ARISTOTLE ON CAUSATION AND CONDITIONAL NECESSITY: *ANALYTICA POSTERIORA* II 12 IN CONTEXT

Inna Kupreeva*

1. Introduction

One of the debated questions concerning Aristotle's theory of causation is whether it presupposes anything like a necessitation of the effect by its cause. Some texts in the corpus suggest that there is no necessitation in any familiar sense of necessity; others imply that there is 'simple', 'unconditional' necessity operating in natural processes. In this paper I attempt an analysis of a relatively little studied chapter of *Posterior Analytics*, II 12, which may prove useful for our understanding of the problem.[1] This chapter is devoted to the question whether all kinds of causation involve necessitation. Aristotle's main concern is specifically with the case where the cause precedes its effect in time. The term 'conditional necessity' is not used, but Aristotle's discussion sheds some important light on the ways in which this concept is used elsewhere in the corpus, notably in the last chapter of the treatise *On Coming-to-Be and Perishing* (*GC* II 11) that is devoted to the question whether there is necessity in the world. The paper falls into two parts. The first contains an analysis of Aristotle's argument according to which inferences about a causal process where cause precedes its effect in time should conclude from effect to cause, and not from cause to effect, if they are to be valid. The second part of the paper uses this argument as a background for the

* Earlier versions of this paper were presented to the ESF conference 'Interpretations of *Posterior Analytics*' in Leiden 2004 and the Philosophy Omega Seminar in Edinburgh 2005. I am grateful to the organisers (respectively, Frans de Haas and Jeffrey Ketland), the audiences, and to Natasha Alechina, Pieter Sjoerd Hasper and David Sedley for helpful discussions of the subject matter. It will be clear from the paper that I am hugely indebted to Bob Sharples's groundbreaking work on the subject (even though I disagree with some of his conclusions). He is badly missed.

[1] The recent commentary by Detel (1993) is an outstanding contribution that almost makes up for the lack of critical studies; it provides bibliographical references to each chapter.

analysis of the discussion of conditional necessity in natural processes in *GC* II 11, and in the work of two ancient commentators, Alexander of Aphrodisias and John Philoponus.

2. Aristotle's Humean Doubts

The difference between the Aristotelian and modern concept of cause is generally well appreciated: Aristotle's causes, unlike modern, are taken to be reasons or 'becauses', 'explanatory items' rather than causal factors (events or objects) necessitating their effects. However, in several places Aristotle does speak of causes in the meaning close to modern; and in *APo*. II 12, this way of speaking is put in a full-fledged theoretical perspective as Aristotle raises a problem similar to Hume's problem of whether there is a 'necessary connexion' between cause and effect.

2.1. *The Problem*

Aristotle begins by drawing a distinction between the causes of being and the causes of coming to be and perishing:

> (T1) What explains why something is *coming about* (and why it has come about, and why it will be) is the same as what explains why *it is the case*: it is the middle term which is explanatory. But if something is the case, the explanatory item is the case; if it is coming about, [the explanatory item] is coming about; if it has come about, it has come about; and if it will be, it will be. (*APo*. II 12, 95a10–14, trans. Barnes.)

The purpose of this distinction between the causes of being (henceforth, B-causes) and the causes of coming to be (G-causes) is to draw attention to the explanations of particulars, which, as Aristotle rightly suspects, may involve some differences compared to the explanations on the level of essences.

The logical form of a full B-causal statement is that of a syllogism, where the explanatory item (the cause proper) is expressed by the middle term.[2] In *APo*. I 13, Aristotle gives us an example of causal demonstration:

[2] Aristotle explains this in detail in *APo*. I 13 and then again (for different types of cause) in II 11. I use the term 'syllogism' loosely, referring to a deduction via a middle term, irrespective of quantification over the terms (what some authors call 'proto-syllogism' or 'deduction' in order to distinguish this form from the syllogism proper discussed in *APr*.).

> (T2) Let C be the planets, B being near, A not twinkling. B holds of C and A of B: hence A holds of C. The deduction gives the reason why, since the primary explanation is contained by the premises.
>
> (*APo*. I 13, 78a39–b4, trans. Barnes)

The relation 'holds of' between the predicate and the subject can be construed as inclusion of the set corresponding to the subject term in the set corresponding to the predicate term. In our example, {C⊆B, B⊆A} ⊢ C⊆A, more generally, {S⊆M_B, M⊆P_B} ⊢ S⊆P_B.[3] This statement explains the resultant state of affairs by showing in its premises the causal constituents of this state. A full G-causal statement does not differ in logical form, but the copula-*cum*-predicate constructions are 'tensed'. Aristotle mentions three tenses covered by G: the present that can be best rendered by present continuous in modern English (γινόμενον), future (ἐσόμενον)and past (γεγενημένον). He treats all the G-predicates with the same tense index as simultaneous, and so the interpretation of premises and conclusion is not affected by the introduction of tenses as such.

> (T3) (i) E.g., why has an eclipse come about [E]?—Because the earth has come to be in the middle [M]. And [the eclipse] is coming about because [the earth] is coming to be there; [the eclipse] will be because [the earth] will be in the middle; and it is because it is. (ii) What is ice?—Assume that it is solidified water. Water C, solidified A; the explanatory middle term is B, complete absence of heat. Thus B holds of C; and being solidified, A, holds of B. Ice is coming about if B is coming about; it has come about [resp.] if it has come about; and it will be if it will be.[4]
>
> (*APo*. II 12, 95a15–21, trans. Barnes)

We can illustrate this in the following way using symbols:

(i) P_G (M)	(the earth's being in the middle comes to be (was, will be) the privation of light)
M_G (E)	(the eclipse comes to be (was, will be) the earth's coming in the middle)
P_G (E)	(the eclipse comes to be (was, will be) the privation of light)

[3] Barnes chooses to express this relation by means of what he calls a 'copula variable', φ, ψ etc. (running over the expressions 'is', 'was', 'will be', and 'is coming to be'). In addition, he introduces something like a sorted variable over things and events that may be covered by the predicate terms which he calls 'dummy subject': mψM will translate as 'the earth is in the middle'; eψE ('there is an eclipse'), etc. I use predicate indices for the same purpose, and traditional notation for the subject terms (although, as will be clear, these will change their reference in accordance with the indexical value of the predicate).

[4] Cf. Barnes (1981), 38n.32, who cites this as an example of syllogistic imposed on the original non-syllogistic deductions of the Apodeictic. 'Eclipse' in this example means 'eclipsed moon'. I am grateful to Mariska Leunissen for the query.

The use of tenses (especially in the minor premise and conclusion) does not mean that the definition of eclipse is 'relativised', but only that it is valid in each of the three 'generational' tenses as well as in the common present. Similarly, we get an explanation of the event of ice-formation, by taking G-predicates instead of B-predicates.

(ii) B_G (C) (water comes to be (was, will be) deprived of heat)
A_G (B) (absence of heat is (was, will be) constitutive of solidification)

A_G (C) (water comes to be (was, will be) solidified)

The G-deductions are valid as long as the simultaneity of events under the same tense index holds good.[5]

> (T4) When an item which is explanatory in this way and the item of which it is explanatory come about, then they both come about at the same time; when they are the case, they are the case at the same time; and similarly for 'have come about' and 'will be'. (*APo*. II 12, 95a22–24 trans. Barnes)

This is probably a methodological device used by Aristotle to prepare the stage for the examination of the main case, namely where premises have predicates with different tense-indices:

> (T5) (i) But what of items which do not hold at the same time as one another? (ii) Can it be that, in continuous time, as we think, one such item is explanatory of another? (iii) Can the fact that this item has come about be explained by something else which has come about, the fact that this will be by something else which will be, the fact that this is coming about by something which has come about earlier?
> (*APo*. II 12, 95a24–27 trans. Barnes)

We have to note the way Aristotle uses the notion of cause (explanatory item) here. So far we assumed that the middle term is an explanatory item; but now the question apparently is whether the deduction holds if the middle term has different tense-indices in the two premises, and the answer is: no, because the syllogism is just not well-formed. Consider the deduction: $M_P(S)$, $P_F(M) \vdash P_G(S)$ (S has been M, M will be P, hence S is now (or will be) P): how does it fare?

The problem Aristotle raises can be compared with Hume's problem of necessary connexion between the two successive events. But unlike Hume, Aristotle does not intend to subvert the explanation on the level

[5] Sorabji (1983), 50, mistakenly says that according to Aristotle in this chapter, a syllogism cannot be constructed if all its terms are future-tensed: in fact, Aristotle says, it can, as long as the simultaneity is preserved.

of essences: that is supposed to hold. The question he raises is what happens with the explanatory value of causal account when we move from essences to 'particularised', individual statements. If the explanatory value is upset on the level of individuals, one can see this as potentially damaging to Aristotle's metaphysics of hylomorphism: we are back to the Platonic question whether there can be any knowledge of particulars, in this case of individual events.

Although he initially formulates the problem using the technique of the syllogism, in his analysis of the solution, Aristotle switches to what seems closer to a conditional analogue of syllogistic deduction. His reply to the question is that not all the logical relations between the antecedent (p) and consequent (q) depend on the truth-functional outcome of q's coming to be when p is true. The converse of the original conditional is also relevant:

> (T6) The deduction (ὁ συλλογισμός) is possible if it starts from what has come to be later (but the principle in this case [i.e. in the case of things that have come to be] is something which has already come about), and similarly with what is coming about; but it [the deduction] is impossible if it starts from what is earlier (e.g. 'Since this has come about, this has come about later'). And similarly for what will be the case. For whether the time is indeterminate or determined it will not be the case that since it is true to say that this has come about it is true to say that this—the later item—has come about: in the interval, when the one item has already come about, the statement will be false. (*APo*. II 12, 95a27–34, trans. Barnes, modified)

Paraphrasing the inference from the earlier to the later in terms of propositional logic, we get the formula $p \rightarrow q$, which is not logically true. And even if it is true of a kind of event *q* that '*q* because *p*', *q* cannot be derived from *p* by a sound inference: for in the case where there is a temporal interval between p and q, $p \rightarrow q$ is liable to be false. The deduction which starts 'from what is later' must have a different logical form. The difference between the two inferences can be described as follows. Let us take the relation 'p is the cause (explanatory item) of q' or 'q because p' to be represented by 'p is the necessary condition of q', which can be paraphrased, using propositional logic, as '$q \rightarrow p$'. (This is the weakest representation, sufficient for the purposes of the current analysis.)[6]

In the case of a direct inference from the earlier (p) to the later (q), given that (i) q because p and (ii) p, it does not follow that (iii) q; or, using symbolism: $q \rightarrow p$, $p \not\vdash q$.

[6] On this representation, see Mackie (1980), ch.2.

Aristotle agrees with Hume that the relation of 'causation' (however understood) between p and q does not convert into a valid inference from p to q. Unlike Hume, however, he apparently believes that the reverse inference, from q to p, as warranted by a causal connexion, can validate a deduction of p as the (antecedent) cause of q. Thus, in the case of 'oblique' inference, given that (i') q because p and (ii') q, it follows that (iii') p, if 'p then q' is necessary (formalising, q, $q \rightarrow p \vdash p$, just a *modus ponens*).

So, Aristotle allows for the 'transition of kind' based on a constant conjunction of cause and effect prohibited by Hume, but with this restriction: it can go only in one direction, i.e. from effect to (antecedent) cause, but not the other way around.

2.2. *Aristotle on the Ontology of Causal Processes in Time*

Setting out the 'Humean' problem, Aristotle asked: 'Can it be that, in continuous time, as we think, one such item is explanatory of another?' Having outlined the conditions of valid inference from effect to cause, he returns to this question and explores the possibility of validating the causal connexion between earlier and later events by taking them to be continuous because of the continuity of time. This possibility ultimately does not work, but the discussion is used by Aristotle, first, to back up the difference between the two kinds of inference, namely the direct inference from the earlier to the later and the oblique one from the later to the earlier, and second, to explain why the latter is and the former is not acceptable on the basis of the ontology of causal processes regarded as events in time.

> (T7) (i) We must enquire what it is that holds things together (τί τὸ συνέχον) so that after what *has come* about there are items which *are coming* about. (ii) Or is it plain that what is coming about is *not* contiguous (ἐχόμενον) with what has come about? (iia) What came about is not contiguous with what came about, since these things are limits and atomic: just as points are not contiguous with each other, so items which came about are not—both are indivisible. (iib) For the same reason, what is coming about is not contiguous with what has come about. What is coming about is divisible, whereas what has come about is indivisible: what is coming about stands to what has come about as a line stands to a point, and infinitely many items which have come about inhere in what is coming about. (But I must discuss this more clearly in my general account of change.) (*APo*. II 12, 95b4–12, trans. Barnes)

The question under investigation is whether it is possible to speak of two distinct events corresponding to cause and effect as continuous in such a

way that if the cause is already present, the effect will necessarily follow because both have temporal structure and thus share in the continuity of time. On this view, the effects would 'flow' from their causes, so to speak, so that positing the occurrence of a cause will necessarily involve an effect in due course.

Aristotle uses two concepts of continuity when speaking about temporal processes. In the statement of the problem he seems to be thinking of the continuity of time in the sense of 'division': for any two moments t_1 and t_2, where $t_1 < t_2$, there will always be a point t_3 such that: $t_1 < t_3 < t_2$.[7] The continuity thus understood is a property of the time-line as a whole or any of its intervals that are not individuated by an actual division.

The continuity referred to in our passage is the continuity 'by limits'.[8] It is a relation between two individuated events or processes in time. This concept of continuity is of interest in the analysis of a causal link between two events in a temporal sequence. Aristotle here invokes his definition of continuity developed specifically for the analysis of change in *Ph.* V 3. According to the series of definitions set out there, A is *contiguous* (ἐχόμενον) with B, if it is *successive to* (ἐφεξῆς) B and *in contact* (ἁπτόμενον) with it (*Ph.* V 3, 227a6). A is continuous (συνεχές) with B if it is contiguous with B and the limits of A and B have become identical and are held together (227a10–12). Aristotle explains that A and B are in succession if they do not have anything *of the same kind* between themselves (one could think, for instance, of numerical successions). A and B are in contact if they have their limits together, i.e., so that nothing (at all) is between them. The continuity involves an additional condition of unity of these limits.

In his solution (see T7 (ii) above), Aristotle explores the possibility of the cause and effect being made *contiguous* by the time's continuity 'of division'. His point of departure is the view of time as a continuum that is potentially divisible at any point. The parts of a time-line could be treated as continuous if they are regarded as the parts of a potential division, i.e., when they are not individuated by an actual process of division which marks the boundaries of events in actuality. The parts of potential division are presumed to be in contact; moreover, they are 'one' because the point in time in this case can be treated as both two and one,

[7] Cf. *Ph.* V 3, 227a10.21; V 4, 228a29; VI 1, 231a22; VI 6, 259a16, 19; *Met.* XI 12, 1269a5, 10.

[8] Cf. *Ph.* I 2, 185b10; III 1, 200b18–20; VI 1, 231a22–b6; VI 2, 232a24; 233a25; 233b17–31.

just as a geometrical point.[9] The actual division of time involves a certain interruption, so that the point of division will be the actual turning point between two processes of change. This kind of interruption imposes a peculiar structure on the individuated temporal parts: a change over an interval of time has no starting point but does have a point at which it is completed.[10] As a result, any actual part of time (and a single process) is represented as an interval which is open and divisible on the left-hand side and closed on the right-hand side.[11]

In T7 (iia) and (iib), Aristotle exploits the thesis that no continuum can come about from the indivisibles: the indivisibles do not have their limits as one, nor do they have their limits together, i.e., they are not even contiguous.[12] The important point of this analogy (in which points, moments, and events are jointly contrasted with sensible things) is that successions formed in the former class cannot amount to a continuum of any kind. In the case of geometrical limits and potential points of division in time, the reason is that there is always a 'stronger' continuum intervening between any two indivisible points of a line or duration (*Ph.* VI 1, 231b6–10)

Thus a causal sequence does not involve the contact of limits, but, as Aristotle tells us next, it has instead the order of succession, whose members are consecutive (ἐφεξῆς). Consecutive events (while being indivisible) do not have anything of the same nature coming between them (there is no corresponding stronger continuum of change or coming to be). If I understand him correctly, Detel takes Aristotle's argument to be the following: the continuity does not hold of coming to be because it does not

[9] *Ph.* IV 11, 220a12–13, *Ph.* VIII 8, 262a12–263b9; Coope (2005), 11–13.

[10] "The first [moment] in which a thing has changed is said in two ways, namely, the first moment in which change has been completed (for at that point it is true to say that a thing has changed), or the first moment at which it has started changing. That which is called first in the sense of the end of change is real and exists (for a change can be accomplished, and there is the end of change, which has been proved to be indivisible on account of its being a limit); but ['the first'] in the sense of the starting point does not exist at all: for the starting point of change does not exist, nor does the first moment of time in which it [started] changing." (*Ph.* VI 5, 236a7–15).

[11] In *Ph.* VI 9, Aristotle uses the term κίνημα, in contrast with κίνησις, to refer to a single unit of change. *Ph.* VI 9, 241a2–4: οὔτε γὰρ ὁ χρόνος ἐκ τῶν νῦν οὐθ᾽ ἡ γραμμὴ ἐκ στιγμῶν οὐθ᾽ ἡ κίνησις ἐκ τῶν κινημάτων. Cf. *Ph.* VI 1, 232a7–9. This contrast is used by both [Philoponus] at 391.1–5 and by Eustratius at 161.32–162.20 (γένεσις and γένημα, 161.39).

[12] Cf. *Ph.* VI 1, cf. VI 6, 237b7–8 and VI 5, 236b11–12.

even hold of time in the sense that is required.[13] But I think this might need a qualification since time generally speaking allows for the analysis in terms of potential division so that between any two points a further point is conceivable. It is the actual division of time, shaped by the events in time, that does not satisfy the conditions of continuity that are required in order to validate the direct causal sequence from the earlier to the later events.[14]

The exact force of Aristotle's argument has been questioned by his commentators. Most ancient commentators take Aristotle's argument to be against the continuity of events in time rather than against contiguity,[15] which they did not regard as problematic. But contiguity is a weaker relation, and strictly speaking it might leave open the possibility for the contiguity of cause and effect. Barnes has made an impressive case for such a weaker relation where two events, the one a completed event A and the other an open-ended event B, can be said to be 'contiguous' in the sense that it should be possible to assume that, although the last moment of the event A (t_a) does not 'contain' any of the event B, any further moment in time t_e (however close to t_a), does contain the event B, so that there will be no 'time' intervening between the two events.[16] But it is not obvious that this weak analysis of contiguity is accepted by Aristotle. He could say that even on this analysis there is always a time *interval* between the end of A and any arbitrary point of B, however close to A. After all,

[13] Cf. Detel (1993), 730 (ad 95b1, a paraphrase of the *Physics* arguments about continuum).

[14] It may be relevant, too, that some ancient commentators (such as Themistius) note the role played by *phantasia* in deriving the concept of the continuity of events from the continuity of time (this has to do with the fact that time does have a different order of continuity compared to change and coming to be).

[15] So Themistius (54.8–15 W.), [Philoponus] (390.5–391.5), Anon. (575.27–576.15 W.), and Averroes. Averroes summarises Aristotle's argument as follows: "He means two things, assuming that the one that is in the prior moment is prior, and the one that is in the posterior moment is posterior; and it is as though he said, the reason why the posterior is not continuous with the prior is that the very last moment of the coming to be of the prior thing, i.e., when it is true to say of the first that it has come to be already, is not continuous with the moment at which it is true to say of the posterior thing that it has come to be.

But since someone might say that since both the prior and the posterior things are bodies, and bodies are divisible, and things that are divisible can be continuous with each other, as a reply, he said: and that which will be is divisible, but the thing which has already come to be is indivisible, i.e., that which is coming to be is divisible, but the moment in which it is true to say that it has already come to be is indivisible." (Averroes, *Comm. magnum*, 494B12–D7 (Abram.))

[16] Cf. also Ross ad loc.

the concept of contact between the limits may be important in indicating the kind of continuity that is relevant in the analysis of real events and causal sequence, and that may be the reason why the case is explicitly made against the contiguity of events in time, even though the key concept in the argument seems to be that of continuity.

On the basis of this ontology, Aristotle formulates his explanation of the kind of deductive reasoning that preserves its validity in the cases where cause and effect are separate in time. He says:

> (T8) Here too (i.e., in the case of events that come about in succession) the middle term and the first term must be immediate. E.g., A has come about since C has come about (C has come about later, A earlier; but C is the starting point since it is nearer to the present moment, which is the starting point of time); and C has come about if D has come about. Thus, if D has come about, A must have come about; and C is the explanation—for if D has come about, C must have come about, and if C has come about, A must have come about earlier.
>
> (*APo*. II 12, 95b14–21, trans. Barnes, modified)

The expression 'things that come about in succession' is noteworthy here, because it refers to an ordered succession where there is nothing of the same kind in-between. C is immediate with respect to D, and A is immediate with respect to C, because in the logical structure of the explanatory account, there can be no further explanatory step between C or D, nor between A and C. Aristotle illustrates this with an example of a sequence of operations in house building:[17]

> (T9) In concrete terms it is like this. If a house has come about, stones must have been cut and have come about. Why? Because a foundation must have come about if a house has come about; and if a foundation has come about, stones must have come about earlier. Again, if there will be a house, in the same way there will be stones earlier. As before, the proof is through the middle term: there will be a foundation earlier.[18]
>
> (*APo*. II 12, 95b31–37, trans. Barnes)

The validity of the oblique inference from effect to cause is grounded by this type of causal sequence. The causal sequence is given beforehand, as a part of a rich explanatory account of the effect. The direct inference from cause to effect cannot be validated in the same way (despite the fact that

[17] Cf. also parallel biological examples, most clearly in *PA* II 1, 646a25–b4; also *PA* I 1, 640a16–19; *GA* I 22, 730a5–8; II 1, 734b13–17; *EN* VII 11, 1152b12–15, cf. *EE* 1219a13–17; *Ph*. III 1, 201a15–19; b9–13; *Met*. XI 9, 1065b16–66a7.

[18] On the use of syllogistic terminology, see Barnes (1981), 38n.32.

the rich explanatory account is the same), because the intervening factors from outside the causal series can disrupt it and block the outcome. Only the completed outcome shows that there has been no interruption of the causal sequence that has preserved the order of 'succession' as defined in *Ph*. V 3. Aristotle's example of house-building should not be taken to mean that the analysis is restricted to the cases of final causality: it applies also in the cases of efficient and material causation. In this respect it may be appropriate to look at two difficulties with this analysis stated by the commentators of Aristotle, both recent and ancient.

The first difficulty is raised by Barnes, who points out in his commentary that we cannot infer 'Socrates drank hemlock' from 'Socrates died', suggesting that Aristotle thinks we can.[19] But in Aristotle's analysis the assumption (the causal statement, 'q because of p') is not derived from a given outcome (q), but is presupposed as an assumption in this further analysis in terms of implication. As Wieland has pointed out, Aristotle's methodological procedure here is not a search for causes using logic as a tool, but rather provides a means for justification, in terms of his theory of deduction, of the already discovered causal relations, thereby providing the independently established explanation with a suitable logical framework.[20] That the object of *Posterior Analytics* is not the 'logic of discovery', but rather the logic of explanation is a common point which has been argued in a more general way by a number of scholars including Barnes himself; it applies in this particular analysis as well.

The second difficulty is potentially more serious. It has to do with the question: Does Aristotle mean to deny direct necessity in causal processes? This problem has been raised by both ancient and modern commentators. Barnes has argued that it should be possible for us to say that someone who has drunk a lethal dose of hemlock will necessarily die. The same point is made in Philoponus' commentary on *GC*,[21] when he says that there are things that are not necessary by themselves but are necessarily caused by some contingent factors.

Detel has replied to Barnes's query by distinguishing between the two meanings of 'lethal': the first is 'bringing about death', in which case, he says, the term is analytic and should not be treated within the scope of causal analysis. The second is 'conducive to death', in which case there remains a room for intervening factors that may block the

[19] Barnes (1993), 235.
[20] See Wieland (1972).
[21] Philoponus *in GC* 308.13–28. Cf. p. 227 below.

causal path.[22] In fact this example could be accommodated by Aristotle's analysis of causal inference in this chapter, *APo*. II 12. Assume that the conclusion is: 'Hemlock causes death', the major premise: 'What destroys the vital organic systems causes death', the minor: 'Hemlock destroys the vital organic systems'. We can get a sound syllogism in the first figure. The next question is whether in virtue of this we can derive 'X will die' from 'X drank hemlock'. This obviously will depend on whether in this particular case hemlock destroyed X's vital organic systems, something that is not to be taken for granted, because hemlock's action could be impeded by some counteracting factors, which would disrupt the causal series and prevent the result from happening. In a 'normalised' version of our syllogism we would work on the assumption of the subject being a normal organism and of the action of hemlock being normal, without in addition assuming any interfering circumstances that might cancel its effect, such as taking an antidote etc. In the case of an individual future event we are not entitled to this assumption, so from 'X drank hemlock' we can only conclude 'X will die' if hemlock's action will not be prevented by some counteraction that does not belong to this causal series. (Think of Rasputin's murder as an example of the case where reportedly the lethal dose of potassium cyanide (KCN) did not work at all.) Thus the version of the syllogism which can claim certainty is either 'analytical' (in Detel's suggestion), or such that the middle term has the same tense-index as the predicative term, i.e., (following Aristotle's analysis) when cause and effect are treated as 'simultaneous', without a time lapse between the *explanandum* and *explanans*.

2.3. *Two Kinds of Necessity: A Distinction Indicated*

In *APo*. II 12, Aristotle does not use the concept of conditional necessity, but his discussion does have bearing on this concept as used by him elsewhere. In particular, Aristotle's discussion here indicates an important distinction that needs to be drawn between the two senses of conditional (hypothetical) necessity in the corpus. In a more familiar sense, conditional necessity refers to the kind of necessity induced in a material process by a final cause. In this sense, conditional necessity is opposed to 'simple' or 'unqualified' necessity that does not presuppose a final cause and which represents another way of referring to material causation.[23]

[22] Detel (1993), II, 738.
[23] The main texts are *PA* I 1 and *Ph*. II 9.

Conditional necessity in this case refers to the necessary conditions of the coming to be of a given natural kind. In biological texts in particular, the term 'conditional necessity' expresses the fact that a certain type of structure is 'essentially' necessary for a certain function: this is how the end 'necessitates' the means. The 'necessity' which is so imparted has to do not with the factual outcome of each individual process of coming to be: an embryo may perish prematurely, and a house may be left unfinished because of some external interference, but the coming to be or presence of the incipient proximate matter of each incomplete compound is necessitated by the final cause to no less an extent. Hypothetical necessity is construed as working backwards, *a fronte*, and contrasted with 'simple' necessity which works directly, *a tergo*. The problem that has been much discussed is whether Aristotle recognises this latter kind of necessity as operating in natural processes: there are important texts which suggest that in the realm of nature, the 'necessary cause', or matter, is always subordinate to the final cause,[24] and moreover, that 'absolute' necessity only belongs to the eternal (viz. heavenly) objects.[25] The main contrast drawn by scholars is between matter and purpose, and those who argue for the existence of simple necessity in natural processes argue for the independent causal force of matter (in this debate, 'unconditional' is understood as not being subjected to teleological conditioning).

The second context in which 'conditional necessity' occurs has to do with the question whether the cause which precedes its effect in time necessitates the outcome. The 'condition' on which this kind of conditional necessity is based is that the effect be realised. An important text is *GC* II 11, where Aristotle argues that things that come to be are divided into those that do so necessarily (i.e., come to be always) and those that are contingent (i.e., sometimes come to be and at other times do not). Simple necessity means in this case that an outcome of a given process of coming to be does not depend on any conditions at all (i.e., cannot be stopped by the interference of any external factors). In this context, Aristotle suggests that *all* sublunary necessity is conditional, while simple necessity is found only in circular processes, such as the circular motion of heavenly bodies. Sometimes the two contexts, teleological and modal, appear to overlap, as in *GC* II 11, where Aristotle illustrates the modal distinction with the example of house-building—an example that is typically invoked in the biological corpus to illustrate conditional necessity in

[24] *Ph.* II 9; cf. Cooper (1987), Johnson (2005), Sedley (2006).
[25] *PA* I 1. cf. Johnson (2005), 162ff.

the causal sense, as the necessity imparted by the final cause. As a result, the concept of 'absolute', unconditional necessity, or necessity *a tergo*, is sometimes taken to refer to all necessity of non-final origin, and Aristotle's argument against the necessity *a tergo* is taken to deny the causal force to any non-teleological factors in nature.[26]

It seems useful to distinguish these two senses of conditional necessity: first a non-modal sense, which refers to conditional necessity as operating in each process of change through several causal factors (so that for each process of change we have a final cause which defines the conditions of material causation, but also the efficient cause which will to some extent rely on matter in contributing to the outcome), and second, a modal sense, i.e., one that is based on the distinction between the necessary and the contingent. *APo*. II 12 supports this distinction, because Aristotle's analysis here follows upon an argument covering all the four types of cause (in *APo*. II 11), and is to be taken as applying to all the four cases. Thus 'necessity' is used in a different sense, i.e., as characterising the relation between the cause and effect *within* any type of causal sequence. The necessity of an individual outcome is made dependent on the validity of an inference from the earlier cause to its effect that comes about later. This discussion shows the grounds of Aristotle's treatment of simple and conditional necessity in *GC* II 11.

3. Conditional Necessity and Natural Cycles (*GC* II 11)

In *APo*. II 12 Aristotle outlines some problems which he considers in detail in *GC* II 11. The main problem discussed in *GC* II 11 is whether there is simple necessity in the coming-to-be and perishing within the cosmos. The chapter has been regarded as controversial because it seems to claim that simple necessity is found only in the heavenly revolutions, while other texts (such as *Phys*. II 9 and the biological treatises) suggest it exists also in sublunary natural processes. We have seen that the concept of simple necessity is used by Aristotle in two different senses and in two kinds of context. In this section, I consider the bearing of this difference on our understanding of the discussion in *GC* II 11, while paying special attention to the interpretations offered by ancient commentators.

[26] This interpretation surfaces in some ancient commentaries (e.g., in Philoponus, on which below), and is suggested in some modern studies (Sharples 1979).

3.1. *Some Further Problems Outlined: The End of* APo. *II 12*

Having discussed the structure of a causal sequence that validates the conclusion from effect to cause through a middle term, Aristotle turns to the case of circular sequences. In the final section of *APo*. II 12, he speaks of natural cycles:

> (T10) (i) We observe among events a sort of circular coming about. This can be the case if the middle term and the extremes follow one another: they must convert (as I proved in the beginning) because the conclusions convert; and this is what being circular is. (ii) In concrete terms it looks like this. If the earth has been soaked (A), necessarily steam came about (B); if steam, cloud (C); if cloud, water (D); and if water came about, it is necessary for the earth to have been soaked. But this was the starting-point, so that things have come round in a circle: if any item is the case, another is; if that, another; and if that, the first.
>
> (*APo*. II 12, 95b38–96a7 trans. Barnes)

The overall purpose of this turn in the discussion is not immediately obvious. In the immediate context of the argument its force is a contrast with the just discussed linear causal sequences with a definite starting and end-point. Such definite limits in a strict sense are not found in the cycles. This example raises a whole number of problems. Does Aristotle perhaps mean to relax the restriction on the inference from the earlier to the later in the case of cyclical processes? As Detel points out, Aristotle cannot be taken to suggest that there is a circular 'demonstration' of any sort available in the case of such cycles.[27] Aristotle already dealed with circular demonstrations in *APo*. I 3 and concluded there that they are generally impossible. However, he does mention one exception, namely a very special and rare case in which the three terms of the syllogism are convertible (follow upon, or are 'counter-predicated' of each other).[28] This case can stand as far as the validity of the deductive procedure is concerned; but the plausible real instances of such a relation are supposed to be virtually non extant.

In our passage (T10), Aristotle describes the terms of circular deduction in question as convertible, thus perhaps referring to the same rare exception as the one he allowed for in I 3. We should note, however, that

[27] Detel (1993), ad loc. Cf. Barnes (1975) and (1976), suggesting that Aristotle reverts to his own earlier view that circular demonstrations are possible (assuming that one of the three 'circular demonstrators' in I 3 is young Aristotle). Barnes (1976), 239, suggests that 'Aristotle toyed with the idea of representing the natural cycles of *GC* II 11 by means of circular demonstrations'.

[28] *APo*. I 3, 73a6–20; see Smith (1986).

this kind of mutual predication differs somewhat from the one licensed in I 3 in that here Aristotle is talking about the G-terms and not the B-terms as in the earlier discussion; i.e., his subject is particulars rather than essential *explanantia* and *explananda*, and this makes a difference. This difference is underscored by the fact that in the current example the particular *explanans* and *explanandum* are always separated by a time interval: air cannot be simultaneous with the water it comes from. The example of elemental transformations is thus special in that, for the elements, we do not have a B-version of deduction, i.e., qua B-terms, the elements are not convertible in a strict sense: air is not water, but becomes water as a result of transformation. This raises some problems about the ontological status of the elements that are represented by convertible G-terms.

The example of elements is adduced to illustrate the kind of deduction that might be different in form from the one discussed in 95b16–35: in these transformations it does not matter what term is chosen as the starting point of reasoning. It is worth pointing out that the force of this illustration is not made explicit by Aristotle, and perhaps we should be careful not to seek to disambiguate it too soon. In particular, we are not warranted in thinking that this case must constitute a breach of the analysis of causal deduction stated earlier. A weak reading of the example is possible if we parse the cycle into regular three-term deductions. Then every such deduction would still contain three (convertible) terms, and proceed from the starting point of reasoning (i.e., the realised effect of a given transformation) to its end-point (the starting point of this partial transformation) via the middle term. A strong reading, according to which this example would constitute a case of circular demonstration, is not suggested by the text,[29] although the problem is clearly flagged, and the possibility is not ruled out. It seems that in this chapter the example is cited to highlight the problem, and that Aristotle deals with it elsewhere, namely in *GC* II 11. His remarks that follow upon this discussion in *APo*. II 12 sound rather general; his wording may well be deliberately chosen so as to leave room for a not-too-strong reading of the illustration. Aristotle says:

> (T11) Some things come about universally (they either are or come about in this way always and in every case), others not always but for the most

[29] As Detel (1993) points out, Aristotle here avoids the use of the terminology of 'demonstration'.

> part—e.g. not every male man has hair on his chin, but they do for the most part. In such cases the middle term must also hold for the most part.
>
> (*APo*. II 12, 96a7–12 trans. Barnes)

Notably, there is no example of things that come about in this way always (but it will come up in *GC* II 11). The claim that the middle term holds for the most part seems to involve some sort of a weighted quantification procedure over the individual causal sequences considered in this chapter, i.e. those with uncertain outcome. It is unclear whether Aristotle has in mind any particular mechanism of quantification.[30] What seems important, particularly in the light of this example, is that the operation of 'simple necessity' understood in the sense of material causality has its way 'for the most part', i.e. not in virtue of 'simple necessity' understood as an absolute necessity of an outcome. This example again indicates the distinction between the two senses of necessity. With this in hand, we can look at the argument of the *GC* II 11.

3.2. *Necessity in* GC *II 11: Aristotle and His Ancient Commentators*

The main problem discussed by Aristotle in *GC* II 11 is whether anything comes to be of necessity or whether everything that comes to be is contingent. Already its first formulation shows that it arises from Aristotle's interest in the natural cycles. The logical structure of these cycles is of the kind described at the end of *APo*. II 12. The main argument of the *GC* chapter also displays a number of parallels with the argument of the *Analytics*: the question Aristotle asks and goes on to discuss is whether we are justified in deriving the necessity of the effect (or a later outcome) from the existence of the cause if the cause (or an earlier event) is regarded as a necessary condition of this effect (later event):

> (T12) (1) Granted that the coming to be of something earlier is necessary if a later thing is to be, e.g. if a house, then foundations, and if foundations, then clay, does it follow that if there have come to be foundations a house must necessarily come to be?
>
> (2) Or can we not yet say this, unless it is necessary simpliciter that the latter itself come to be? In this case, if foundations have come to be, it is also necessary that a house come to be; for such was the relationship of the earlier thing to the later, namely, that if there is to be the latter, necessarily there will be the former, earlier thing.

[30] For discussion of the meaning of 'for the most part', see Mignucci (1981).

> (3) If, accordingly, it is necessary for the later one to come to be, it is necessary also for the earlier one, and if the earlier one comes to be, it is accordingly necessary for the later one to do so—but not because of the earlier one, but because it was assumed that it was necessary it should exist. So in those cases where it is necessary for the later one to exist, there is conversion, and it is always necessary, if the earlier has come to be, that the later should also come to be. (*GC* II 11, 337b9–25, trans. Williams)

The example used in (1) is the same as in *APo*. II 12 95b32–37; notably, it is not an example of a cycle, but of a linear causal sequence which grounded the oblique inference from effect to cause. Aristotle approaches the modal problem of necessity by re-stating the question resolved in the *Analytics*: given that (i) $q \rightarrow p$ and (ii) p, does it follow that (iii) q comes to be of necessity? The necessity mentioned here does not need to be rendered as a modality of p: the question is about the logical form of inference; its solution does not depend of whether we prefix the operator of necessity to (i) and (iii).

In (2) Aristotle sketches a solution which is then stated in a somewhat different form in (3). The solution involves postulating the necessity of the later outcome of an earlier event. Formalised proof would look like this:

(i) $\Box q$
(ii) $q \rightarrow p$,

Hence,

(Δ) $p \rightarrow q$ (supplying steps (iii) $\Box q \rightarrow q$; (iv) q ((i),(iii)); (v) p ((ii), (iv))).

The necessity of inference in the conclusion is not modal, but deductive: it is shown that it is properly derived from the assumptions. In other words, the deduction (i) q; (ii) $q \rightarrow p$; (Δ) $p \rightarrow q$ would also be valid; and a more general case (i) q^*; (ii) $q^* \rightarrow p^*$; (Δ) $p^* \rightarrow q^*$, where the asterisked letters stand for metavariables running over modal and non-modal atomic sentences.

In (3) the assumptions seem to be, from generalised version:

(i) $\Box q$
(ii) $\Box q \rightarrow \Box p$;

Hence,

(Δ) $\Box p \rightarrow \Box q$.

The conclusion says that in the processes in which the later is necessary, the earlier is also necessary, not by itself but by conditional necessity

operating *a fronte*, as a condition of an outcome which is necessary by itself. From arguments (2) and (3) it is clear that conditional necessity represented in two conclusions has a deductive nature, even though in (3) both the antecedent and the consequent have modal operators.

But there is also a question about the nature of the simple necessity of the outcome, postulated as the first premise (□q) in both arguments (2) and (3). Aristotle in fact goes on to show that such necessity can accrue only to the things whose coming to be has a cyclical pattern, and of these only heavenly bodies possess the simple necessity of coming to be because only this type of cycles involves continual repetition of the same pattern with the same individuals. The concept of simple necessity referred to here clearly differs from the concept used to describe material necessity operating in natural processes. Here Aristotle speaks of the kind of natural necessity that could validate the deduction of an outcome from its antecedent, as in □p → p, only stronger, validating an conditional of the form: □p → q. Most processes do not have this kind of necessity; only the phenomena of heavenly rotation have guaranteed outcomes for any earlier state of affairs.

Aristotle's argument for circularity is constructed as a refutation of other options, namely finite generation and infinite generation in a straight line.

The former option is ruled out because it lacks one essential attribute of necessity—the provision of eternity. Aristotle illustrates this case with house-building example which previously served to illustrate the necessity of the oblique inference from effect to cause when the effect has been realised (*GC* II 11, 337b29–33). Here Aristotle emphasises that the effect does not possess simple necessity. The reason he gives is that this kind of effect does not possess eternity: eternity is a required attribute of necessity.[31] The notion of eternity he uses deserves some attention. This is not just a provision that the number of future instances of house-building is infinite, or that house-building as an activity will never cease—as it probably will not, on Aristotle's view of the permanence of the human species. It includes a stronger requirement of continuous recurrence in the same order, so that the eternity should be realised in a continuous series of individual processes of coming-to-be which follow upon each other. Such are processes in nature, where there can be no gap between

[31] τὸ γὰρ ἐξ ἀνάγκης καὶ ἀεὶ ἅμα (337b35). As Philoponus' commentary says, ἅμα is to be understood as 'the same' (310.32). This is the emphasis of the discussion of simple necessity, rather than its lacking any conditions at all (so Sharples (1979), 32).

the two comings-to-be, in the sense that each one presupposes another one upon which it follows and provides a condition for its possible successors in the series.[32]

The next part of the argument deals with the second option, proving that simple necessity cannot be realised in an infinite rectilinear series, i.e. open series, where each member is followed by a different one without significant repetitions. Such series, being unlimited, would seem to be able to accommodate the eternity or permanence required by simple necessity. But this is not the only condition of necessity in a required sense. Aristotle is looking for a kind of structure that would satisfy the conversion, whereby $q^* \rightarrow p^*$ would entail $p^* \rightarrow q^*$. This would be allowed by the case where q is necessary *per se*, i.e. q (or $\Box q$, depending on the version) could be added to the set of assumptions, giving $\{q^* \rightarrow p^*, q^*\}$, to derive $p^* \rightarrow q^*$. Such a structure is not available in an infinite rectilinear (i.e. non-recurrent) series. The reason is that in the open series q will itself depend of some member different from the previous sequence, e.g. r^*, such that $r^* \rightarrow q^*$, that on some further, e.g. s^*, $s^* \rightarrow r^*$, etc. So to get the required derivation we would have to include in our set of premises all those further antecedents: this would make it infinite. As Aristotle says,

> (T13) If, then, it proceeds to infinity downwards, it will not be necessary *simpliciter* for this (one of the later ones) to come to be, but only conditionally; for there will always have to be some further thing in front of it on account of which it is necessary for it to come to be; so, given that the infinite has no principle, there will be no first member on account of which it will be necessary for it to come to be.[33]
>
> (*GC* II 11, 337b25–28, trans. Williams)

It has to be pointed out that this kind of infinite structure does not preclude the existence of conditional necessity operating *a fronte*. It is perfectly legitimate to take a completed section of this series which begins with a cause and ends in its effect, and conclude from the effect to the (earlier) cause. But this will not give us simple necessity in the sense required by Aristotle, because this derivation will account only for this individual case, but not for the permanence of this outcome in a continuous series.

---

[32] This is how the scope of investigation has been described at the start of the chapter: τό ἐφεξῆς ὄν καὶ γινόμενον τόδε μετὰ τόδε ὥστε μὴ διαλείπειν (337a34–b1).

[33] Εἰ μὲν οὖν εἰς ἄπειρον εἶσιν ἐπὶ τὸ κάτω, οὐκ ἔσται ἀνάγκη τῶν ὑστέρων τόδε γενέσθαι ἁπλῶς, ἀλλ'* ἐξ ὑποθέσεως· ἀεὶ γὰρ ἕτερον ἔμπροσθεν ἀνάγκη ἔσται, δι' ὃ ἐκεῖνο ἀνάγκη γενέσθαι. Ὥστ' εἰ μή ἐστιν ἀρχὴ τοῦ ἀπείρου, οὐδὲ πρῶτον ἔσται οὐδὲν δι' ὃ ἀναγκαῖον ἔσται γενέσθαι. * ἀλλ' EWF^{1}J^1 Alexander (presumably on the basis of *Quaest*. II 22, III 5) ἀλλ' οὐδ LHV Philop^c ἀλλ' οὐκ M. Cf. p. 225n39 below.

This part of Aristotle's proof was found difficult by commentators, ancient and modern. Alexander of Aphrodisias apparently gave two explanations of the arguments which were known to his school. *Quaest*. II 22 has a title 'Explanation of a passage from the second book of Aristotle's *On Coming-to-Be and Passing Away*, differing from that in the commentary on it'.[34] In this treatise, Alexander reconstructs Aristotle's argument as follows:

> (T14) If [i] it were shown that it is only ⟨in⟩ those cases where what is later is of necessity that 'if what is first, of necessity what is later' is true, and [ii] in the case of coming-to-be to infinity that which is later is not of necessity, because it does not even come to be the same; [then] [iii] neither would anything before that which came to be last come to be of necessity without qualification, [iv] because the things that precede the final [member of the series] only derive necessity without qualification from the end if this comes to be of necessity without qualification. (71.9–15, trans. Sharples)

In [i] Alexander seems to give a modal version of the main result of *APo*. II 12, which was that an inference from the earlier cause to the later outcome is valid only when the outcome has occurred. There we had $p \rightarrow q$ true if (i) q because p; (ii) q, via (iii) $q \rightarrow p$. Here we have $\Box(p \rightarrow q)$ if (i) q presupposes p; (ii) $\Box q$; via something like (iii) $\Box q \rightarrow \Box p$. Here it is the validity of a modal inference that is in question, hence the requirement that the later outcome be necessary, not just true.[35]

In [ii], Alexander apparently is trying to capture the lack of regularity in an infinite linear series. It is not entirely clear from our text what he means by 'never coming to be the same'. The possibilities include: (a) a series with no type or token repetitions at all: this will make the point, but is too strong; (b) a series with type but not token repetitions,[36] with two versions: (b') type-regular sequence constituted by different tokens; (b") a series of random type-repetitions instantiated by different tokens. Now, (b') can hardly be Alexander's choice: in *Quaest*. III 5, as we shall see shortly, he considers this to be a weakened variant of cyclical pattern; this would not be a clear example of rectilinear infinite series. (a) would be a clear refutation, perhaps too strong, but conceivable, particularly given that Aristotle himself uses something like this in his proofs of circular character of elemental transformations in *GC* II 5, where he uses different letter characters for different elements to illustrate the (refuted) hypothesis of linear infinite change (see Appendix 2 and 3). But in fact (b") would

[34] 71.3–5 Bruns; trans. Sharples.
[35] On Alexander's commentary on *APo*. in general, see (Moraux) 1979.
[36] Along the lines of Sharples's suggestion, Sharples (1994), n. 123.

suffice to make the required point, namely, that in this sequence later outcomes do not possess necessity. The reason is precisely the random character of the series in question which does not allow us to assume a regular completion of each 'minimal' causal sequence ⟨p; q⟩. In *APo*. II 12, the inference from the earlier cause to its later effect is not logically true because there can be an instance when the antecedent representing cause is true and the consequent representing the effect false. In this case, the inference will not be necessary even if its one non-modal instance is true because the necessity requires that it be true always.[37]

The conclusion [iii] is that there is no simple necessity in this kind of infinite series. A couple of lines down, Alexander explains that later outcomes (the right hand-side members of successive causal pairs ⟨p, q⟩) in such series can possess conditional necessity insofar as they themselves are necessary conditions of the outcomes already completed. Here Alexander uses the notion of simple and conditional necessity in the logical sense, borne out by the analysis of 'causal' deduction in the *Analytics* chapter. The thesis he attributes to Aristotle is that direct necessity from cause to effect does not obtain except in the circular processes for logical reasons. That this is his interpretation of the main claim of *GC* II 11 is clear also from his analysis in *Quaest*. III 5, which does not differ in its main points from *Quaest*. II 22.

The thesis attributed to Aristotle at several points in Philoponus' commentary seems to be rather different. The composition of Philoponus' *GC* commentary is complex and the authorship of different arguments possibly varies. *In GC* is considered to be one of the earliest writings by Philoponus; technically, it is a set of notes taken in Ammonius' seminars. Ammonius, himself an original thinker, makes ample use of the lost commentary by Alexander, which Philoponus also might have consulted when writing his notes. Philoponus sometimes registers disagreement with his master's argument and adds his own comments to his reports of Ammonius' discussions.[38] The commentary on *GC* II 11 is an example of such a 'layered' discussion: we have the main interpretation which must have been stated by Ammonius, and a number of brief interjections which could be by Philoponus himself. The main interpretation is given twice, with minor differences: in the *theôria* which forms a preface to the

[37] One could draw a parallel with a modern semantic definition of necessity in modal logic as truth in all possible worlds; of course, Aristotle's 'possible worlds' have rather unusual constraints, but the basic type of justification seems to be similar: in a modern case, truth in all possible worlds; for Aristotle, the infinite set of convertible instances.

[38] For a more detailed survey of the composition of this commentary see Kupreeva (2005), 1–16.

whole chapter and in the discussion of the text of Aristotle's argument. The author argues that neither simple nor conditional necessity is possible in the infinite linear regress:

> (T15) For if (a) necessity *simpliciter* is this: when upon that which is first, when it [already] exists, what is posterior of necessity follows, having necessity due to its own nature, and [if] (b) it is not possible to assume the first and the posterior in the infinite, it is manifest that (c) in this case there will be no necessity *simpliciter*. (d) But nor [will there be necessity] *ex hypothesi*, as he says,[39] which was: 'if what is posterior, then of necessity also what is prior'. If, therefore, there is no prior and posterior in an infinite straight line, then not just necessity *simpliciter* is not there in it, but nor is there necessity *ex hypothesi*, for the same reason.[40] (309.10–19)

The argument is that because there is no first and last member in an infinite series as a whole, it is impossible to speak of a relation of prior and posterior in a proper sense, and therefore the definitions of conditional and simple necessity (understood here as necessity operating between the earlier and the later) do not apply. It is not clear to what extent this argument depends exclusively on the crucial reading of line 337b26 in Aristotle's text that was accessible to the source of this argument in Philoponus' commentary (perhaps Ammonius). The *theôria* version of the argument contains an additional reason in support of the impossibility of conditional necessity:

> (T16) For it is generally impossible that anything could have come to be in the infinity. For each of the things assumed as having [so] come to be has an infinite distance from the beginning, such that it would have been impossible for this thing to traverse it and end up at the point at which it would have come to be. (304.25–28 = (ad)–(ae) in Appendix 1 below)

This reasoning shows a number of parallels with the discussion of Aristotle's proof of the cyclical pattern of elemental transformations in the commentary on *GC* II 5 (see Appendix 2, 3). The point of drawing this parallel would be to deny the linear, non-cyclical pattern in the natural processes. This point seems to be picked up in the objection to the described interpretation of Aristotle's argument against infinite series stated by Philoponus:

---

[39] This shows that the text of 337b26 that the author has is ἀλλ' οὐδ', cf. p. 222n33 above.

[40] This is the version in the comment ad loc. The version in the preface to chapter contains separate arguments for the 'ascending' and 'descending' regress, showing, respectively, that there can be no first member among the antecedents and no last one among later outcomes, with the same conclusions about the impossibility of both simple and conditional necessity in an infinite straight line (see Appendix 1).

> (T17) [309, 20] But we should realise that the argument like this apparently is not well formed to reach its goal.[41] For if (i) someone wanted to eliminate an infinite straight line and said that nothing can come to be in an infinite straight line, since everything that comes to be comes to be for the sake of some end, but every infinite thing is[42] without an end, he would say something that is both true and irrefutable (for there can be no infinite straight line at all). But if (ii) he assumes an infinite straight line, and since the infinite is without a beginning and without a limit, he says that on it, it is impossible for the second to follow upon the first of necessity, his argument will not be plausible. (iii) For although there is no beginning and no end, there still is succession and the coming to be of this after this. In this way, at any rate, Aristotle while saying that time has neither beginning nor end still says that succession and the first and the second are observed in it.[43] So, as far as Aristotle's claims are concerned, necessity will not be eliminated. (309.20–31)

The authorship of this objection is not entirely clear; but Philoponus himself is a very likely candidate. This objection is a one-off criticism of the weak points of the twice-repeated main argument, and does not seem to start a new line of argument.[44] Philoponus endorses the view that no process in the world can have an infinite regress as its model because of the teleological causality by virtue of which there must be an end to each process. But he points out that once the infinite straight line is assumed (perhaps for the purposes of analysis), we can speak of things being prior and posterior in a linear order, and earlier and later in time (the time represented in its totality by an infinite straight line). He shows no awareness at all of the analysis of the kind given by Alexander, nor of a different reading of Aristotle's text.[45] It is possible that he took

[41] 309.20: μὴ κατωρθωμένον: i.e. there is an ambiguity in the statement of the argument leaving open the possibility of a reading on which the argument is either not valid or not sound. Cf. Philop. *In GC* 98.20 where κατορθοῦν is opposed to ἀμφιβάλλεσθαι (cf. Vitelli's *Index verborum* s.vv.).

[42] 309.23, reading ἔστιν instead of εἶναι as Vit. suggests in apparatus.

[43] 309.28, 30: *Ph.* IV 11 (on definition of time); *Ph.* VIII 1, VIII 8 (on time being infinitely divisible).

[44] This is characteristic of several departures from the main line of the argument in the commentary; in Kupreeva (2005) 161n473, I suggest that these might be Philoponus' ἰδίαι ἐπιστάσεις, against Sharples (1994), 121n120 who treats the whole passage as a part of the 'mainstream' argument which he attributes to Philoponus. I believe that tracking the difference between this objection (T17) and the preceding argument (T16), apart from its being in order for mere pedantic reasons, can allow us to trace a common position behind some of these objections.

[45] There is a tantalising question of the role of Alexander's commentary in Philoponus' commentary on this chapter. From reference at 314.10 it is clear that it has been consulted in some way, either by Ammonius or by Philoponus, but it is unclear what different

the concept of conditional necessity used by Aristotle in the sense of hypothetical necessity imparted to the process by its final cause. In the commentary on 337b22–23, after the explanation of Aristotle's thesis that direct necessity between the earlier event and its later outcome obtains only if the outcome is necessary simpliciter, we find a long objection based on the examples of natural processes whose outcomes, the author says, are necessary, even though they are not necessary simpliciter:

> (T18) But it will seem that many facts are in conflict with Aristotle's arguments. For we see that many among natural things, generable and perishable, of those that necessarily follow upon certain things that have preceded them, have the necessity of coming to be not because of themselves but because of the things that have preceded them, e.g. upon starvation, emaciation follows of necessity, not because it by itself has necessary coming to be, but because of the starvation that has preceded [it]. For emaciation does not come to be out of necessity by itself, but when starvation has preceded. In the same way, too, when much has been eaten, beyond the capacity to digest, slow digestion will follow, not because of itself, and after a blow [inflicted] upon a fleshy part a weal will follow, not because of itself, but because of the blow, and after water has been poured over the earth mud will follow out of necessity, not because of itself. And there are many other [cases] where the second follows upon the first of necessity, not because of itself, but because of the first.
>
> And neither is this always convertible: for it is not the case that if [there is] slow digestion, then always too much has been eaten, but this happens to come about also because of worries, insomnia and other causes; and upon killing death follows of necessity, however, it is not the case that if there is death, killing has also preceded. (308.13–28)

These objections have been considered in the first part of the paper (see p. 213n21 above). Underlying is the view of simple necessity as operating independently from any teleological factors in a narrow sense, not the simple necessity in a logically precise sense required by the analysis of causal nexus in *APo*. II 12 and *GC* II 12. There is no reply to these objections in the commentary. Moreover, as we have seen, the objections seem to persist in modern scholarship. And even though they can be partly answered by this clarification of Aristotle's project in these two texts, there still remains a question about the relation between the modal simple necessity in logic and the necessary patterns in natural processes.

---

interpretation of Aristotle's argument it offered (cf. p. 223n34 above) and whether it influenced the 'main' argument of Ammonius/Philoponus.

3.3. *Necessity in Natural Cycles?*

As we have seen, in *APo*. II 12 Aristotle uses the example of elemental transformations as a tentative illustration of the type of counter-predication among the terms that is required to validate the deduction of the later outcome from the earlier cause. In *GC*, we find a discussion of all the three main types of cycle found in Aristotle's physical corpus: elemental transformations, biological reproduction, and heavenly rotation. Only one type (heavenly rotation) is shown to be an adequate model for the deduction of the later outcome from the earlier state of affairs.

The cyclical pattern is essential for elemental transformations. Earlier in *GC* II, Aristotle provides a detailed argument which involves a refutation of a 'linear infinite' model for these transformations. In *GC* II 2, Aristotle constructs the notion of an element ('simple body') that is constituted by a pair of elemental qualities, so as to arrange for mutual transformations between the elements based on the exchange of elemental qualities (which are two pairs of contraries).[46] In *GC* II 5, he describes three different paths of transformation, all based on the same principle of prevalence: two opposite qualities, hot and cold, constituent of two different elements, exercise their action upon each other: the greater intensity of action (which depends on the mass of respective elements) determines the direction of transformation.[47] So, in the case of elemental transformations, the role of cyclical pattern is not to establish a definitive path matching the logical concept of necessity. On the contrary, contingency of the outcome of each elemental 'reaction' seems to be an important result of this proof. Nonetheless, in *GC* II 11, a particular type of elemental transformation is considered as a candidate for some sort of sublunary necessity:

[46] This arrangement is not equivalent to a *demonstration* of each element from every other one: the commentators understand this correctly when they emphasise that the elements are not the elements of each other, but of all other things, cf. Philop. *In GC* II 5, 243.18.

[47] (1) The fastest one follows the order of subsequent transmission of the active σύμβολον; when less fire is overridden by more air, and the resulting air is overcome by more water, and the resulting water is overcome by more earth: (hot+dry) + (hot+moist) + (cold+moist) + (cold+dry)—⟩ (cold+dry). (2) The second way consists in the simultaneous transformation of the two qualities into their opposites: hot+dry ⇔ cold+moist; hot+moist ⇔ cold+dry. (3) The third way consists in the transformation of the two non-neighbouring elements into the third one by removing one quality in each of the initial agents: (hot+dry) + (cold+moist) ⇔ (dry+cold) + hot + moist.

(T19) (1) If that which is moved in a circle moves something continually, the movement of these things must also be in a circle. For example, the locomotion above it being in a circle, the sun moves in this way, and since it moves in that way, the seasons because of it come to be in a circle and return upon themselves, and since these come to be in this way, the things affected by them do so in their turn.

(2) Some things, then, are obviously like this; water and air, for instance, come to be in a circle, and if there is a cloud it is bound to rain and if it rains there is bound also to be a cloud. Men and animals, on the other hand, do not return on themselves in such a way that the same one comes to be again (since there was no necessity, given that your father came to be, that you should have come to be, only that he should have, given that you did), and it seems that this coming to be is in a straight line.

(3) Why is there this difference? This again is where the investigation begins: do all things return on themselves in the same way, or not, but rather some in number and some only in form? It is obvious that those whose substance, i.e. what is moved, is imperishable will be the same in number, since movement follows the thing moved, but those whose substance is, on the contrary, perishable, must necessarily return on themselves in form, not in number. That is why water from air and air from water is the same in form not in number; but if these too are the same in number, still they are not things whose substance comes to be, the sort, namely, that is capable of not being.

(*GC* II 11, 338b1–20 transl. Williams)

Having drawn a distinction between the things that move in a circle and the ones whose coming to be is in a straight line (1) Aristotle cites the elements as an example of the former class versus biological species whose reproduction, it is suggested, follows a linear, non-recurrent pattern (2). In (3), he introduces a further distinction, between the individual recurrence and the recurrence of species. In the light of this distinction, it seems, both elemental and biological cycles can be considered in two ways, namely, at the level of species and at the level of individuals. However, Aristotle does not say here that the recurrence of species in biological reproduction has necessity. This generalisation is proposed by Alexander of Aphrodisias in *Quaest*. III 5.

(T20) (a) Someone might raise the difficulty whether the consequence 'if what is first, [then] also what is later' is true of the things that come to be in a cycle and return again. Well, this consequence is true of the things that are brought about in a determinate fashion by the bodies that move in a circle. If the winter solstice, [then] also the [spring] equinox, and if the [spring] equinox, [then] also the summer solstice, and if the summer solstice, [then] also the [autumn] equinox, and if this, the winter solstice; and ⟨also⟩, if winter, [then] also spring, and if spring, summer, and if sum-

> mer, autumn, and if autumn, winter again. And in the case of the things mentioned first [i.e. the solstices and equinoxes] the ordered sequence is determinate and permanent and is never retarded or advanced, because the sole cause of the being and ordered sequence of these things is the movement of the primary [i.e. heavenly] bodies, nothing else contributing towards it; and for this reason it is possible in the case of these things to determine the time, too, and say [not only that they will be of necessity, but] when they will be of necessity.
>
> (b) But summers and autumns and winters no longer possess determinacy in a similar way, although they come to be in a cycle and they too themselves follow on the motion of the eternal [bodies], because matter too contributes to their coming to be, being affected by the movements of [the eternal bodies]; and since [matter] does not in every respect, in the way in which it is affected, follow the movements and revolutions of [the eternal bodies] in a similar way, [for this reason the seasons] are not determined in the individual details of the way in which and time at which they come to be] in the same way [as are the solstices and equinoxes].
>
> (c) And indeterminacy is still more [present] in those things that need more things to contribute to their being; and among these is the coming to be of living creatures. And for this reason it is true of them, [speaking] generally, that each of them is everlasting as regards the species (and the cause of this [eternity] is the revolution of the divine [bodies]), but [as for] the coming to be of individuals, in the case of which the cause from the proximate efficient [causes] has the greatest influence, of these 'if what is first, [then] of necessity what is later' is not true, but 'if what is later, [then] of necessity what precedes it' is true. (89.2–24, trans. Sharples)

Several points made by Alexander are important for understanding the force of Aristotle's analysis of necessity. Alexander combines in one picture the causal role played in the generation and corruption by heavenly bodies and proximate sublunary causes. The somewhat paradoxical result of this combination is that in the grand scheme of things, on the ladder of being (from the sublunary to heavenly processes) the increasing role of necessitation *a tergo* is accompanied by the diminishing causal, explanatory link between the successive processes in question. The only pure case of such necessity properly speaking is that of heavenly rotations. But different stations of heavenly bodies do not cause each other *qua* individual events, they are only co-ordinated with each other, and that coordination possesses necessity. As Alexander says, solstices and equinoxes do not cause each other, but have one, to a certain extent common, cause.[48] The greater the causal import of proximate causes, the less the 'necessitation'

[48] Notably, Averroes makes this remark in his commentary on *APo*. II 12 in his discussion of the last part of the chapter, to do with universal coming to be.

between the elements of a series. Necessity and causation seem to get an epistemological reconciliation at the expense of the ontological split. The recurrence at lower levels (seasons corresponding to the level of elements (b), and biological reproduction (c)) is less determinate, although the role of causal nexus between the processes in one series is much more tangible. We may recall at this point Aristotle's claim in *APo*. II 12 that in some cases the middle term holds 'for the most part'.

Further, according to Alexander, matter is the factor of indeterminacy in the sublunary realm, i.e. matter is responsible for the fact that the regular pattern of recurrence can be violated. So, the necessity induced by material causation is shown to have a source distinct from that of logical necessity. The regularity 'for the most part' is thus a result of a combined action of two kinds of necessity, and this result by itself does not possess necessity strictly speaking in either the first or the second sense. One can see the metaphysical significance of this result, both in Alexander's own discussion of the problems of determinism and in the later Neoplatonic discussions of the hierarchy of principles. It is important to note once again that the starting point and the outcome of Alexander's discussion are closely related to the problem Aristotle discussed by Aristotle in *APo*. II 12 and the method of that discussion.

Appendix

1. Philoponus *In GC* II 11, 304.16–32 (second version of reconstruction of Aristotle's argument)

> (i) To [things] that move in an infinite straight line necessity *simpliciter* will not belong because we say that necessity *simpliciter* is where the second follows upon the first of necessity,[49] but in the case of infinity, there is no prior and posterior,[50] for it is without a beginning and without a limit.[51] For (a) it is not possible to assume necessity *simpliciter* in the case of

[49] 304.18–19: p→□q, but from the argument (ia) below it is clear that the intended logical form of this is □(p→q), because the necessity *simpliciter* is supposed to be imparted by *p* to *q*. Aristotle's idea of simple necessity does not necessarily include the condition of its being imparted. But in minimal modal logic, the claim □ (p →q) → (p → □q) is not valid, and in order to get □q from valid □ (p→q) →(□p →□q) we need to assume □p.

[50] 304.19: τὸ πρῶτον καὶ τὸ ὕστερον, can be taken in either temporal or causal sense.

[51] 304.20: ἄναρχον γὰρ καὶ ἀπεράτωτον: καί is treated by our commentator here not as epexegetic, so that ἀρχή is understood as πέρας (as it is correctly explained by Alexander in his exegesis of 337b27–29, Alexander, *Quaest*. II 22, 71.23 Br., cf. Joachim ad loc., Sharples (1979), 37), but as a real conjunction.

past events, so as to say that *because* of the first that which came to be second followed: (aa) for it is totally impossible to assume the beginning and the first of infinity.[52] (ab) But if it is not possible to assume the first, the second will not be there either. (ac) Hence, nor will there be necessity *ex hypothesi*.[53] (ad) For it is generally impossible that anything could have come to be in the infinity. (ae) For each of the things assumed as having [so] come to be has an infinite distance from the beginning, such that it would have been impossible for this thing to traverse it and end up at the point at which it would have come to be.[54] This, then, is the case with past events. In a similar way, (b) in the case of future [events] necessity *simpliciter* is impossible in infinity: for there is no 'posterior' in infinity.[55] Hence it will not be possible to say that if this came to be, that which is posterior will follow. Thus, it is clear from this that necessity *simpliciter* does not belong to things that move along a straight line in infinity.

2. Aristotle, *GC* II 5, 332b30–333a13

That it is not possible to proceed to infinity … can be shown from the following. If the next move of F, i.e. fire, is to change into something else (and not turn back), e.g. into X, there will be a contrariety between fire and X other than those mentioned, because by assumption X is identical with none of the group EWAF. Let G belong to F, B to X. G will belong to all the group EWAF, since they all change into one another (this, however, ought not at this stage to be taken as proved). But so much, at least, is clear: if X in its turn is to change into something else, another contrariety will belong to both X and F, i.e. fire. Equally, it will always be the case that as a new member is added to the series a contrariety attaches to the previous members, so that if the series goes on to infinity the number of contrarieties which attach to a single member will also be infinite.

In this case it will not be possible for anything either to be defined or to come to be; (i) for, if it is to be one from another, it will be necessary for that many contrarieties to be gone through, and still more. So there will be some things into which there will never be change. (ii) This will happen if the number of intermediate stages is infinite, and this will necessarily be the case if the elements are infinite in number. (iii) Again, there will be no change from air into fire if there are infinitely many contrarieties. (iv) Furthermore, everything will come to be one.

[52] 304.22–23.

[53] 304.23–24: this possibly reflects Aristotle text at 337b26–27; Bruns, followed by Sharples, takes this to be an evidence that Philoponus is committed to the view that there can be no hypothetical necessity on an infinite line, but this may in fact be a view of his source (see p. 222n33, p. 225n39 above).

[54] The argument (ae) differs from the preceding one in understanding infinity as infinite divisibility. Cf. the argument in *GC* II 5 against the view that elements could be infinite in number in the following text (Appendix 2).

[55] 304.29.

(This proof presupposes that by 'change' we understand the change of any arbitrary element into any arbitrary one. The methodological role of these proofs is to back up the selection procedure by which the four elements were established in *GC* II 3, showing the advantages of this system and the disadvantages of the alternative systems. This point is underscored by Philoponus who argues that any change among the elements will be impossible if change is taken to be an asymmetric relation.)

3. Philoponus *In GC* 257.7–13

But it is best to understand this argument in this way, as elaborating the theorem *per se*, in its own nature, rather than as continuous with the preceding argument. (a) For if you say that (i) F changes into X and (ii) X neither turns back to F, nor changes [into it] then (iii) X has no contrariety in relation to F. (iv) For had it had one, X would have been liable to change into F; (v) it follows that it will be the same [as F]. In this way going through the sequence [of elements], you will prove that all [the elements] are the same with one another. (b) But if someone says, 'I have assumed this very thing in the beginning, that these, F and X, have contrariety in relation to one another, just as I assumed that F changes into X', I will say to him: 'This very contrariety that you assumed you yourself are destroying when you say that X no longer changes into F. And in this way, in general, having assumed that the [elements] above do not change into the ones below you destroy their contrariety which you assumed in the change of every one in relation to the following one. And the contrariety having been destroyed, all will be the same with each other, as he has reasonably concluded.'

(The two assumptions attributed to the proponents of linear change are: (i) F changes into X; (ii) X does not change into F. From this, our author concludes (iii) X does not have a contrariety with F (conclusion explicated in (iv)). (v) follows from the fact that X has no contrariety with F. The author seems to assume that relation 'has contrariety with' is asymmetric either, so that while F may have contrariety with X, X does not have contrariety with F, and therefore may be considered as 'the same as' F. 'Is the same as' thus also may lose its symmetry in this interpretation (although this is never explicitly stated).)