Aristotle’s psychology – what he calls the “study of the soul” (ἡ ἑττ ὁς ψυχῆς ιστορία) – occupies a prominent place both in his own philosophy and in the Western philosophical tradition as a whole. In his own system, psychology is the culmination of metaphysics and natural science. For Aristotle, living things are the paradigm of natural objects and substances in general, and so offer the best case for the application of his theories. Psychology also serves as a foundation for the rest of his philosophy, in so far as it provides a framework for understanding thought, speech, and action, and to various extents his logic, rhetoric, politics, and ethics all draw on these views. Its influence on subsequent philosophers has also been great. During the Roman empire, later Platonists appropriated many of his doctrines, especially those concerned with cognition. In medieval scholasticism, they become part of a common legacy, shared widely by philosophers who differ in many other respects. They thus belong to the backdrop against which early modern philosophers attempted to distinguish themselves, a point of reference and a point of departure to which some twentieth century philosophers sought to return, in their efforts to shake free of Descartes’ grip. Many still regard Aristotle’s theory as offering an attractive middle course, which avoids the extremes of both extravagant dualism and crude materialism. With so much at stake, it is not surprising that there has been heated controversy about Aristotle’s psychology in recent years, as regards both its precise nature and its viability.

Some misunderstandings can be avoided at the outset by being careful about the words “soul” and “mind.” We tend to associate the first with religious traditions, in which the soul is regarded as a separate substance, divinely implanted in us and capable of immortal life. The second term, “mind,” is often construed in a similarly dualistic way, although it owes more to Descartes than to religion. Here the assumption is that “the mental” and “the physical” are mutually exclusive, where the mental is characterized as ineradicably private and subjective. Neither of these preconceptions is at work in Aristotle. In inquiring after the nature of the soul or ψυχῆ, Aristotle is looking for nothing other than the basis or ἄρχη of life – that in virtue of which living things are alive and behave in the ways distinctive of living things, whatever that should turn out to be. He does not prejudge the question of whether this is a separate immaterial substance (as Plato thought), or a material one (as Democritus thought), or something else altogether. For all we know at the outset, the soul might be inseparable from the body, or alternatively, it might be capable of separate existence. Such conclusions are to be established by argument, based on what it is to be alive and the characteristics that distinguish living from nonliving things. The fact that this is
Aristotle’s starting point, and not questions about skepticism and certainty, explains why he is not preoccupied with many of the issues that preoccupy Descartes.

The soul, then, is that by which we perceive, feel, think and act, since these are all activities peculiar to living things. To this extent, Aristotle’s use of “soul” is quite similar to our use of “mind.” We commonly say that we perform these activities “with our minds,” without implying anything Cartesian about dualism or privacy. But the two are not precise equivalents. As the basis of life, the soul for Aristotle is also that in virtue of which we grow, digest, reproduce, and breathe. In fact, he criticizes his predecessors for not having paid attention to this, and pointedly insists that plants also have a soul, even though he regards them as incapable of perception, thought, feeling, or action. In so far as it studies vital capacities as well as mental ones, psychology has a broader scope for Aristotle than philosophy of mind does for us. But apart from this difference in extension, they will have much in common. Many of the issues that arise in connection with the mind – for example, mechanism, reduction, and emergence – have structural analogues for vital phenomena more generally. The differences between these phenomena will make a difference, of course, as to how we answer these questions in each case. But this is something Aristotle himself emphasizes. The unity of the phenomena that psychology studies is only a loose one. We cannot give a single, general account of the soul, he believes, that will apply across the board and still be substantive and illuminating. Instead, we must attend to the specific activities that are characteristic of different kinds of living thing. In certain ways, perception and thought differ from digestion. But in other ways, perception has more in common with digestion. and Aristotle will regard thought as the odd one out. By keeping all of these phenomena together within a single area of study, we are less likely to generalize falsely about the nature or peculiarity of the “soul” and the “mind.”

The Soul–Body Relation

One of the chief concerns of Aristotle’s treatise De Anima is the relation of the soul to the body. Aristotle wants to account for the way in which they can be said to be “one,” as well as the way in which they differ; and he repeatedly returns to the question of whether there is any sense in which the soul can be said to be “separable” (chôristos). His own solution appeals to the central concepts of his metaphysics. He regards the body as the matter and the soul as the form of a living thing (De An. II.1.

2. De An. I.5, 410b16–411a2, 411a30; II.2, 413a21–b8; II.3, 414a31; II.4, 415a22–b2, b23–26, 416a6–b29; III.12, 434a22–26. See also “On Length and Shortness of Life” and “On Youth, Old Age, Life and Death and Respiration” in his Parva Naturalia, as well as the treatment of reproduction in Generation of Animals.
4. For Aristotle’s treatment of this topic in his biological works, see Lennox, ARISTOTLE’S BIOLOGY AND ARISTOTLE’S PHILOSOPHY, esp. the “Biology and Metaphysics” section, in this volume.
412a15–b6) – hence, the description of his view as “hylomorphism,” literally, “matter-form-ism.” The two are correlative to one another. The parts and materials that make up a concrete object are its matter, while the way they are organized into a whole that can function in the appropriate ways is its form. If we grant this, he claims, then there is no more point to asking whether soul and body are one than there is in the case of wax and its shape (412b6–8). The soul is not itself a certain kind of body (as the Atomists think), but neither can it exist without a body (as the Platonists think). It is something that belongs to a body of a given sort and so inheres in it (II.2, 414a17–22).

Most, if not all, of what living things do or undergo will thus be the activity of body and soul together (I.1, 403a3–19, b17–19). The only exception Aristotle is willing to consider is the understanding (nous); and even this will not be separable, if it always requires the power of representation, which is necessarily embodied (403a8–16). In general, psychological states are also to be understood hylomorphically, just like the composite living thing that has them, as “ennastered structures” (enhuloi logoi) that must be defined in terms of both their formal organization and their material realizations. Being angry, for example, is at once a desire for reprisal and a boiling of the blood around the heart (403a24–b9).

Hylomorphism is widely regarded as the sort of middle course for which Aristotle is famous, a nuanced solution that does justice to the intuitions on each side, without going to either extreme. Others see it instead as a beguiling, but ultimately hopeless, attempt to have things both ways.5 We cannot hope to resolve this question, though, without a more precise understanding of the position itself, and on this antecedent question there is a dismaying range of responses in the literature. Scholars differ over even quite basic questions, such as whether Aristotle is a materialist or a dualist,6 not to mention more refined ones, like whether he is committed to functionalism,7 psychophysical supervenience,8 or emergentism.9

Before exploring these debates, it may help to see why one simple answer will not do. At the opening of De Anima II.1, Aristotle suggests that body and soul are each

5. This accusation is leveled in Williams (1986), although the worry is already present in Ackrill’s seminal article (1972–3). A different charge of incoherence is pursued at length in Granger (1996).
9. For the suggestion that Aristotle is an emergentist, see Ackrill (1972–3) 1979, p. 74); Caston (1997, pp. 326–8, and 1999b); Heinaman (1990, p. 91); Robinson (1983). (Note that both Robinson and Heinaman take emergentism to be incompatible with materialism and supervenience, while I take it to imply both). Against an emergentist reading, see Miller (1999); Sisko (2000).
“substances” or ousiai and that the individual composed from these two is a third (412a6–b9). There could not be a clearer statement, one might think, of substance dualism. But the traditional translation of “substance” is misleading. The Greek, as Aristotle uses it, signifies something’s being, what something is, and the point he is making is that there is a systematic ambiguity when we speak about what a thing is: in one way, it is the kind or type of thing it is and what it is to be that sort of thing; in another way, it is the matter that constitutes or composes such things; and in yet another, it is the compound of the two, that is, this sort of nature as realized by that sort of matter.10 Body and soul are therefore not “substances” in the way that substance dualism requires. They are not two species of a single genus, each of which can exist independently on its own. Rather, each is a substance in a fundamentally different sense, as matter and as form, and both must be conjoined for there to be an individual at all. We can approach this point in another way. If hylomorphism committed Aristotle to substance dualism, he would be a substance dualist not only with regard to humans, animals, and even plants, but all natural bodies. The distinction between matter and form is not peculiar to living things, but applies to all natural substances across the board. In applying it to living things, Aristotle is therefore treating them on a par with the rest of the natural world. In none of these cases can a form exist independent of matter. Form is not an additional ingredient alongside the materials that constitute an individual, which can be separated from the rest. It is rather the organization of these materials into a certain kind of thing, that in virtue of which they are such a thing and possess the relevant capacities (Met. Z.17). It cannot exist apart on its own.

These same considerations suffice to show that Aristotle cannot be an eliminative or reductive materialist either. His analysis presupposes that there is such a thing as the soul, not to mention psychological phenomena; and that there is a fundamental, irreducible distinction between the soul and the body, and more generally between their formal and material contributions in the analysis of psychological phenomena. On the other hand, because the soul is necessarily the form of the body and psychological phenomena have both a formal and a material aspect, he must nevertheless be some kind of materialist, though of a nonreductive sort.11 We might say that Aristotle rejects “type identity,” while accepting a certain “token identity.” Soul and body constitute distinct types: what it is to be animate is different from what it is to be embodied, since one consists in being organized in a certain way, the other in being composed of certain kinds of material and parts. Nevertheless, an individual living thing is both animate and embodied – each of these is rightly said to be something the living individual is (and thus its being or ousia) – and so the same token instantiates both types. Aristotle extends the same analysis to psychological phenomena. My anger on a particular occasion is both a desire for reprisal and a boiling of my blood, even though what it is to desire reprisal differs from what it is to have one’s blood boil.

10. For more on being and substance, see M. L. Gill, FIRST PHILOSOPHY IN ARISTOTLE, in this volume.
11. I am bracketing for the moment (as Aristotle often does himself) the question of whether thought constitutes a genuine exception to this rule. See pp. 335–41 below.
Between these two extremes, there is still a wide range of options, depending on how the matter–form relation is to be understood. It is in this connection that technical concepts such as functionalism, supervenience, and emergentism have been invoked in recent years.

**Functionalism**

Functionalism, perhaps the predominant approach in philosophy of mind over the past 30 years, has also been the subject of intense debate among Aristotle scholars. The functionalist interpretation of Aristotle was initially embraced quite widely, encouraged by the endorsement of Hilary Putnam, one of functionalism’s early pioneers. But during the past decade, it has suffered something of a backlash, triggered by Myles Burnyeat’s polemic, “Is an Aristotelian Philosophy of Mind Still Credible?” (1992). It ends with this defiant challenge: “Hence all we can do with the Aristotelian philosophy of mind and its theory of perception . . . is what the seventeenth century did: junk it . . . new functionalist minds do not fit into old Aristotelian bodies” (p. 26).

The subsequent debate has produced a large and wide-ranging literature. But little of it in fact concerns functionalism as such. The focus has instead been on other theses, which are allegedly presupposed or implied by functionalism, such as supervenience or multiple realization. But if we consider functionalism in its own right, two things become immediately clear: (i) such an approach contains deep insights about Aristotle’s psychology, which deserve recognition; yet (ii) in the technical sense, Aristotle cannot be a functionalist.

According to functionalism, each mental state is defined by its functional role, that is, by all the ways in which it can interact causally, in mediating between sensory inputs, behavioral outputs and other mental states. What plays or realizes this role will be — or, at least, is generally expected to be — some material state. But it is not specified as such in the definition. What is required is that it be a state whose causal powers suit it to play the functional role in question. Consequently, even if a role is always realized by a certain type of material state, it will not be identified as such. In fact, there might even be several types of material state that are suited to play the same role. Functional definitions thus leave open the possibility of multiple realization. The specification of a functional role constrains the kind of states that could play it, without guaranteeing that there is a unique realization or, for that matter, several. Nothing in the definition

---

12. Putnam (1975c) describes his own view as “substantially the same” as Aristotle’s, “although stated a bit more precisely with the aid of the vocabulary of contemporary scientific methodology and cybernetics” (1975b, p. xiv; cf. 1975c, p. 302 and 1975a, p. 279). He never intended the precise details involving Turing machines to be attributed to Aristotle, only the general strategy of functional analysis (Nussbaum and Putnam, 1992, p. 48). See also Block (1980, pp. 171, 177).

explicitly requires either outcome.\textsuperscript{14} Functional explanations are, to this extent, autonomous from material ones. The essence of mentality consists not in what realizes or constitutes these states, but in what they do.

If we consider Aristotle’s psychology from this perspective, three themes become salient. First, Aristotle’s conception of form in general is a functional one, in a suitably broad sense of the term. In Meteorology IV.12, he characterizes forms, and hence what each thing is, in terms of what each thing can do. Nothing can properly be said to belong to a kind \( K \) unless it is able to perform the activities characteristic of \( K \)s – if it is not able to satisfy this condition, it can be called a \( K \) only “homonymously” or by courtesy, due to extrinsic similarities. Aristotle applies this view explicitly to living things (De An. II.1, esp. 412b10–413a3), and employs functional analysis to explore the specific organization of different kinds of souls. Each, in fact, is defined by its capacities, such as the capacity to digest, to perceive, to think and to move (De An. II.2, 413b11–13). Second, matter plays a subsidiary role for Aristotle too. Every living thing, as something that undergoes change, must be embodied. The type of matter is constrained only by its function, through what Aristotle calls “hypothetical necessity” (Phys. II.9: PA I.1). A substance of a certain kind \( K \) must have a suitable type of matter, if it is to perform the activities characteristic of \( K \)s. Just as a saw cannot be made of just anything, but must be made of a metal such as iron (Phys. II.9, 200a10–13), so too a given soul cannot exist in just any type of body: it must be in a body of the right sort (De An. I.3, 407b20–26; II.2, 414a22–27). Beyond meeting this suitability requirement, however, the type of matter is not important to what a thing is, and, in the abstract, more than one type of matter might conceivably do. An axe might be made from various kinds of metal (PA I.1, 642a9–11), and perhaps a human could be constituted from something other than flesh and bones (Met. Z.10–11). Third, for Aristotle psychological explanations are “from the top down” – they give primacy to form. One begins by looking at the type of activities a living thing can perform, in order to develop a functional analysis of the capacities required for such activities. Only then can one turn to the details of how these activities are implemented in specific materials, and how these might malfunction in various ways. Psychological explanation thus presupposes a material account. But it also possesses a kind of autonomy, which constrains the explanatory role of the material account, rather than vice versa.

Each of these themes is central to Aristotle’s psychology. But they still do not amount to functionalism. Functionalism is primarily a thesis about mental states: they are to be defined holistically, by specifying the functional role of all the states simultaneously via the Ramsey-Lewis method of theoretical identifications, purely in terms of their causal relations to one another.\textsuperscript{15} This removes all mental terms from the definiens.

\textsuperscript{14} It is worth emphasizing that multiple realizability only entails epistemic possibility: for all we know, there might be several types of material state that satisfy the definition, since it does not explicitly specify a certain type of material state, much less one that uniquely satisfies it. But it does not follow that multiple realization is a genuine physical possibility in any given case, as is sometimes alleged. Epistemic possibility is just a question of what is “left open” by the definition, due to its abstractness.

\textsuperscript{15} For the classic statement of the Ramsey-Lewis method, see D. Lewis (1970); and for its application to functionalism, D. Lewis (1972), and Loar (1981).
leaving a topic-neutral characterization of mental states that can be used to identify
the underlying material states. Aristotle’s main emphasis, in contrast, is not on mental
or psychological states, but on the soul and the capacities that constitute it. And while
he does discuss the proper definition of psychological states (De An. I.1, 403a25–b19),
the schema he endorses does not approach a functional definition either in spirit
or detail. It makes no attempt to capture the holistic nature of psychological states or
their functional roles. Nor does he define them by their causal roles, by specifying
all of their causes and effects in different circumstances. He does require that the
definition state the efficient cause of a psychological state, along with its matter, form,
and aim – that is, each of his four atit or explanatory factors (403a26–27). But
his own emphasis, moreover, is not on the efficient cause, but on matter and form,
and this directly conflicts with functionalism. The example he provides, a definition
of anger, shows this clearly (403a30–b1). It lacks the abstractness of functionalist
definitions, because it specifies the matter as the blood’s boiling. It also violates topic-
neutrality, because it specifies the form as a desire for reprisal. Aristotle therefore
does not attempt to eliminate either psychological or material vocabulary from his
definitions, one of the key motivations behind contemporary functionalism.

Supervenience and homonymy

Even if we put functionalism to one side, the debate has raised important questions
about related issues, which are of independent interest. Burnyeat’s two main objections
to functionalist interpretations, for example, turn on the modal status of the relation
between the body and the soul. One of them claims that on Aristotle’s theory it lacks
the necessity that functionalism requires. The other claims that it lacks the contingency
that functionalism requires.

The first objection concerns whether the mental or the psychological supervenes on
the physical, whether, that is, a subject’s psychological states necessarily covary
with its physical or material states. If there is supervenience, then given certain material
states, certain psychological states must obtain of necessity: there cannot be a difference
in the latter without some difference in the former. Against this, Burnyeat and others

16. In Rh. II.1–11, Aristotle does offer some relational characterizations of various emotions.
Even if these were intended as proper definitions for purposes of psychology (or parts of such
definitions), they do not make the critical move towards picking out psychological states in terms
of their causal role. They mainly concern the structure of the emotion’s intentional content.
17. Someone might suggest that the aim or telos of a psychological state corresponds to its
causal effects, of which it is the efficient cause. But this is still a far cry from specifying a state’s
causal role. What is at issue for functionalism is the full range of effects a state can have within
one’s psychological economy, not just the ultimate aim intended (such as reprisal in the case of
anger).
19. It is worth emphasizing that “supervenience,” as I am using it here, signifies nothing more
than this pattern of covariation. To claim that the mental supervenes on the physical, then, is
much weaker than to claim that it is a causal consequence of the physical, much less a logical
consequence or somehow metaphysically dependent. For more on supervenience, see the classic
have argued that Aristotle rejects supervenience, citing De An. I.1, 403a21–22.\textsuperscript{20} It is possible to be in same sort of material state that one is in when angry \textit{without being angry}: here would be a difference in one’s psychological condition without a corresponding difference in one’s material conditions. On Burnyeat’s interpretation, a change in perception likewise does not require any change in one’s material state, but only a change in external objects. Others argue that understanding does not supervene on bodily states.\textsuperscript{21}

Aristotle does not, of course, speak directly of supervenience. But various remarks commit him to it. To begin with, he thinks that souls supervene on bodies. He castigates his predecessors for speaking as if any type of soul could be present in any type of body. On the contrary, he urges. “each body seems to have its own form and structure” (De An. I.3, 407b20–24; cf. II.2, 414a22–28). Anything, that is, which has a certain kind of body must have a certain kind of soul: there cannot be difference in souls without a corresponding difference in bodies. The generation of substances quite generally, in fact, supervenes on material changes, and so \textit{a fortiori} the generation of living things and their souls (\textit{Phys.} VII.3, esp. 246a4–8); and similar remarks apply to the demise and destruction of living things. Aristotle does not make a blanket statement with regard to the supervenience of psychological states. But the passage cited above, as evidence against supervenience, actually favors it on closer inspection. Aristotle says that \textit{whenever} we are in that same bodily state, we \textit{are} angry, even if the sorts of external stimuli that normally provoke anger are absent. Indeed, we can have emotions such as fear even when there are \textit{no} external stimuli of the relevant sort; but if we are not in the right state, even pronounced stimuli will not result in an emotional reaction (403a19–25).\textsuperscript{22} We will consider perception and thought more closely in the sections below. But nothing there is incompatible with supervenience either.

If anything, Aristotle is faced with the opposite problem. The fact that a body of the relevant sort must have a soul seems to imply that it is \textit{essentially} alive or ensouled. But then when a living thing dies, \textit{that body} will cease to exist as well – it cannot exist as such, whatever else survives, without the soul that necessarily accompanies it. What was flesh will no longer be flesh except “homonymously,” that is, in name alone.\textsuperscript{21} If so, then living things seem to be quite unlike the artifacts that figure in Aristotle’s examples, such as a bronze sphere, where the matter has that form \textit{only contingently} and can be identified independently of the compound into which it enters. This poses a second problem for the functionalist interpretation, Burnyeat argues (1992, p. 26; cf. 17), because it presupposes “the artifact model”: it requires that material states be identifiable independently of psychological states and contingently related to them.

If homonymy is a genuine problem, though, it would threaten a good deal more. Aristotle’s account of substantial change, of how substances come to be and cease to

\textsuperscript{20} Burnyeat (1992, p. 23); Heinaman (1990, p. 101).
\textsuperscript{22} For a closer examination of the evidence for supervenience, see Caston (1997, pp. 332–7).
\textsuperscript{23} \textit{Meteor.} IV.12; \textit{GA} II.1, 734b24–31; \textit{Met.} Z.10, 1035b23–25; cf. \textit{GC} 1.5, 321b28–32.
be, holds that the matter of a substance underlies these transformations. To do this, though, the matter must exist before the substance has come to be and must remain after the substance has ceased to be. In short, it must be capable of existing independently of the substance to which it belongs. And since living things are the paradigmatic substances for Aristotle – indeed, they may be the only genuine substances for him – the same should apply here. But it does not, if what was said about homonymy above is correct: the matter of a living thing cannot exist independently. If so, then what Aristotle says about homonymy is at odds with one of the main tenets of his metaphysics, concerning substantial change. Either (a) Aristotle is grossly inconsistent, or (b) he has abandoned his views about substantial change, or (c) he has a different understanding of homonymy, which allows that some of the matter of a living thing can exist independently of it. Of these, (c) is plainly the best option.24

Yet how can some matter exist independently of a living thing, while some does not? One attractive solution suggests that we distinguish two kinds of matter: (i) functional matter, such as hands or uniform parts, like flesh and bones, which are specified by reference to the function of the substance as a whole, and thus its form; and (ii) compositional matter, such as earth and fire, from which the functional matter, and thus ultimately the substance, is formed. Functional matter cannot exist apart from the functioning whole, except homonymously, in name alone, because it cannot play that function apart from the substance in question. But compositional matter can exist independently of that substance, because it can be specified without reference to the functioning whole – indeed, it must be able to, if it is to underlie substantial change. In fact, Aristotle draws just this sort of distinction between two senses of the word “flesh.” We use the word in one sense to refer to the form, to what it is to be flesh and be able to do things characteristic of flesh. But in another sense, we use it to refer to the matter that constitutes flesh (GC I.5, 321b19–22). Aristotle can therefore appeal to homonymy when speaking of functional matter, while accounting for substantial change in terms of compositional matter.25

Compositional matter is not only important for problems concerning homonymy. It is important for supervenience as well. The soul and its capacities trivially supervene on the functional matter of living things, since functional matter by definition implies the presence of the functional whole. It is a substantive question, however, whether the soul and its capacities supervene on the compositional matter of living things, and so whether psychological states supervene on material states. It is significant, then, that the evidence for supervenience mentions the elements, and not merely flesh, bones, or vital heat (see Caston, 1997, p. 336).

25. For a full and clear statement of this solution, see Whiting (1992), based on her (1984), esp. pp. 223–34; the same solution appears in Irwin (1988, pp. 241–5), framed in terms of “proximate” and “non-proximate” matter. Chapter 4 of Gill (1989) appeals to the distinction (pp. 127–30, cf. 163), but Gill does not regard it, or the rest of Met. Z, as representing Aristotle’s final view; rather, it poses a problem for him, which is only solved later in Met. H and Θ on a different basis. For a close examination of different conceptions of matter and the ways in which they relate to supervenience and homonymy, see Lewis (1994).
Another objection to functionalism, which has broader ramifications, concerns causation. Aristotle cannot be a functionalist, it is argued, because he regards psychological states and the soul itself as efficient causes. Functionalists, in contrast, locate the real causal power in the underlying material states: for them, it is claimed, causal explanation works “from the bottom up,” not “from the top down,” as Aristotle prefers. Some think that supervenience also precludes psychological states from being efficient causes in their own right. If so, then questions of mental causation will still pose a difficulty for these interpretations, even if functionalism has been abandoned.

Part of the problem here is due to the metaphorical use of “up” and “down,” which is bound to cause confusion unless we distinguish carefully the different issues involved. If psychological states supervene on material states, there is a sense in which the “lower” material states of a living thing determine or even necessitate which “higher” psychological states it has: given certain material states, it must have certain psychological states. But these higher states are not caused by the lower ones. Consider the analogy Simmias offers in Plato’s *Phaedo* (92e–93a; cf. 85e–86c). Whenever a musical instrument like a lyre is in a certain physical state, each string having a specific length and tension of its own—follows necessarily that the instrument has a certain tuning or *harmonia*. But the tuning is not a causal consequence of the strings’ tension, so much as something that covaries with it: for the strings to have a different tuning, they would have to have a different tension. Determination “from the bottom up” is an ontological issue about which properties or states a thing has, given others that it has. It says nothing about causal responsibility. It might be that causal responsibility always resides on the lowest level, and that these lower states are primarily what drives things. But it also might not. There is nothing inherent in supervenience that requires higher-level states to be epiphenomenal, incapable of bringing anything about in their own right. In some cases, it might be the higher-level states, and not the lower-level ones, which are causally responsible—there might, that is, be downward causation, even though there is determination from the bottom up. Which higher-level states a thing has will be determined by the lower-level states it has. But the causal powers of the lower-level states themselves are not sufficient to explain the result. In this sense, the higher-level states have genuinely new, emergent causal powers that are not reducible to the lower-level ones, even though they supervene upon them. In the nineteenth and twentieth centuries, emergentists such as John Stuart Mill and C. Lloyd Morgan argued that this was in fact the way that chemistry was related to physics: which chemical properties a substance had depended on which physical properties it had; but their causal powers could not be explained by their physical properties alone, at least not before the advent of quantum mechanics.


27. For an excellent treatment of emergentism, historically and systematically, see McLaughlin (1992).
Aristotle believes that the soul, unlike the tuning of an instrument, has causal efficacy in its own right, which is not reducible to bodily properties (De An. I.4, 407b34–408a5). The most dramatic case concerns organic rather than mental phenomena. Aristotle argues that the soul alone is responsible for the cohesion of a living thing: on their own, the material elements that constitute it would disperse to their natural places (De An. II.4, 416a6–9). But he likewise regards mental states, such as desire and phantasia, as the proper efficient causes of action and speech: the underlying material states have a merely instrumental role (MA 6, 700b17–20; De An. III.10, 433b13–27). Aristotle is thus committed to “downward causation.” Not all of the effects of a living thing are brought about by it in so far as it is material. Some are the result, at least in part, of a psychological state, precisely in so far as it is psychological.

Whether or not downward causation is compatible with functionalism, it is compatible with supervenience; to accept both supervenience and downward causation is just to accept emergentism. Downward causation, then, need not commit Aristotle to either vitalism or dualism; and indeed neither view fits his overall approach in psychology. The four material elements jointly exhaust the constitution of living things, as they do everything else in the sublunary world. In this sense, all living things are fully material, without remainder. Form is not reducible to matter, but it is not a ghostly element injected into the material world either. Aristotle ridicules any theory that explains mental phenomena by introducing “mind dust” – a soul-element, in effect – whether material or immaterial. Form is rather the organization of existing material at each level of complexity into new substances with new capabilities (Met. Z.17). The substantial natures that result are ultimately based on configurations of the four elements, even though they are irreducible to them. This is not peculiar to life and consciousness. It holds equally for inanimate matter as well. The 18 chemical properties Aristotle discusses in Meteorology IV.8–9 already go well beyond the four elemental qualities of the hot, cold, moist and dry, even though they are based on them. The theory of the elements is not enough, therefore, to account for the behavior of even relatively simple bodies. For Aristotle, chemistry, biology, and psychology have a crucial explanatory role as well. He thus tends towards a form of emergentism: a position committed to downward causation, while upholding the supervenience of higher states, including psychological states, on lower, material ones. If so, then for Aristotle psychological states would have genuinely new causal powers of their own, which are not reducible to those of the underlying material states, without being basic. Which psychological states a living thing has will still be a function of its material states.


29. For a closer examination of the evidence, see Caston (1997, pp. 326–32).  

Aristotle’s Psychology

Perception

In addition to the soul–body relation, the topic of perception constitutes one of Aristotle’s central concerns. Nine chapters of *De Anima* are devoted exclusively to it (*De An. II.5–III.2)*, as well as a separate essay, “On Perception and Perceptibles,” in the short essays referred to as the *Parva Naturalia*. These views have ramifications as well, since he regards many of our other mental abilities – such as *phantasia*, desire, memory, and dreams – as grounded in the perceptual system, and considers understanding to be analogous to perception in certain important ways (*De An. III.4, 429a13–18*).

In *De Anima II.6*, Aristotle distinguishes three kinds of perception, depending on whether what is perceived is “exclusively,” “commonly,” or “extrinsically” perceptible. Something is *exclusively perceptible* if it can be perceived by only one sense, such as sight or taste. Colors, flavors, tones, odors and so on fall into this class. Something is *commonly perceptible* if it can be perceived by more than one sense. Aristotle lists motion and rest, number, shape, and extension as belonging to this group. Both of these kinds of object are intrinsically perceptible. Other kinds of object, such as the son of Diaries, are also perceptible, but only *extrinsically*. They are something that we perceive only in virtue of other features, which are themselves intrinsically perceptible. Aristotle regards what is exclusively perceptible as fundamental, and it is with respect to these that the nature of each sense is to be defined (*418a24–25*). Such perception is always true, while the others are subject to error in varying degrees. Given its exclusive concern with sensible qualities, its infallibility, and its foundational role, it is not unreasonable to think of this basic form of perception as “sensation,” even though the terminology is not Aristotle’s, in contrast with the complex activity that constitutes our ordinary perceptual experience (and typically involves all three forms of perception). In developing his causal model of cognition in *De Anima II.5, II.12*, and *III.4*, Aristotle is concerned almost exclusively with perception in this most basic sense.

**The causal model**

Aristotle takes sensation to be about what brings it about, namely, a particular sensible quality. The color of an object – jade green, say – intrinsically has the ability to affect a transparent medium like air or water, which in turn affects the eye of a perceiver, causing him to see that color. Aristotle is thus committed not only to the reality of sensible qualities, but to their causal efficacy. It is color, primarily, which is capable of being seen, because it is this quality that can affect the eye in the right way so as to produce sight (*De An. II.7, 418a26–b2*). He explicitly rejects the view of subjectivists,

---

32. See e.g., *De An. III.7, 431a12–14; Mem. 1, 449b24–25, 450a10–12; Insomn. 1, 459a16–19.*
33. *De An. II.6, 418a17–18; III.1, 425a14–20; Sens. 1, 437a8–9.* Contrast Plato’s list of “commons” at *Thet. 185c–d.*
34. *De An. II.6, 418a11; III.3, 427b12 (but cf. 428b18–25); III.6, 430b29–30; Sens. 4, 442b8–10; Met. Γ. 5, 1010b2–3.*
35. This is well-emphasized by Sarah Broadie (1992). For alternative accounts of his realism and its ramifications, see Everson (1997, ch. 3), and Broackes (1999).
who hold that there would not be colors, flavors, and so on, if perceivers did not exist. But his rejection is qualified. On Aristotle’s view, sensible qualities would still exist without perceivers. But since they would not actually be sensed, they would not fully realize their nature. They would instead exist only at the “first level” of actuality: they would actually be capable of producing sensation, but they would not actually be producing it.36

When a sense is acted on by a sensible quality, both are brought into activity, one sensing, the other being sensed. Aristotle explains this as he does any agent–patient interaction (Phys. III.3). The activity of the patient (the sense) and the activity of the agent (the sensible quality) are “one and the same” and take place in the patient. But their “being is different” – what it is to be the one and what it is to be the other are distinct (De An. III.2, 425b26–426a26). And just like other agent–patient interactions, it consists in the patient (the sense) taking on the form of the agent (the sensible). Initially, they are only potentially alike, but as a result of the interaction the sense comes to actually be the sort of thing the sensible is already.37 This change does not alter the nature of the sense, causing it to lose its perceptual capacity. To the contrary, the sense realizes its nature through this change, by exercising its capacity, like a builder building (II.5, 417a21–b19). But sensation differs from other changes, such as a plant’s being warmed or cooled, in so far as it receives the form “without the matter.” Just as a bronze signet ring produces a sealing in wax, but does not make it another seal, so a sense is “informed” with the form of the sensible, without becoming an exact replica of its object (De An. II.12, 424a17–24, a32–b2). Aristotle ridicules views that require an internal replica. We do not have a stone in our soul, he replies, but only the form of the stone (De An. I.5, 409b23–410a13; III.8, 431b20–432a3).

The literalism–spiritualism debate

The precise implications of these views are, not surprisingly, the subject of considerable controversy. Some maintain that the sense organ literally takes on the form of the sensible quality it senses, so that, when I look at a rose and then an azure sky, some part of my eyes literally turns crimson first and then azure. Receiving form “without the matter” is, on this view, an entirely ordinary alteration. Aristotle’s phrase indicates that the organ receives only the form of the sensible object and not its matter, so that there is no influx of matter from the object (as Empedocles imagined). But in addition to being an ordinary alteration, this change is also a perceiving, because it takes place in the appropriate part of a sentient being, namely, a functioning sense organ. Such a view has, for obvious reasons, come to be called “literalism.”38

Against this, Myles Burnyeat and others have argued that there is no physiological change that occurs during perception. In coming to see azure, all that occurs is my

36. De An. III.2, 426a15–26; Met. Γ.5, 1010b30–1011a2; Θ.3, 1047a4–10; Cat. 7, 7b22–8a12. One might worry whether these views compromise Aristotle’s realism: see Irwin (1988, pp. 313–14).
coming to see it. Perception is still an embodied process on this view – it is a change of or in a bodily organ. And it is a “physical” change in so far as Aristotle views it as a completely natural change. But there is no material process underlying this change that can be described in topic-neutral terms. The only material conditions necessary for perception are standing conditions, such as having an undamaged eye, a clear and illuminated medium which is not obscured by intervening objects, and so on. The ability to perceive, on Burnyeat’s view, is a basic power of matter that does not involve any further underlying changes. It is for this reason, he urges, that Aristotle’s philosophy of mind is no longer credible. It presupposes a view about matter, and its basic powers, that we can no longer accept, much less find intelligible: for us, matter is not “pregnant with consciousness,” something that is just, in itself, capable of perception. For Aristotle, Burnyeat claims, it is. Such a position has come to be called “spiritualism,” because it regards perception, with Thomas Aquinas, as a “spiritual” change – one that is physical and embodied, but not a material change. 39

Much of the debate here turns on close analyses of key texts and cannot be treated, even cursorily, in such a brief space. 40 Even so, there is good reason to think that literalism and spiritualism do not exhaust the possibilities, and that both in fact are mistaken.

The arguments for spiritualism are largely negative. The most ambitious arguments claim that certain of Aristotle’s doctrines, such as the distinction of perception from ordinary alterations or the causal efficacy of sensible qualities, preclude literalism. 41 But both overstate the case. Perception is not an alteration in the customary sense, according to Aristotle, because it does not destroy our ability to perceive, but rather realizes it, “just as a builder does when building” (De An. II.5, 417b6–15). But we wouldn’t infer from this that a builder can build without any material change going on, seated, as it were, with arms folded. On the contrary, the builder realizes his building ability precisely through material changes such as hammering and sawing. Such changes are not only essential. They in some sense constitute his activity as a builder. But then this doctrine cannot rule out such changes in perception either. The second argument, in contrast, which concerns the efficacy of sensible qualities, does preclude material alterations from playing an independent causal role. But it does not rule out the presence of underlying material alterations, as a supervenience base – that is, it overlooks the possibility that Aristotle is an emergentist. Material alterations might well underlie perception, then, without being in causal competition with sensible qualities and so undermining their causal role.

Aristotle is not in fact silent about material changes in perception, as is sometimes thought. Having classed perception as a state or “affection” of the soul (De An. I.1, 403a3–8), he says that the body undergoes something along with all the affections of the soul, a change that can therefore be distinguished from the affection itself (a16–19, cf. b17–19). In Generation of Animals he also clearly states that visible objects affect the eye both in so far as they are moist and in so far as they are transparent (V.1, 39. Broadie (1992); Burnyeat (1992, 1995, 2001, 2002); Johansen (1998); Magee (2000).
41. For the first argument, see esp. Burnyeat (2002), and for the second, Broadie (1992, pp. 143–5, 150–1).
779b26–780a7). Change in the latter respect constitutes seeing. But this does not preclude change in the former respect, a material – indeed, elemental – change, and these changes, he goes on to explain, have consequences for how we see.

Literalism is not without its difficulties. Literalism is not merely committed to underlying material changes in perception. It is committed to a quite specific change: it claims that when we perceive a sensible quality $F$, our organ comes to be $F$ in just the same sense that the object is $F$. But Aristotle seems to reject this. If the exact same qualities were exemplified, the object would be replicated within the subject, and Aristotle ridicules such views mercilessly. We do not have a stone or a human in our soul when we come to know one, or good and not-good, or any other predication (De An. I.5, 409b23–410a13; III.8, 431b21–29). Whatever it means when Aristotle claims that the sense “becomes like” the sensible quality $F$, therefore, the sense cannot become $F$ in the same sense that the object is, since that would result in an internal replica. But then Aristotle rejects literalism.

This leaves a crucial gap. Aristotle might be committed to underlying material changes in perception, without being committed to internal replicas. This is clear from his account of understanding. When we think of objects which are large and far away, we do so by means of changes within us that embody the proportions of the objects in question, without having the exact same magnitudes, or even necessarily magnitudes along the same dimensions (Mem. 2, 452b9–16). These internal changes model or represent the objects we think about, by having physical magnitudes that embody the same proportions and, perhaps more fundamentally, by sharing certain causal powers in common. Having the same proportions is one way of being similar to something in a precise sense without necessarily being a replica. It is worth noting, then, that Aristotle appeals to proportions extensively in his account of sensible qualities. Colors, sounds, odors, and temperatures are all treated as proportions of contrary qualities. Crimson, for example, is a mixture of white and black in one proportion, purple another.42 A sense organ could thus become like a sensible quality by taking on the same proportions as the quality possesses, yet avoid replicating it by exhibiting those proportions in a different set of contraries: the eye might take on the proportion that crimson has, but exemplify it not in black and white, but in other contraries, such as hot and cold or viscous and runny. So there is textual as well as logical space for alternative material changes in perception.43

**Consciousness**

There has been debate within the scholarly literature as to whether Aristotle had a concept of consciousness.44 In a sense, each side is right. Aristotle clearly distinguishes

---

42. *De An.* III.2, 426b3–7; *Sens.* 3, 440b1–26; 4, 442a13–31; *GA* V.7, 786b25–787b20; *APo.* II.2, 90a18–22; *GC* II.7, 334b14–16. On this point, see Sorabji (1972).
43. For this line of interpretation, see Bradshaw (1997); Bynum (1987); Caston (2005); Lear (1988, pp. 109–16); Modrak (1987, pp. 56–62, cf. 28); Price (1996); Silverman (1989); Tweedale (1992, pp. 226–30); Ward (1988).
being awake and alert from being asleep or knocked out, where the notion of consciousness comes close to that of perceiving. On the other hand, he does not use any single word to pick out the phenomena we have in mind, and he does not share the epistemological concerns distinctive of the Cartesian conception, such as privacy or indubitability. But in another sense, these observations are all beside the point. Aristotle has a good deal to say about the higher-order awareness we have of our own mental states, and this is a sense of “consciousness” which is clearly of interest today.

Aristotle asks how we are able to perceive that we see at the opening of De Anima III.2 (425b12–25), and he offers a set of arguments to show that it is by perception itself. The details are controversial, but virtually all commentators have taken them to rule out a distinct capacity of higher-order awareness, such as an inner sense. This would put De Anima in conflict with his essay “On Sleep and Waking,” though, where he argues that we do this not in virtue of vision, but rather a common sense (2. 455a15–22). But the arguments in De Anima III.2 are better read as ruling out a distinct activity of being aware that we are perceiving. Instead, we perceive that we see an azure sky by the very same perceptual act by which we see the sky. Higher-order awareness is thus intrinsic to each act of perception. It does not belong to any particular sense modality as such, but to the common perceptual ability that underlies the system as a whole.45

Phantasia

After Aristotle completes his discussion of perceptual abilities in De Anima III.2, he turns to the topics of thought (III.4–6) and desire and action (III.9–11). But not immediately. In De Anima III.3, he introduces a new ability, distinct from all the rest, which he calls “phantasia.” This ability will play a central role not only in his accounts of thought and desire, but also in his accounts of memory, dreams, passions, and aspects of perceptual experience that go beyond mere sensation. But as the first chapter of De Anima already makes clear, it has especially significant consequences for the understanding. This, among all mental capacities, is the most likely to be “separable.” But, Aristotle warns, if it requires phantasia, even the understanding cannot exist apart from the body (1.1, 403a8–10). And it does in fact require phantasia. In the last part of De Anima and in the Parva Naturalia, Aristotle repeatedly claims that we do not ever think without a phantasma, the state we have when we are using phantasia (III.3, 428a1–2).

“Phantasia” and “phantasma” are most commonly rendered “imagination” and “image.” But these translations are loaded and potentially misleading. Like the cognates “fantasy” and “fancy,” they have a long history of theoretical usage in philosophy.

45. See e.g., Hamlyn (1993, pp. 121–2); Kosman (1975, pp. 500–2); Modrak (1980–1); Osborne (1983).
46. For a detailed defense of this reading, see Caston (2002).
47. De An. III.8, 432a9, and Mem. 1, 449b31–450a14; cf. also De An. III.7, 431a15, b2; and III.8, 432a4–5, a12–14.
VICTOR CASTON

psychology, and poetics, especially in the past four centuries, which strongly colors our associations. But Aristotle stands at the beginning of this history, which evolves well beyond his original concerns. When we speak of imagination, for example, we often have in mind a source of creativity and invention. But these are not a part of Aristotle’s concerns when he introduces phantasia. And while it has connections to mental imagery and visualization, it may be wrong to think of a phantasma as something which is viewed with the mind’s eye. Many interpretations do make just this assumption, conditioned by the translation “image.” But it is not part of the meaning of the Greek. Such questions can only be resolved by looking at Aristotle’s actual usage and the details of his theory.

Although Aristotle is not the first to use the term, “phantasia” is a technical term. Derived from the passive verb “phantazesthai,” it signifies the capacity through which things are made to phainesthai, to appear or seem to us to be the case. It thus has more to do with things appearing a certain way in experience than with inventing imaginary scenes. Plato actually defines phantasia as “a belief that comes about through perception” (Sph. 264a–b; cf. Tht. 152a–c). Aristotle rejects the claim that it is a kind of belief, but he too is thinking along similar lines. The sun appears to be a foot wide, he argues, even though we believe and in fact know that it is not, but larger than the earth (De An. III.3, 428a24–b9), just as the two lines in a Müller-Lyer diagram (to use a modern example) continue to look unequal even after we have convinced ourselves that they are in fact equal. The way things appear to us has a certain independence from what we believe. Because of this, Aristotle regards phantasia as more rudimentary than belief and even more closely connected with perceptual experience than Plato had claimed.

What enables us to have experiences of this sort? It is not simply perception, since things can appear to us other than they are. Yet it is something like perception and clearly plays a part in perceptual experience, more broadly conceived. This, according to the last third of De Anima III.3, is what phantasia is. Aristotle proposes that it is a trace or echo of perceptual activity. It thus bears a similar content to perception, even after the original perceptual encounter has ended, and so is able to falsely represent how things are in the world (428b10–429a9). In the Parva Naturalia, Aristotle explicitly compares phantasmata to representations, suggesting that they are reproduced from perceptual activity, like an impression from a signet ring (Mem. 1, 450a27–32). In these essays, he also discusses the physiology more extensively, beginning with how the traces persist in the peripheral organs, where they can bring about after-images and other perceptual illusions (Insomn. 2), and then how they proceed to the central perceptual organ (which, for Aristotle, is the heart). Along the way, they are subject to distortion, resulting in dreams that can deviate wildly from our waking experience (Insomn. 3). Because these phantasmata can be stored for long periods in the walls of the heart (Mem. 1, 450a32–b11), we are able to remember experiences long past, as well as search for and retrieve particular items (Mem. 2). There are even a few tantalizing details about the mechanisms of representation itself, at least as regards magnitudes and their ordering (Mem. 2, 452b7–453a4).

The richness of detail in the last third of De Anima III.3, corroborated by the Parva Naturalia, has rightly led many interpreters to emphasize this part of Aristotle’s account. But it has also traditionally been construed in terms of mental images, especially before the rise of behaviorism, when introspection was still dominant in
psychology. On this interpretation, *phantasmata* are not merely representations that bear the content of mental acts. The way they represent objects is by being viewed *internally*, by being themselves the *objects* of an internal mental act. Consequently, they are like the percepts from which they are copied by subjectively resembling them, though they are fainter and less vivid (cf. *Rh.* I.11, 1370a28–29). The echoes of British empiricism here are not accidental. Such interpretations often allude explicitly to Hobbes’ characterization of the imagination as a kind of “decaying sense” or Hume’s description of it as a “faint and languid perception.”

Growing concerns about mental images in the twentieth century have led to a re-examination of this reading. The most influential critiques were advanced by Malcolm Schofield (1978) and Martha Nussbaum (1978). Both allow that Aristotle in places treats *phantasmata* as mental images, but deny that this is essential to their function. Schofield emphasizes Aristotle’s sensitivity to ordinary language, especially skeptical or non-committal uses of the phrase “it appears that” in waking experience, which is more evident in the middle section of *De Anima* III.3 (427b6–428b9). He concludes that *phantasia* is a “loose-knit, family concept” (p. 106), best understood as a passive capacity for having “non-paradigmatic sensory experiences” (pp. 101–2). Nussbaum similarly regards Aristotle as lacking a “canonical theory” (p. 222). But she emphasizes the more positive role *phantasia* plays in Aristotle’s account of action, both later in *De Anima* (III.9–11) and in *De Motu Animalium* (6–11). Here, she contends, Aristotle treats *phantasia* as a capacity to perceive objects as certain sorts of things, in particular as something worth pursuing or avoiding. Both stress the interpretive character of *phantasia* and its pervasive role in ordinary perceptual experience, as opposed to dreams and visualization.

With the resurgence of interest in representational theories of mind in the last quarter of the twentieth century, the pendulum has swung back, at least in part. Deborah Modrak (1987), Michael Wedin (1988) and Dorothea Frede (1992) defend the overall coherence of Aristotle’s theory against Schofield’s and Nussbaum’s critiques. But while both Modrak and Frede accept the traditional view that *phantasmata* are images, Wedin’s account marks more of a new departure by construing Aristotle’s account along “cognitivist” lines. *Phantasia* is not a full-fledged faculty in the Aristotelian sense, but a system of internal representation that subserves the other faculties, where representations or *phantasmata* are not themselves objects of *phantasia*, or indeed Humean images at all. Instead, they are to be understood as physical states of the body, which possess their content in virtue of their similarity to an object, together with their role in the cognitive system as a whole.

There is much to be said for this general approach. Aristotle explicitly treats *phantasmata* as representations that underwrite the content of mental states generally. This naturally includes imagistic experience and visualization. But active visualization is not necessary for these representations to bear their content and so perform their cognitive role. To use a scholastic distinction, they are that “by which” (*a quo*) mental states are about objects. But they are not in general something “towards which” (*ad quem*) mental states are directed – in general they are not themselves the objects of mental states. Criticisms of representational theories, from the early modern period on, often assume that a representation must be the object of some internal mental state in order to represent another object. But Aristotle does not appear to think that this is
necessary. It may simply be that by which our mental states are directed at objects, without itself being an object of a mental state at all. The question of subjective resemblance, therefore, need not arise: phantasmata do not represent by being looked at and compared to the objects they represent.\textsuperscript{48} To the extent that Aristotle does appeal to similarity, it is objective, physical similarities that matter, such as possessing magnitudes with the same proportions as those of the object or having similar causal powers with respect to the cognitive system.\textsuperscript{49} In fact, Aristotle also explains similarity in artistic representation by reference to causal powers, in distinguishing signs from likenesses (\textit{Pol.} VIII.5, 1340a18–35).\textsuperscript{50}

The suggestion that \textit{phantasia} is a form of internal representation that underlies mental states quite generally also fits the contexts in which Aristotle invokes it. On several occasions, he raises puzzles concerning intentionality and mental content, which he uses as a basis for rejecting other theories – any adequate theory, he believes, must have a solution to them. He rejects both pre-Socratic and Platonic accounts, for example, because they cannot solve the problem of presence in absence in its full generality – the problem, that is, of explaining how we can remember or think of objects that are absent, whether they are simply absent from our immediate environment, or no longer existent, or have never existed at all. But a solution can be found, he believes, if we posit internal representations or phantasmata.\textsuperscript{51} Similarly a theory must have a solution to the problem of error, of explaining how it is possible for the content of our mental states to deviate not only from immediate stimuli in our environment, but from the way things are in the world more generally. This is the task Aristotle sets explicitly in the opening section of \textit{De Anima} III.3 (427a17–b6). This passage, though overlooked by most discussions, makes clear the structure of the entire chapter. Aristotle taunts his predecessors for not being able to explain how error is possible given their simple causal account of cognition, according to which “like is known by like.” On such a view a mental state is invariably about what brings it about and so always corresponds to actual conditions in the world – it cannot err or deviate from the way things are. What makes this critique especially interesting is that Aristotle’s own account of the most basic forms of perception and thought does not differ in this regard and Aristotle accordingly takes both sensation and understanding to be incapable of error.\textsuperscript{52} The difference is that Aristotle does not think that all cognition can be reduced to these two basic activities – the simple causal model of cognition that underlies them does not account for content in general. A different kind of activity is required: \textit{phantasia}. In the second section of the chapter (427b6–428b9), he argues that \textit{phantasia} is distinct

\textsuperscript{48} There are of course cases where we do reflect on the content of our mental states – for example, when we wonder whether we are genuinely remembering or not (\textit{Mem.} 1, 450b20–451a14), or whether we are dreaming (\textit{Insomn.} 3, 462a5–7) – and in such cases, we do consider the similarity of their content to that of perceptual experiences. But it does not entail that phantasmata possess their content in the first place by subjective resemblance.


\textsuperscript{50} For more extensive discussion and defense of these claims, see Caston (1998a).

\textsuperscript{51} \textit{Mem.} 1, 450a27–32; 2, 452b11–16; \textit{Peri Ideōn} 81.25–82.6 Harlfinger.

\textsuperscript{52} For sensation, see n. 34 above; for understanding, n. 55 below.
from perception, thought, or any combination of the two. The way it is generated from perception, he argues in the third section (428b10–429a9), explains why its content is similar to perception, yet also capable of deviating from actual conditions in the world. *Phantasia* can represent the world falsely as well as truly (428b17) and thus is a key factor in explaining the complex behavior of animals (428b16–17, 429a4–8).\(^{53}\)

**Thought**

Aristotle’s “noetic” – his account of *nous* or thought (*De An.* III.4–8) – is one of the most influential parts of his entire psychology. It is also one of the most controversial, as it is decisive for several issues of larger importance, including dualism and personal immortality. Given the predominance of these metaphysical issues in the literature, it is worthwhile to start with his views on content and intentionality instead, which have received comparatively less attention.

**The content of thought**

Aristotle’s treatment of thought resembles, in certain large-scale features, his treatment of perception. Just as he distinguished a basic form of perception, which we called “sensation,” from other forms of perception, he also singles out a basic form of thinking from more complex ones that include propositional thought and reasoning. This basic form of thinking or *nous* is perhaps best thought of as “understanding.” Its object is always a nature or essence,\(^{54}\) about which one cannot be in error: either one grasps it or one doesn’t.\(^{55}\) Its infallibility, like the infallibility of sensation, can be traced to the simple causal model that underlies both accounts. Understanding is about the object that brings it about, which causes the understanding to become like it in form, without becoming the object itself (*De An.* III.4, 429a13–18; *Met.* Ψ.7, 1072a30).

For humans, each act of understanding is grounded in *phantasia* and so ultimately perception. Without any sensory experience, humans could not learn or grasp anything (*De An.* III.8, 432a3–8). But we also retain the contents of such experiences in memory, which allows us to have the objects of understanding available within us and so think whenever we want (II.5, 417b19–26; III.4, 429b5–9). The objects of understanding are said to be “in” *phantasmata*. Hence, their contents in some sense depend on quasi-perceptual content (see p. 332 above) which therefore constrain what we are capable of understanding.\(^{56}\) But even if concepts are not without *phantasmata*, they are not reducible to *phantasmata* either (III.8, 432a12–13), since understanding is “of the universal” (II.5, 417b22–23). This difference is plainly due to the interaction of *phantasmata* and the understanding, but on this crucial question Aristotle says very little. According to one common interpretation, it consists in the literal “abstraction”

---

\(^{53}\) A close reading of the chapter and its structure can be found in Caston (1996).

\(^{54}\) *De An.* III.4, 429b10–22; III.6, 430b27–29.

\(^{55}\) *De An.* III.4, 430a26–27, b27–28; *Met.* Ω.10, 1051b15–32, 1052a1–4; cf. *Int.* 1, 16a9–13. The metaphor of touch is used at *Met.* Ω.10, 1051b24–25; cf. Λ.7, 1072b21.

\(^{56}\) See e.g., *De An.* III.7, 431b18–19; *Mem.* 1, 450a7–14.
of intelligible forms from material *phantasmata*, by stripping away or removing the matter to yield disembodied forms, freed from their particularity. But the only process Aristotle explicitly describes is more like selectively attending, or better ignoring, parts of a *phantasma’s* content, as we do when we use diagrams in geometry and ignore those features which are irrelevant to our purposes (*Mem.* 1, 450a1–10).

This basic form of thought is contrasted with a more discursive form he refers to as *dianoia*. This type of thinking involves the “combination and division” of basic concepts, to produce a new compound unity, which is capable of falsehood as well as truth. The analogy Aristotle draws with words and sentences suggests that he takes both combination and division to be forms of predication, where one concept is either applied to another or withheld from it. Understanding, in contrast, is like uttering a single word. Aristotle offers few further details. Apart from the cryptic remarks at *De An.* III.6, 430b20–23, for example, very little is said about how the mind “divides” concepts. Even combination is not entirely clear, as “combining” and “dividing” ordinarily signify symmetric operations, whereas predication is nonsymmetric. One would also like to know more about the basic “simple” concepts, which are uncombined and undivided (or perhaps even indivisible). The metaphysics of understanding

Aristotle devotes considerably more attention to the metaphysics of understanding: the nature of the understanding as both a capacity and an activity, its relation to the body, and the existence of a second understanding, the so-called “agent intellect,” which he says is alone immortal (*De An.* III.5, 430a23). It will be possible here only to outline the major issues.

Aristotle begins *De Anima* III.4 by confronting a question that has dogged him throughout the treatise, namely, whether the understanding is “separable” (*chôristos*). Until this point, he has offered only hedged and qualified remarks. Plainly, it is separable in so far as it is conceptually distinct from other capacities. The question now is whether it is spatially distinct as well (429a10–13). He concludes that it does not have an organ of its own, but is “uncompounded” with the body (Anaxagoras’ phrase), on the grounds that if it had any actual qualities of its own prior to its exercise – as it would, if it were compounded with the body – these would block it from understanding things that we can in fact grasp (429a18–27). Prior to grasping something, there

58. *De An.* III.6, 430a27–28, b1–4; III.8, 432a10–12; *Int.* 1, 16a9–18.
59. *Int.* 1, 16a9–18; *De An.* III.7, 431a8–10, a14–16. On Aristotle’s psychosemantics more generally, see Charles (2000) and Modrak (2001).
60. For some speculations along these lines, see Caston (1998b).
61. On this question, see Aubenque (1979) and Berti (1978).
62. *De An.* I.1, 403a3–b19; I.4, 408b18–29; II.1, 413a4–7; II.2, 413b24–27; II.3, 414b18–19, 415a11–12.
63. Aristotle actually claims that the understanding grasps *everything*. But his argument only requires that being mixed with the body would prevent us from thinking *some* things we actually can think. See Caston (2000).
ARISTOTLE’S PSYCHOLOGY

is nothing more to the understanding beyond the ability itself, like a slate on which nothing has yet been written (429a24–27, 429b31–430a2). This makes it even less vulnerable than our perceptual abilities. If anything, highly intelligible objects strengthen our ability rather than debilitate it (429a29–b5).64

On the face of it, this conclusion sounds like a strong affirmation of dualism. But its exact import is less clear in context. To begin with, it does not imply that the human understanding can exist on its own, independently of the body. In the very next chapter, he claims that this intellect is perishable, in contrast with another, which “alone is immortal and eternal” (De An. III.5, 430a22–25). Nor does it imply that the understanding can function independently of the body’s involvement. As we have seen, all human understanding presupposes phantasmata – it actually grasps its objects “in” phantasmata – and so cannot take place without certain bodily activities.

Aristotle’s stated conclusion is more modest and limited in any case. It need mean no more than this: that there is no organ of understanding, that is, no discrete part of the body that is dedicated to its functioning, as there is for each of the other capacities that make up the soul. It is in this sense that Aristotle can claim that there is nothing more to the understanding, prior to actually grasping something, than its “nature,” namely, the mere ability itself to understand. Beyond the equipment we already possess for other functions, there is no special apparatus for understanding that exists even when it is not being exercised. To go back to the question Aristotle raises at the beginning of the chapter, the ability to understand would be separate only in the sense of being conceptually distinct from our other abilities. It cannot further be located in some particular part of our bodies. It is part of the form of the body, but it is not the form of part of the body, as he had intimated earlier (cf. II.1, 413a4–7). Such a claim is compatible with various forms of materialism (even if it runs counter to our own view that the brain is the seat of cognitive activity).

Some have felt that the understanding must be an “immaterial faculty” in a stronger sense, though, which no materialist could countenance.65 But it is difficult to give these claims a precise meaning. There is a temptation to imagine an incorporeal organ, something analogous to a bodily sense organ, but dematerialized and, as it were, diaphanous. Such a view would conflict with Aristotle’s stated argument, however. For an immaterial organ would have to be something actual, beyond the mere ability of understanding, which would exist even when it is not being exercised – otherwise, it would be an “organ” only in name, and there would be nothing to distinguish this view from the minimalist reading given above. Yet if this organ is something actual prior to acts of understanding, it will have characteristics of its own, which would obstruct the full range of understanding, against what Aristotle claims.

Any interpretation, in fact, other than the minimalist reading will confront the same objection. If the understanding, prior to its exercise, is nothing more than the mere ability to understand, there cannot be anything more to the understanding, whether material or immaterial, that exists between episodes of understanding. The capacity to understand is something that belongs to the human being as a whole, without any additional special apparatus.

64. For close examinations of this argument, see Caston (2000); Shields (1997); Sisko (1999).
65. See Hamlyn (1978); Robinson (1983); and now Sisko (2000).

337
Aristotle’s argument has consequences for his views about cognition more generally. One of the most interesting features of the argument is the way it links the constitution of the understanding to the contents it is able to entertain. It rests on certain assumptions about how forms must be received for cognition to take place, and how the constitution of material organs affects performance in specific ways. Such assumptions are clearly important in trying to evaluate his views on mental content.66

The “agent intellect”

These difficulties pale, however, when we come to De Anima III.5, a chapter of a mere 16 lines. In it, Aristotle argues that there must be a second understanding, traditionally referred to as the “agent intellect” (nous poiēikos or, in Latin, intellectus agens), which alone is “immortal and eternal” (430a22–23). There is not a phrase in the chapter whose interpretation has not been disputed. But plainly it is decisive for many of the issues we have been raising. If each person has an agent intellect of their own, Aristotle is committed not only to personal immortality, but to a genuinely robust substance dualism, where the human soul, or part of it at any rate, can exist after our demise, independent of the body. It is not surprising that over the last 2,300 years it has occasioned more controversy than any other passage in the corpus.67

The chapter consists of two parts: an extended inference, arguing for the existence of the second understanding (430a10–17); and a compendious list of its attributes (a17–25). In every kind found in nature, Aristotle claims, there is (i) something that serves as matter and has the potential to become each of the things in that kind and (ii) something that is the productive cause that makes all the things in that kind, comparable to the way that art (technē) is related to matter and light makes potential colors into actual ones. This distinction, he argues, is also found in the soul: one understanding has the capacity to become all things, another to make all things. The latter understanding is not only separate, inviolable, and unmixed, it is also by its very essence in actuality – it is not the case that it sometimes thinks and sometimes does not. Thus, even though the capacity to understand precedes the activity of understanding in the individual, in the universe as a whole the activity of understanding is prior. This second understanding, taken separately just by itself, is alone “immortal and eternal.” The other understanding is perishable.

The identity of the second understanding is very much in dispute. One tradition, championed by Thomas Aquinas and dominant throughout the twentieth century, holds that (a) each human being has an “agent intellect” of his own, which guarantees some form of continued existence after death.68 But this isn’t the only way to construe Aristotle’s conclusion that the distinction between agent and patient can also be found

66. For an attempt to spell out some of these assumptions more precisely, see Caston (2000).
67. The most detailed account of ancient interpretations is still Kurfess (1911). In English, see Blumenthal (1996, ch. 11); Brentano (1977, pp. 4–24); Kal (1988, pp. 93–109).
68. Thomas Aquinas ST 1a q. 79 a. 4–5; SCG II.76–78; In III De An. lect. 10; Quaest. De An. a. 5; De Spir. Creat. a. 10. For representative examples from the twentieth century, see Rist (1966), and Ross (1949, pp. 148–53).
“in the soul.” It need not imply that both understandings can be found within each human soul, but only among souls generally. The fact that this second understanding is supposed to be eternal and in actuality by its very essence suggests that (b) it belongs to a different kind of soul, either (1) one of the higher intelligences, as many of the medieval Arabic and Italian Renaissance commentators held, or (2) God himself, as Alexander of Aphrodisias maintained (second–third century CE). There is extensive overlap, in fact, between the attributes of the second understanding and those of the divine understanding listed in *Metaphysics* A.7–9, which is held to be unique. On either (b1) or (b2), personal immortality is out of the question, as is substance dualism.

On any of these readings, though, Aristotle’s naturalism would be in doubt. For they all take the second understanding to play a direct and essential role in the production of ordinary human thoughts, and this would seem to preclude a naturalistic account, whether the second understanding is supposed to be a higher intelligence, like God, or a human intellect capable of existing independently from the body. This consequence has been combated recently by Michael Wedin, who argues that Aristotle’s account is “stubbornly naturalistic” (1988, p. 194). In fact, it is thoroughly functionalist and cognitivist in spirit, he claims, especially the distinction between two kinds of understanding. Aristotle first explains the basic features of cognitive activity by positing a single, unified faculty of understanding in *De Anima* III.4. But he goes on to raise two puzzles at the end of the chapter (429b22–430a9), which prompt him to offer a deeper analysis, at a “lower, sub-personal level.” *De Anima* III.5 offers us, in effect, a distinction between two “subsystems” within a single, unified mind that together allow us to think spontaneously, as self-movers. This move does not threaten naturalism, according to Wedin, because the “productive mind” is not literally divine, eternal, or independent of the body. It is active whenever it exists, but it does not exist forever or function continuously (pp. 178–9, 189–90). It is eternal only in the sense that a mathematical object is, in so far as both are defined abstractly without reference to the body; yet neither can exist or function without a body (pp. 190–3). Productive mind is separable only in so far as it is not the actuality or form of a discrete organ (pp. 182–3, 186).

Some may reject such an interpretation as overly deflationary: it is difficult not to take “eternal” as meaning existing at all times. But it would be wrong to blame Wedin’s commitment to “stubborn naturalism.” The real culprit is the nearly universal assumption that the second understanding is instrumental in the production of ordinary thoughts. Absent this, it is easy to offer a naturalistic reading without being deflationary. The second understanding can be literally eternal, and even God, if the causal mechanisms of human thought involve nothing more than the first intellect together with the lower faculties. One of the greatest difficulties for the tradition, in fact, has always been to specify what the second understanding is required to do. Suggestions run the gamut, including the abstraction of universals from images (which is itself

69. If anything, the logical structure of the argument requires that they belong to distinct kinds of soul: see Caston (1999a, pp. 205–11).
70. See Davidson (1992) and Kessler (1988).
71. See Moraux (1942).
explained in diverse ways), selective attention, the ability to think spontaneously, and even free choice. But these are entirely speculative. Aristotle says nothing determinate about how the second understanding would produce thoughts – in fact, this intellect is never expressly referred to outside De Anima III.5. And that chapter, beyond its initial distinction between an understanding that “becomes all things” and another that “makes all things,” only offers analogies to art and to light, which have proven extremely malleable throughout the tradition. So commentators have searched for lacunae in Aristotle’s account that need to be filled. But if the second understanding played such a critical role in the production of human thoughts, it would not have been to his credit to have introduced it in this way. It would be little more than a deus ex machina, a magical problem solver, mentioned only in an exceedingly telegraphic and cryptic manner. Worse, many proposals are difficult to reconcile with what Aristotle actually does say, in particular about the parallel between understanding and sensation. Whether one takes the “agent intellect” to act directly on our receptive understanding or indirectly by acting on an object of understanding (which in turn acts on the receptive understanding), nothing comparable is found in sensation. Aristotle makes no similar call for an “agent sense,” whose causal intervention is required if sensible objects are to have any effect on our capacity to sense. Instead, the sensible object acts directly to produce a sensation of itself. If the object of understanding cannot act similarly, as is traditionally assumed, then the parallel Aristotle draws between sensation and understanding and their respective objects cannot run deep at all: the simple causal model used to explain sensation will be inadequate for understanding. But what exactly is inadequate about this model in the case of understanding? What is it about the object of understanding that prevents it from acting in a parallel fashion? And which differences between the two activities demand the introduction of a new agent that is itself a kind of understanding, rather than an object? What, finally, will remain of Aristotle’s prominent claim at the opening of De Anima III.4 that the understanding is “related to its objects in just the same way” that sensation is to its objects (429a15–18)?

These difficulties are avoided if we deny that the second understanding is part of what we would call the causal mechanisms of thought. The capacities Aristotle discusses at length in the De Anima are sufficient on their own to produce thought. Human understanding does in fact work on the same model as sensation: the object of understanding is able, on its own, to act on the receptive understanding, causing us to grasp it, just as objects of sensation act on the senses. The second understanding is not a part of this account at all, but is introduced only against the backdrop of “the whole of nature” (430a10). At this point in his discussion, as at the climax of his other great works (Met. A.7–9, EN X.7, and EE VIII.2–3), Aristotle considers his subject in its larger, theological context. On this reading, the second understanding is simply God, who is said to be eternal and pure activity (Met. A.7, 1072a25–26, b27–28.

73. Although this did not stop some medieval commentators from worrying about this: see Dewender (1996) and Puttin (1988), and in English, Kennedy (1966); MacClintock (1956).
74. See Caston (1999a), as well as Frede (1996), which gives an allied reading, though different in both its motivations and details.
Aristotle’s Psychology

cf. Λ.9, 1075a10). His role as the Prime Mover, moreover, as something that is ultimately responsible for all movement in the universe, shows how Aristotle might have regarded him as “producing” all thoughts. According to Aristotle, God makes the heavenly spheres move “in the way a beloved does” (Λ.7, 1072b3), by being the object of their striving, an endpoint towards which all their efforts tend. Aristotle explicitly regards this kind of final cause as an efficient cause (poiešton kai kinešton, Λ.10, 1075b8–10, b30–35). But it is not a triggering cause, as moved movers are, which can only bring about change by direct contact (GC I.6–7). God is an unmoved mover, a standing condition that helps to explain certain general patterns of change, rather than the occurrence and peculiarities of particular episodes. The intermittent and imperfect exercise of our capacity for understanding is something that can be fully appreciated only by reference to the most complete and perfect example of understanding, God. It is only then that we understand what understanding really is. But episodes of understanding – why I succeed in understanding something on this occasion rather than fail, or why I have an understanding of this thing rather than something else – must be accounted for in entirely non-transcendent terms, by reference to the objects in my surrounding, my cognitive history, and my very human capacities. These are, one and all, the capacities of an embodied being, even if they are not all forms of a specific part of the body, and they all perish with the body at death. In this way, a naturalistic account of the causal mechanisms of human thought can be preserved. The idiosyncratic sense in which the second understanding is “productive” depends heavily on Aristotle’s views about explanation, and in particular the central importance that teleology has for him. But it does not interfere with what we would call the causal account of thought.

Bibliography

Works Cited


VICTOR CASTON


Aristotle’s Psychology


**Further Reading**


In addition:

