις are not as easily isolated from each other as I have pretended for the sake of exposition. This is especially true of a developing science; but Aristotle's interest in the question of how we arrive at a conception of something's φύσις on the basis of observed regularities is not itself well-developed. See *An. Post.* II. 19; Robert Bolton's paper in this volume discusses many of the questions which that chapter poses.

[59] And thus to acknowledge consistent runs of good or bad luck, once νοῦς was also taken into account. Aristotle discusses whether such runs of luck are possible in *EE* VIII. 2.

it is likewise flawed. But the simple view, and the consequent account of chance, are not absurd views, and it is not hard to imagine a philosopher holding them. Moreover, as I argued in section III, even the simple view makes much better sense than the doctrines to which Aristotle is committed on the absolute-frequency interpretation. So I think that even if it turns out that, given my interpretation of the meaning of the 'always/for the most part/rarely' classification, Aristotle has to be seen as holding the simple view, this would not be a serious objection to the interpretation: it would just be that Aristotle failed to see the need for (or consciously rejected) the more sophisticated view.

I shall explore view (2) further elsewhere, and (to let the cat out of the bag) attempt to defend the claim that Aristotle embraces it (or something like it) and rejects the simple view. But the interpretation of his analysis of chance and of the meaning of the 'always/for the most part/rarely' classification which I have defended in this paper does not depend on this claim, for the reasons which I have just given.

VI

It is a commonplace that individual substances occupy a central position in Aristotle's natural science and his philosophy alike. The interpretation advanced here of Aristotle's threefold classification of the totality of events reveals yet another way in which this is true: this classification hinges on the natures and choices of individual substances, and its application to the empirical world is entirely dependent upon what natural science discovers about the workings of those substances. This suggests that the centrality of individual substances will itself be central to understanding Aristotelian positions which invoke the threefold classification. The conditional-frequency account of 'happening always', for instance, suggests that the doctrine that what happens always does so of necessity is rooted not in his conceptions of time and modality, but in his conceptions of ἐπιστήμη and φύσις. Likewise, the 'equation' (whichever way we take it) of *E*'s happening rarely and its happening incidentally relative to all the natural processes of its subject should lead us to expect that the centrality of individual substances is itself central to other issues that involve chance and related notions; and I think that this is in fact the case. The defence of teleology against Empedocles in II. 8, the idea of *regular* spontaneous generation in the biological works, and the argument in *Met. E.* 3, I would argue, all turn on the assumption of the centrality of substances. But these too are stories for another day.[60]

[60] An earlier version of this paper was read at a conference on the *Physics* held at the University of Southern California in December, 1989; I am grateful to the participants at that conference for their comments. I am also grateful to John Ackrill, Jonathan Barnes, David Charles, Malcolm Schofield, and Jenny Tyler for their helpful suggestions at various points in the genesis of the paper.

5

Teleological Causation in the *Physics*

DAVID CHARLES

I

Aristotle's discussion of the fourth cause is introduced in *Phys.* II. 3 as follows.

> Again a mode of cause is the goal [τὸ τέλος]. This is that for the sake of which [τὸ οὗ ἕνεκα]: as health is that for the sake of which there is walking. ('Why is he walking?' We say 'in order to be healthy', and saying this we think we have given the cause.) The same applies to all things that come to be on the way to the goal [τὸ τέλος] through the action of someone else: reduction of flesh, purging, drugs or surgical instruments in the case of health. For all these are for the sake of the goal, even though they differ as some are activities, others instruments. (*Phys.* 194^{b}32–195^{a}3)

Later in the same chapter Aristotle wrote,

> the others are causes of the goal [τὸ τέλος], that is the good [ἀγαθόν] of the other things. That for the sake of which [τὸ οὗ ἕνεκα] means that which is the best, that is the goal of the things that lead up to it. And it makes no difference whether we say the good or the apparent good. (195^{a}23–6)

The fourth cause—the teleological cause—plays a distinctive role in Aristotle's discussion of physics and of metaphysics. He held that in these areas, as in biology and psychology, one must employ teleological causation in addition to material and efficient causation. Many of his predecessors erred, in his view, in thinking that such phenomena could be explained in terms of material causation alone. Thus he (famously) wrote, 'It is not likely that fire or earth or any such element should be the cause of things manifesting goodness and beauty both in their being and their coming to be' (*Met.* 984^{b}11–14). In most recent discussions of Aristotle's account of teleology, the major interest has been to discover why he thought that teleological causation was indispensable in studying natural processes, and why he thought that his predecessors had erred in attempting to avoid using it. Some have argued that this was because he believed that it would never be possible to characterize conditions sufficient for the existence of an organism or its parts if one employed only the resources of material causation and efficient causation (involving the matter alone).[1] Others have suggested that Aristotle believed that because the essences of certain psychological or biological phenomena are determined by their role in a teleological system they could not be reductively defined, even if there was a complete account in terms of material and efficient causation of the conditions under which they existed. On this view, Aristotle held that teleological causation was a type of causation which could not be understood on the basis of material and efficient causation alone.[2]

These discussions are not without interest. But there is a prior question which easily can be overlooked. Is Aristotle's basic account of what teleological causation is defensible? This is a major issue in assessing central areas of his philosophy. In this essay, my aim is to address only some aspects of this question:

(1) to seek to understand how Aristotle conceived of teleological causation, and
(2) to consider some of the resources he had to justify his employment of teleological causation.

Further, since my present focus is on *Phys.* II, I will, in the main, ignore considerations which, although relevant to these questions, are drawn from his other works. This will allow for further discussion elsewhere of whether (and to what extent) his conception of teleological causation remained unchanged throughout his writings.

If we are to achieve goal (1) we must consider at least three sub-issues.

(1*a*) In the examples cited, there is some diversity among teleological causes. Health (a state) and being healthy (a non-intentional activity) are both mentioned in 194^{b}32 ff. Elsewhere in *Phys.* II, there is an even wider variety of teleological causes:

organisms: men (200^{b}3);
artefacts: houses, axes (200^{a}10, 24);

[1] For a clear statement of this view see Allan Gotthelf, 'Aristotle's Conception of Final Causality', *Review of Metaphysics* 30 (1976–7), 226–54.
[2] See D. Charles, 'Aristotle on Hypothetical Necessity and Irreducibility', *Pacific Philosophy Quarterly* 69 (1988), 1–53; F. Lewis, 'Teleology and Material/Efficient Causes in Aristotle', *Pacific Philosophical Quarterly* 69 (1988), 54–98.

intentional actions: cutting with an axe (200^b5), obtaining money (196^b33);
natures of things: (198^b17, 194^a28).

So what, if anything, unifies this class? Do they form one ontological category? Or is there some further way in which they are unified?[3]

(1*b*) In the examples cited, there is also some variety among what is caused teleologically. Some effects are actions (walking, purging), others are instruments (drugs, surgical instruments). Elsewhere in *Phys.* II, there is an even greater range of (possible) teleological effects:

processes: raining (198^b20 ff.);
parts of organisms: teeth (198^b25);
things that come into being: (199^a8);
things being as they are: (199^a8).

How is this class to be unified? Do they form one ontological category? If not, what makes them all teleological *effects*? Is there a consistent pattern present in this apparent plethora of teleological causes and effects? What, if anything, constrains what can be included in the relevant lists?

(1*c*) What is the relation between teleological cause and effect when the former is the cause of the latter? When Aristotle writes 'walking is for the sake of health', is the relation 'for the sake of' primitive? Or did he think that it could be further explained? If so, how?

Consideration of these issues raises in a sharp form the question of how far Aristotle succeeded in justifying his employment of teleological causation. In particular, it gives rise to the following sub-issues.

(2*a*) What grounds does Aristotle offer for thinking that teleological causation is a genuine type of causation which is distinct from (e.g.) efficient causation? And this involves consideration of what, for Aristotle, makes something a genuine type of causation.

[3] There is considerable variety also in Aristotle's examples of efficient causes. Some are objects (e.g. the doctor (195^a21, 30), the builder (195^b5, 21, 23), the seed (195^a21), Polyclitus and the sculptor (195^b10)), some are objects in activity (e.g. the builder building (195^b6), this man building (195^b19), this man doctoring (195^b19)), while others appear to be states or attitudes of agents (e.g. the art of building (195^b23), wishing to see something (197^a16)). However, some of these will only be potential, not actual, causes. The latter appear to be *substances in activity* (195^b16 ff.), while the former are the substances themselves which are potentially active in the way required to bring about the relevant change. Further, the connection between the builder actively building and his exercising his skill as a builder is a close one. By contrast, the examples of teleological causation are too diverse to be ordered in this way.

(2*b*) To what extent does Aristotle succeed in establishing that teleological causation is a genuine type of causation? Or, more minimally, does he succeed in establishing that it is as defensible a mode of causation as efficient causation?

These five sub-issues are distinct from a further question which I will attempt to exclude (as far as possible) from the present discussion.

(3) Does Aristotle give good grounds for taking teleological causation to be *indispensable* in his account of biological or psychological phenomena?

Favourable answers to (2*a*) and (2*b*) would not, by themselves, lead to an affirmative answer to (3). For Aristotle might have taken teleological causation to be a genuine type of causation, but have thought that was not essential for an adequate explanation of biological or psychological phenomena.

II Teleological Causes and Effects

The *relata* in *Phys.* II. 8 fall into two broad categories. The first concern agency, the second natural organisms. They may be classified (somewhat roughly) as follows:

(1) *Cases of agency*

- (i) Actions done for the sake of something, which may be either a state *or* a further action;
- (ii) Objects (or features of objects) which exist for the sake of actions done for the sake of something, which may be either a state or a further action.

(2) *Cases involving natural organisms* (and not involving purposive agency)

- (iii) Processes of formation of organisms which occur for the sake of the whole organism;
- (iv) Parts of organisms (or features of those parts) which are present for the sake of the whole organism.

In the *Physics*, the first category includes cases both of human (e.g. craftsmen: 199^{a}12 ff.) and of animal agency (e.g. spiders: 199^{a}22). The second (as I define it) includes cases such as the formation of teeth (198^{b}25 ff.) or the location of the roots (199^{a}27 ff.) which do not involve imagination, *nous*, or anything (psychological) of this type. Aristotle says little about the cases of sub-human animal action in the *Physics*. In what follows, therefore, I will compare his account of human agency with that of natural phenomena not governed by imagination, or anything (psychological) of this type. Aristotle's discussion of spiders, swallows and ants raises important issues, but they lie outside the scope of the present paper.

In the *Physics*, Aristotle does not seem concerned to organize these cases further. Thus, he treats (i) and (ii) as on a par (195^a1–3) as both occur 'for the sake of a goal'. He does not even attempt to argue in the cases involving *agency* that (i) is more basic than (ii) in the order of teleological causation. While he notes that both (iii) and (iv) obtain in cases of organisms (199^a6–8, 30), he makes no attempt to argue that either (iii) or (iv) is the more basic (in the *Physics*). It is true that in *Phys.* II. 8 he focuses most insistently on (i) (the case of the craftsman), and (iii) (the case of natural genesis). But this seems to be only because he is pursuing the analogy between artistic and natural production, rather than because he is taking these to be the central cases of teleological causation in terms of which (ii) and (iv) are to be explained. When the point of the analogy is completed, he alludes to cases where the explanandum is an organism in existence rather than one coming into existence. But there seems to be no attempt here to take (iii) as the basic case of teleological causation in terms of which (iv) is to be understood.

Outside the *Physics*, Aristotle does on occasion claim that in the order of exposition (iv) should be stated before (iii), and notes that because man has certain properties, the process of his formation must be such and such and take place in a given manner (*PA* 640^b1–3). But Aristotle is, it seems, doing nothing more here than saying that you cannot understand the goal or nature of generation if you do not understand what the goal or pre-existing agent is—a man with his parts. So we need to understand what man is before we understand the processes of generation in which other men are produced. But this in no way commits Aristotle to thinking that he can or should define

> *A*'s *coming to be* for the sake of a goal (e.g. man)

in terms of

> *A*'s *being* for the sake of a goal.

Rather they seem to be treated as separate cases of teleological causation.

It appears that Aristotle accepts a plurality of cases of teleological causation without attempting to explain them in terms of one more basic case. If he had tried to do this, he could have unified his account of teleological causation by showing how all other cases could be explained on the basis of cases of (e.g.) the processes of formation of the organism which occur for the sake of the whole organism. This would have been a major task. It would have required him to show that parts of animals exist for the sake of an end solely because they are the focus of capacities which play a role in producing whole organisms. Also, it would have been incomplete unless he could also show that human actions are done for the sake of a goal only when they are steps towards the formation or maintenance of the relevant organism (or thought to be such). But there are no indications in the *Physics* or elsewhere that Aristotle actually engaged in attempting to carry through this type of project. Nor is there clear evidence that he was convinced that it could be carried through, or even that he thought it a project worth pursuing.[4]

This raises the question of whether there is one unified conception of teleological causation in Aristotle's account. If he did not attempt to take one of these four cases as basic, and explain the others as derived from this, he could have established the unity of teleological causation only by showing (as an alternative) that all these cases exemplified a common form. Indeed, this might appear to be the strategy he wishes to follow in the *Physics* when he treats agency and natural-organism cases as analogous with one another, with neither as more basic. Thus, he argues both 'as in art, so in nature' and 'as in nature, so in art'. (*Phys.* 199^{a}9–10, 15–16) However, superficially, cases (1) and (2) seem

[4] For a contrasting (but unsupported) claim see Gotthelf, 'Aristotle's Conception', n. 19. Gotthelf reiterates this claim in the 1986 Postscript to his earlier article in A. Gotthelf and J. Lennox (eds.), *Philosophical Issues in Aristotle's Biology* (Cambridge, 1987), 239 ff., and provides one consideration, based on one introductory text in *PA* I, in favour of it. It should be noted that even if this argument were successful, it would not establish Gotthelf's general claim, which concerns *all* cases of teleological causation, as the relevant text does not discuss examples of human agency. However, the consideration he adduces is at best inconclusive. In *PA* 640^{a}22 ff. Aristotle notes that Empedocles failed to recognize that in cases of animal generation the producer was a pre-existing agent ('as man produces man': 640^{a}25–6). He then continues (after a break) in 640^{a}34: 'Hence we should say that because man is such and such, therefore a man has these things'. But this sentence appears to emphasize *not* that man has certain parts because they come to be in a given way (as Gotthelf claims), but rather that generation occurs as it does *because* man has certain parts (cf. 640^{b}1–3). And this appears to favour, if anything, *some* form of priority of case (iv) over case (iii) (even if only in order of exposition). Gotthelf noted in 'Aristotle's Conception' that he needed to establish that Aristotle accepted the general explanatory priority of case (iii) over cases (i), (ii), and (iv). This claim, which is central to Gotthelf's important thesis, has yet to be supported by detailed analysis of either *Phys.* II. 8–9 or the *Parts of Animals* (with its wealth of specific examples). A full investigation of the latter cases would be a major, and worthwhile, undertaking.

different. In (1), the goal appears to be the object of an agent's desire or choice. And this encourages the thought that an action is done for the sake of a goal *G* if and only if the agent does that action with the further desire to produce *G*. Similarly, an instrument may exist for the sake of goal *G* if and only if that instrument is made by an agent with the desire to achieve *G*. And in these cases the agent is sensitive to the value of the goal. It is because the relevant state seems valuable that the agent pursues it (*Met.* 1072^{a}27 ff.). Were it not to seem valuable, she would not do so. But in (2), desire or choice are lacking. Aristotle rightly shows no sign of attributing desires or choices to all natural organisms. So in (2) the goal is not itself the object of an agent's desire or propositional attitude. And it appears that the organism need be not sensitive to the value of the end state produced. Were it to cease to be valuable, the organism would not immediately cease to pursue it.

In cases of agency, the agent is also sensitive to what is the best *means* to the goal. In a variety of cases, were the best means to achieve a goal to be different, the route the agent would follow would be correspondingly different. And this shows that the agent acted as she did *because*, for instance, the action she chose was the best means to the goal. It was not that the action chosen merely happened to be the best means to the goal. The agent's internal states (e.g. of perception or belief) were sensitive to the fact that this was the best route to her goal and she chose it for that reason.

The situation seems different in the cases of natural organisms, considered in (2). Parts of organisms (such as hearts) exist for the sake of the whole organism in that they serve a function in (e.g.) the maintenance of the life of that organism. Thus we say (e.g.) that the heart produces certain consequences which contribute to the survival or flourishing of that organism. This thought is sometimes strengthened by the further claim that organisms of this type have survived and reproduced in the past *because* they possessed hearts of this type. On both these views, a part has a function if and only if it tends (in one or these ways) to ensure the survival or other characteristic goals of the organism.[5] But the goal is obviously not the propositional object of the

[5] Similarly, a process would have a function if and only if it tends to produce a given state which is beneficial to the organism's survival or flourishing. And we might add to this the additional historical claim that in the past processes of this type played this role, and this contributed to the survival of the species. For claims of this general form, see L. Wright 'Functions', *Philosophical Review* 82 (1973), 161, and Ruth Millikan, *Language, Thought and other Biological Categories* (London, 1984), 28 ff.

relevant part's 'desire', nor is it selected by the part *because* it seemed or was good. Rather it is determined by an external characterization of what flourishing or survival for that organism consists in. The part in operation contributes to that end, but need not itself be sensitive either to the fact that the state produced contributes to the organism's flourishing, or to the goodness of the state produced. Further, it is not claimed that were the best means to that state to be different, the part would function in a correspondingly different way. Thus, if the best way for the heart to function to contribute to the organism's survival were to change (and it was required that it pump blood in the opposite direction), the heart's activities would not alter accordingly. The heart itself is not sensitive to the fact that the route it follows is the best way to achieve its goal. Even if it does take what is, in fact, the best route to its goal, it does not do so *because* it is the best one to achieve that goal.[6] At most, the heart has a tendency to produce a state which is a good for the organism to which it belongs by a route which reliably produces that goal. But in this case neither the route itself nor the goal is produced by an internal state of the organism which is sensitive to the fact that this is good or the best route to the relevant goal.

There are, it appears, major differences between the types of teleological causation at work in the cases of agency and natural organisms (in this account). In the case of agency, both the goal and means are selected because they are good for the agent and the agent is sensitive to this fact. In the examples of natural organisms, parts may indeed function in ways that are good for the organism. But neither the goals nor the means need be selected just because the organism is sensitive to their goodness. The goodness of the goals or route is not a necessary feature of the explanation of what occurs. The part (or organism as a whole) may have a tendency to promote a state, which is in fact good. But it need not itself be sensitive to the goodness of the end state.[7]

[6] A similar point could be made if the route was specified as the 'one required for the goal'.

[7] This is not, of course, to say that one can determine what things are goals without reference to the good of the organism. It is to say only that the parts themselves are not *sensitive* to the goodness (or goal-status) of the end-state produced. In 'The Place of the Good in Aristotle's Natural Teleology', in John J. Cleary and D. C. Shartin (eds.), *Proceedings of the Boston Area Colloquium in Ancient Philosophy* iv: 1988 (Lanham, Md. 1989), Gotthelf argues for the more radical thesis that Aristotle thought that one could determine what an organism's end is (e.g. its full actuality) without reference to what is good for it. This claim is difficult to sustain in *Phys.* II as Aristotle here introduces the notion of an end in a way which is at least partially explained by what is good (195^{a}23–6 ff., 194^{a}28–

This point can be brought out by the following example. In certain cases, a species (e.g. the mosquito) may develop over time new methods for overcoming its environmental challenges (e.g. DDT) and, eventually, find a route which allows it to escape this danger. This is, on one account, because DDT-resistant mosquitoes have higher successful reproduction rates than those that are not resistant, and eventually come to predominate in the mosquito population. This differs from the agency case in that it takes a period of time for the adjustment to occur. But more radically, the adjustment occurs because (e.g.) the non-DDT-resistant mosquitoes die out, and the ones which survive and reproduce are DDT-resistant. There is nothing in the mosquito (species) itself which is sensitive to the fact that the latter route is better for its survival chances. Rather the environment blindly destroys those mosquitoes which are not DDT-resistant, without the species itself selecting *as good* a means to overcome the challenge of the environment.[8] And this contrasts sharply with cases of purposeful agency. While in both cases one may use the locution '*a* occurs for the sake of goal *G*', it involves something completely different in examples of functional and purposive explanation.[9]

The difference in ontology between Aristotle's varying teleological causes and effects is startling enough. But the apparent differences within the genus of teleological relation between cases of agency and of natural organisms' function are even more threatening. And this naturally prompts the question: is there one unified account of teleo-

33). Elsewhere, Aristotle characterizes full actuality (partially) in terms of certain intrinsic goals (*Met.* 1048^{b}17 ff., 1050^{a}21–3) rather than explicating these goals in terms of the relevant actuality. But what are these goals if not what is good for the relevant organism? (I intend to argue for this claim in more detail elsewhere. It is, in my view, essential for a proper understanding of Aristotle's strategy in *Met.* Θ.)

[8] This is not to suggest that Aristotle himself countenanced or developed this type of causal story. The issue is whether his account leaves a gap which could be filled by a story of this type.

[9] There is a further possible difference between the cases of agency and natural function which should be noted. Animals' natural flourishing has some specific content in Aristotle's account which remains relatively unchanged under different conditions. Human agents, by contrast, deliberate towards the apparent good, and their actions can be caused teleologically even when they do not, in fact, aim at their real good. (cf. *Phys.* 195^{a}25–6). Since their real and apparent good can differ, their desires are sensitive to the goodness of their (proposed) goals in a way not paralleled in the case of non-deliberating animals. For it is this, and not (e.g.) their tendency to promote their full actuality, which is common to all their relevant desires. (Lindsay Judson alerted me to this.) However, this is a separate point from the one on which I am concentrating; for mine concerns the flexibility of means pursued for ends taken to be good rather than the variability of ends of a human agent.

logical causation at work in Aristotle's discussion of these topics? Did he believe that these two kinds of teleological causation were the same? Alternatively, did he think that the second, natural organism, model was the basis for understanding all cases of teleological causation? Or did he take the first, agency model, as his starting point, and represent all cases of function as exemplifying this stronger form of teleology? Or perhaps the differences between them did not matter for his concerns? It is only when we have an answer to these questions that we can begin to assess Aristotle's account of teleology, and grasp the distinctive kind of causation it involves.

Recognition of the unreduced plurality of types of teleological causation at work in the *Physics* serves to render untenable one otherwise attractive view of what its operation consists in. If one focuses exclusively on discussion of case (iii)—the process of formation of natural organism—and adopts the 'function-model' interpretation, it is natural to construe Aristotle's teleological causes as examples of efficient causation with a distinctive type of cause (e.g. the potentiality for or *nisus* towards the fully-formed organism) or a distinctive type of effect (fully functioning organism, or its constitutive activities). On this view, the basis for teleological causation would be a *nisus*, or potentiality, to bring into being a fully formed organism. And the connection between the *nisus* and its effect need only be that the former is the efficient cause of the latter.[10] One may doubt whether this is the correct model to understand cases of type (iii).[11] But even if it were, it

[10] This model for understanding teleological causation is influential. Gotthelf, 'Aristotle's Conception', adopts the view that what is distinctive of teleological causation is the presence of a certain type of efficient cause (e.g. an irreducible potential for form). Martha Nussbaum has regarded teleological causation in certain cases as marked out by its distinctive effects (*Aristotle's De Motu Animalium* (Princeton, 1978), 84).

[11] This view would account for cases of type (i) if one took (e.g.) desires to exemplify a *nisus* of this type, and gave up the idea that (e.g.) the *desire for* ϕ was itself essentially teleologically grounded in ϕ *being* or *seeming good*. But it would appear to require a major change in Aristotle's view of desire (cf. *Met.* 1072^{a}29–31). I argued for the alternative, good-based, interpretation of Aristotle's account of desire in *Aristotle's Philosophy of Action* (London 1984), 85–8, 229–33, 237–40. If it is correct, Aristotelian desires are essentially defined in terms of the goodness (real or apparent) of the state at which they aim and the good-based rationalizing connections with other psychological states in which they stand (e.g. in practical reasoning). Indeed, these two teleological features (sensitivity to the goodness of the goal and the means) serve to define desire in a way which (I claimed) is inconsistent with the standard functionalist account of the goal-directedness of desire (in terms merely of the state it tends to produce). For in the latter, the goodness of the end-state and the good-based rationalizing connections play no essential explanatory role. If my account of Aristotelian desire is correct, it seems better to argue by analogy that case (iii) is mishandled on any view which takes the existence of the relevant *nisus* as a basic datum (e.g. a potential for Form), and fails to explain its presence or nature in

could not account for cases (iv) or (ii). For here there is no possibility of the existence of parts being the *efficient cause* of the existence of the whole. There is no process of which the part is the starting-point or controller which culminates in the existence of the whole. If so, the general form of Aristotle's teleological causation cannot be given by construing it as an *efficient* cause with a distinctive type of antecedent or consequent.

III *Physics* II. 8: How Many Models at Work Here?

In discussing teleological causation in this chapter (and the next), Aristotle is addressing a specific type of opponent. This opponent, in effect, makes three separate claims.

(1) Everything that occurs of necessity, occurs as a causal consequence of the nature of simple bodies and their movements.[12]

(2) The necessity involved in the nature of simple bodies and their movements is simple, non-hypothetical, necessity.[13]

(3) Whatever occurs not as the result of the nature and movement of simple bodies, occurs by chance. Amongst the latter category are the development and continued propagation of a given species of animal.[14]

Given that Aristotle's opponent holds both (1) and (2), he will hold that everything that occurs necessarily is an example of simple necessity. If anything occurs which is not a case of simple necessity, it occurs 'by chance'—(3). So his world is divided into two exclusive and exhaustive sets of phenomena: those that occur by simple necessity,

terms of its being (or seeming to be) for the good of the organism. (For a contrasting view of these issues, see Susan Suavé's Review of D. Charles, *Aristotle's Philosophy of Action*, *Philosophical Review* 97 (1988), 413. Suavé regards as 'not foreign to Aristotle' a minimal view of teleology according to which a process is teleological if and only if the essence is to produce a given result, which need not itself be specified as good. In the account offered here, the goodness of the good is explanatorily essential in Aristotle's account of action, desire and function. See p. 122 below.)

12 *Phys.* 198^{b}12–16. The opponent does not take seriously the possibility that there are other causes beyond the hot, the cold, and things of this sort. This latter claim need mean no more than this: whenever there is a causal relation whose relata are not (e.g.) the hot and the cold, but (e.g.) a man, this causal relation depends on one which involves (e.g.) the hot and the cold. The opponent is not saying that *everything* that exists occurs of necessity. He is making the weaker claim that everything that occurs of *necessity*, is caused in a given way. This is consistent with holding that some things occur by chance.

13 *Phys.* 199^{b}35–200^{a}4.

14 *Phys.* 198^{b}27–32.