The Ghost of the Machine

An analysis of the conceptual pressure toward dualism in Descartes and its modern manifestations

Matthew Su

Faculty of Arts and Social Sciences

The University of Sydney

December 2021

A thesis submitted to fulfil the requirements for the degree of a Master of Arts by Research.

ACKNOWLEDGEMENTS

I am grateful for the guidance, generosity of time, and patience shown me by my primary supervisor, Professor Peter Anstey, in the conception, composition and revision of this thesis. By your feedback and assistance I hope that I have managed to find something to say which is both interesting and not completely daft.

I am also grateful for the patience and assistance of my secondary supervisor, Professor David Braddon-Mitchell, in reviewing the modern philosophy of mind portions of this thesis in Part 3. It was a major gear-shift in the composition process and I am grateful for the reassuring feedback.

Additional thanks to the friends and family who have humoured me these past two years, and forced me to find pithy and comprehensible ways to talk about souls and machines and why it matters.

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Matthew Su

Abbreviations

Unless otherwise attributed, references in this thesis to the primary works of Descartes are taken from Rene Descartes, *The Philosophical Writings of Descartes*, translated by Cottingham, John, Robert Stoothoff and Dugald Murdoch, Cambridge University Press, 1984. Volumes 1 and 2 are abbreviated as CSM I and CSM II respectively followed by page numbers. Volume 3, with Anthony Kenny as additional translator, is abbreviated CSMK III.

CONTENTS

Part 1: On Descartes and His Precursors

Introduction

This thesis examines the origins of mind-body dualism in Descartes, the consequences of such dualism for Descartes, and their continuing impact in modern philosophy of mind. **Part 1** traces the history of the Descartes' intellectual forbears, in order to frame the significance of the Cartesian turn toward mechanism for understanding the relation between mind and body. It is argued that Descartes' dualism was not the inevitable working-out of the inherent tensions within scholastic hylomorphism (though it did have antecedents), but the result of a decisive shift in philosophical priorities. **Part 2** examines in detail Descartes' particular arguments for dualism, and shows firstly that dualism does indeed follow given the particular choices he has made in setting up his substance-mode ontology to allow for the mechanisation of the physical, and secondly, that this dualism in turn faces an insurmountable and indeed fatal interaction problem. **Part 3** examines the legacy of Descartes' choice to reify a mechanistic-cum-physico-mathematical conception of physical reality in the modern discourse, focusing on the 'structure-and-dynamics argument against materialism.' The chief aim of Part 3 will be to situate the conceptual pressure toward dualism within a complex and crowded field, so the arguments made in this part are made chiefly to show how the conceptual pressure I identify interacts with various contemporary approaches to naturalism, rather than to arrive at decisive conclusions. It is argued that in important ways Descartes' inadequate conception of the physical haunts us still and is a key source of pressure toward dualism. It is suggested that unravelling that conception and diminishing the pressure toward a theoretically unsatisfying dualism (that is, to effectively diminish our surprise that there should be 'non-physical' aspects to reality) will require fundamental metaphysical reflection to recover a sense of ontological complementarity between the 'physical' and the 'non-physical.'

Section 1.1 sets forth the general features of scholastic hylomorphism in order to note the most striking differences between the scholastic synthesis and the mechanistic notions which would follow. In particular, it highlights the fundamental ontological complementarity between 'substantial form' and 'prime matter' at the core of scholastic hylomorphism, which would in turn prove crucial to the sense which prevailed before Descartes that mental and non-mental nature were 'of a piece.' **Section 1.2** traces developments within Scholasticism which foreshadowed the dualism Descartes would later adopt, in terms of the gradual increase in the independence of matter from form as exemplified in the Scotists, the increasingly independent substantiality of substantial form, and the change of the notion of substantial form from a ground of substance to a source of physical behaviour under Suarez. **Section 1.3** follows the contributions of the anti-Aristotelian atomists to the project of mechanical philosophy, and explains the similarities and differences between the reduction to quantity of Cartesian mechanical philosophy, and the atomistic reductionism favoured by the Atomists. **Section 1.4** traces the conceptual evolution of the ideas of mechanism and mechanics, the latter of which progressed from a 'servile art' concerned with the manipulation of nature to a true science explicating nature itself by means of quantity and the relations between them. **Section 1.5** ties together these influences to yield a picture of the Cartesian move to dualism within its intellectual context. It is argued that the most enduring contribution of the Cartesian vision is the rejection of the hylomorphic synthesis which provided the 'metaphysical scaffold' of the medieval scholastics, and the reification of the distinction between quantitative and non-quantitative reality as the fundamental division in nature.

1.1. Some Remarkable Features of Scholastic Hylomorphism

Crucial for understanding the appeal of Cartesianism and the dualistic problems it creates, are the discussions within scholastic hylomorphism (Literally, "matter-form-ism") in the Middle Ages and Renaissance. To understand these developments is to better grasp the significance of

the Cartesian mechanisation of nature, and to better compare the problems facing the pre-Cartesian hylomorphic synthesis with the residual dualistic problems Descartes created in their stead. While the metaphysics of medieval Aristotelianism includes more than hylomorphic themes, for example the important act/potency distinction, the essence/existence distinction and the notion of the four causes (efficient, material, formal and final causes), this account will focus on hylomorphic themes as providing the conceptual backbone for understanding the place of mind in nature, and will invoke the other concepts mentioned as needed. In particular, the following features of scholastic hylomorphism will be noted: 1) The friendliness of a hylomorphic metaphysic to mind, 2) the *metaphysical complementarity* of the chief explanatory principles of mind and body which heads off interaction problems, and 3) the relative unimportance of the quantitative/non-quantitative distinction in delineating the fundamental kinds of entities there are.

Before we proceed, a useful caveat to keep in mind is that 'hylomorphism' is not an actor's category.[1] The term was developed long after the scholastic milieu had broken down to refer to the themes of 'form-matter-ism' which characterised scholastic discourse. While it is important to resist the impression that scholastic hylomorphism was a single consistent synthesis, nevertheless some consistent themes in the medieval discourse may be picked out, particularly in respect of the complementarity of matter and form as ontological principles, which are consciously rejected by Descartes, with this rejection in turn informing Descartes' dualistic metaphysics. By understanding the significance of Descartes' rejection of these hylomorphic themes and the internal developments within the hylomorphic discourse which led to the possibility of a mechanical physics, we can better understand the philosophical

[1] See Gideon Manning, 'Three Biased Reminders about Hylomorphism in Early Modern Science and Philosophy', in Gideon Manning (ed.), *Matter and Form in Early Modern Philosophy*, Brill, 2012, p.1.

origins and costs of his dualism, the better to identify the origins of analogous moves in the present discourse on the philosophy of mind.

Hylomorphism enters the medieval discourse as a significant force at latest with the acceptance of the synthesis of Christian theology and Aristotelianism in the writings of Thomas Aquinas (1225–1274), and it is from Aquinas, therefore, that we derive some of the basic notions of scholastic hylomorphism. From Aristotle Aquinas adopts the idea that material substances (i.e., those capable of undergoing change, and belonging to kinds which are multiply-instantiable) are composed of *substantial form*, a metaphysical principle of unity and universality; and *matter*, a principle of individuation and substrate of change which is defined by and 'receives' form. Form, for Aquinas, is more than mere shape, which Aquinas understands merely as an accident of an already-existing substance.[2] Rather, form is the principle of *actuality* in a substance, the principle of 'intrinsic definition' that helps make a substance what it is.[3] Form of this kind is chiefly motivated by the desire to distinguish substantial change (by which things-in-themselves are generated or destroyed), and accidental change, which only change derivative features of things-in-themselves. Matter, being but the potency to receive form which individuates one instance of the form from another, as lacking form in and of itself, has no existence in and of itself—it is akin to a bare particular, but one which exists only in the context of lending its particularity to some form or other. This general metaphysical outlook is not specifically a thesis in the philosophy of mind, as it is motivated by other concerns applying to inanimate as much as animate reality (i.e., it is motivated by the need for a metaphysical account of change, and of similarity and difference). However, the hylomorphism of Aquinas

[2] E.g., in Thomas Aquinas, *Commentary on Aristotle's De Anima,* translated by Kenelm Foster and Silvester Humphries, Dumb Ox Books, 1999, Book 2 lecture 1, at [218], p.73.

[3] Thomas Aquinas, *Commentary on Aristotle's Physics*, translated by Richard J. Blackwell, Richard J. Spath, and W. Edmund Thirlkel, Dumb Ox Books, 1999, Book 2 Lecture 2 at [153], p.8.

(and, arguably, that of Aristotle), as a theory of inanimate reality, is also *friendly* to a metaphysics of mind in a way that mechanical philosophy is not.

In Cartesian thought, there is a vast gulf between the purely quantitative reality of the *res extensa,* the 'extended thing,' which embraces the whole of nature except the human and divine nature, and the *res cogitans, '*thinking thing,' peculiar to the human soul and the divine nature. By contrast, the emergence of mind given scholastic hylomorphism does not require the sudden introduction of radically new kinds of basic metaphysical principles. The peculiar properties of minds, for scholastic thinkers like Aquinas, are easily situated as aspects of certain kinds of form. Sense is explained by the communication of 'sensible species,' i.e., forms which unify the thing sensed to the sensitive mind, in a material medium. This material medium is *retained* by the sensitive mind (since it retains sensitive species but does not abstract them from particularity, the sensitive mind is considered a *material* phenomenon in Aquinas).[4] Intellect, while an unusual metaphysical operation in that it involves separating the universal forms from their particularisation in matter, nevertheless does not introduce a new fundamental ontological principle unanticipated in nature. The form, already regarded as the principle of 'actuality' relative to the 'potentiality' of matter in inanimate, unthinking nature,[5] is merely liberated in the intellect from its confinement in matter. Form, in the intellect, is treated as '*more itself,*' an unusually potent version of something already found to an extent in inanimate nature,[6] rather than something wholly novel. Even Aquinas' speculative angelology, far removed from everyday experience, requires no novelty as far as the metaphysical components of angels are concerned: angels are, for Aquinas, pure forms (and therefore, they are purely intellectual, operating on understanding alone rather than sense, each its own unique kind of being because

[4] *Summa Theologiae* I, Q78 art. 3.

[5] *De Anima 2.1,* 412b.

[6] *Summa Theologiae I,* Q75 art. 5.

it could not be individuated by matter).[7] The concept of form (with its implicit metaphysical complements) is thus a kind of versatile 'metaphysical scaffold' on which a unified ontology of animate and inanimate nature may be built. The point of noting these theoretical benefits of hylomorphism is not to endorse hylomorphism or medieval angelology, but to note how remarkably *unsurprising* mind is on a hylomorphic metaphysic. If there are distinctions between the rungs on the ladder of being, to the scholastic it is nevertheless obvious that the rungs belong on the same ladder.

It is of course true that scholastic hylomorphism contained dichotomies of its own. The notion of form and matter could be considered a kind of 'dualism.' However, the dualism between matter and form as the basic metaphysical constituents of material things, is very unlike the dualism between the 'extended thing' and the 'thinking thing', and even between 'mental properties' and 'physical properties' in contemporary property dualism, in at least one respect: that the members of the form-matter dyad are defined by *complementary* metaphysical functions with respect to each other. Matter on its own, or 'prime matter,' as a pure principle of individuation and a conservation principle in change, could not (at least for the early medieval Aristotelians like Aquinas) be reified as something which exists in its own right, independently of form: matter without form would be akin to a bare particular, without any particular mode of being. Similarly, form, while it could in particular instances subsist somewhat apart from matter (i.e., in the intellects of rational animals, or in the purely intellectual being of angels), in the ordinary case of physical objects capable of individuation by matter, could not exist on its own – the form of a material thing could have no existence apart from the material substance it informed, except in the mind of some intellect which considered it.

[7] Summa Theologiae, I Q50 art. 2.

This built-in metaphysical complementarity, where the posited principles play mutually-dependent roles in the constitution of a substance, has implications for the prospect of interaction between matter and form: in an ordinary material thing, animate or inanimate, the form is that through which the potentiality of matter, is *realised*—as Aquinas put it, the form of a thing is the cause of the thing's existence 'absolutely,'[8] the realisation of which is the existence of the thing per se. This means that nothing can act as itself without the concurrence of its formal cause, and that the powers of matter to act as a cause of some kind or another are constrained by the form which informs the matter. Because matter is never considered a 'complete' substance in itself, its causal isolation from form is never an issue. This relation of matter and form in a substance implies that 'form' exerts a kind of 'control' over the activity of the substance, yet not as one 'efficient' cause among others. Form is not an internal engine generating internal forces, but the unifying principle which helps to ground a thing's very existence, and *thereby* imparts existence to the forces which are proper to it. As the principle of life in a living thing, and in which thought is embedded in an intelligent thing, the *form* of a living thing is identified with the *soul* of a living thing, and it is as part of such a soul that mind, conceived of as that which specially mediates the interaction of the living thing with the forms of other things, has the kind of causal efficacy proper to it. In the other direction, i.e., in the action of corporeal bodies on an immaterial intellect, there is an issue in the Scholastics at least from the time of Aquinas of the interaction of body with mind, which is the problem of how a mere body can affect a higher, purely immaterial entity or faculty such as the intellect. This problem's significance is not as acute as the problem which (so I will argue) confronts Descartes. Without going too far into the detail of their solutions, it bears noting that Scholastics such as Aquinas and Suarez developed solutions to this issue—e.g., Aquinas's

[8] Thomas Aquinas, *Commentary on Aristotle's Physics*, Dumb Ox Books, 1999, Book 2 Lecture 10 at [240], p.119.

solution that the corporeal impression determines the content which the active intellect derives from it[9]—which draw precisely upon the fundamental ontological aptitude of even relatively material, 'corporeal' nature for becoming understood.[10] It makes sense that ordinary material objects should be a fitting determinant of the content of understanding, precisely because such objects have, in their substantial forms, a kind of 'content' to be liberated.

If the principles of life and being and thought are in scholasticism associated with the formal principle in the substance, the scholastic has a ready answer to the question of what mind contributes to body in their interaction. The scholastic may answer that the body, considered apart from mind, is an incomplete abstraction, and that the body of a living thing acts as such only when under the direction of that particular pattern of activity which in a living thing is called life, and in an intelligent being has sub-faculties or aspects which may be called mental. The form is not an efficient cause acting as an invisible organ, but a pattern *toward* which and *within* which the causal activity of the whole is directed. It is quite consistent with such a view that, within the organism constituting such a pattern of activity, its movements should, as far as their efficient causes be concerned, be referrable entirely to things other than the soul. In the other direction (i.e., thinking about casuality of body upon intellect), even if there is some difficulty as to how something with a lower degree of reality than the 'higher', purely immaterial faculty of the soul, can affect the higher, there remains a fundamental affinity between the formal aspects of corporeal nature and the 'higher,' 'purely immaterial' faculties of understanding which makes such an ascent (involving complex interactions between the 'higher' and the 'lower') tractable. Whatever the merits of such an account of the interaction of body and soul, it can be seen that the underlying metaphysical complementarity of matter and form make it much more difficult to motivate an interaction problem between mind and

[9] *Summa Theologiae I Q.74 Art 4 ad 4.*

[10] See Marleen Rozemond, 'Descartes on Mind-Body Interaction: What's the Problem?' (1999), *Journal of the History of Philosophy 37(3), pp.435-467.*

body. The explanations for their co-emergence and interactions, moreover, are immanent and (in principle) knowable, rather than transcendent and wholly mysterious.

It also bears noting that, in this kind of pre-mechanistic metaphysics, the difference between the quantitative and non-quantitative aspects of reality does not have the huge degree of metaphysical significance which it would acquire in Descartes. Quantity (in the form of extension) is considered, not the 'essence' of material substance, as Descartes might have it, but an 'accident' of substance – something which is, in itself, only a dependent part of the thing which it quantifies. There is no question, then, of bridging an inherent divide between the quantitative and the qualitative world – whatever the relations between particular qualities and quantities, quantity is not understood as something independently subsistent. Quantity, after all, has to quantify *something*, and apart from that something, is a mere abstraction.

Whatever subsequent developments, the sense that the mental and physical components of a human being had this kind of intrinsic complementarity would persist as long as hylomorphic themes predominated in discussions of fundamental ontology. In the rest of Part 1 I trace the unravelling of this complementarity along four lines: first, the gradual conceptual independence given to matter and substantial form, which foreshadowed the substance dualism Descartes would eventually propose; second, the evolution of the place of form in nature represented by Suarez, who under the pressure of sceptical worries gave substantial form an empirical foundation as an internal efficient cause of particular physical behaviours; third, the contribution of the anti-Aristotelian atomists; and fourth, the development of Aristotelian mechanics and the gradual emergence of a fundamentally geometrical and quantitative vision of physical reality.

1.2. The Independence of Matter and Form in Scholasticism

1.2.1. Scotism and the Independence of Matter

One source of controversy in medieval Scholasticism was the independence of matter. Thomas Aquinas (1225–1274) denied that matter could in principle have any independent existence—for Aquinas, matter considered apart from any form whatever is pure potential, and hence, not actually anything at all. Duns Scotus (1266–1308), however, disagreed, for he held that since matter is not nothing, it must have some kind of intrinsic positive quality apart from form. As such, matter had to have some mode of subsistence proper to it, distinct from that of form. This dispute, while seemingly arcane, plays an important role in understanding how the complementary metaphysical functions (and the corresponding principles) of Aristotelianism as received by the medieval scholastics evolved. These developments in hylomorphism could be seen as attempts to set metaphysical operations and notions such as individuation, specification, substance, accident, universal and particular, together with theological concerns such as the omnipotence of God, the verity of the Eucharist and the immortality of the soul, in a coherent system. Insofar as such disputes persisted without making clear progress, they would create an important motivation to get rid of the whole edifice of hylomorphism in general.

To return to the issue of the independence of matter, the key motivation for Scotus is that anything with positive existence unto itself was a possible object of divine creative intent, and therefore had to have, at least miraculously, a possibly-independent sort of existence.[11] Looked at another way, divine omnipotence served as a vehicle for the Scotists to consider just how

[11] John Duns Scotus, *Ordinatio II*, dist. 12, quaest. 2, art.14, 'Every absolute thing that God produces among creatures by the intermediary of a second cause, he can create without this second cause, which is not part of the effect. Now, the form that confers existence on matter is a second cause and not part of the essence of matter insofar as it is matter. Hence, God can create the matter without the form.'

much positive existence prime matter could be said to have – if it had any positive existence at all, then God could realise it, and therefore it could in principle exist apart from form. Omnipotence would be a way of pumping their modal intuitions, and thus clarifying the metaphysical status of matter. Eustachius a Sancto Paulo (1573–1640) and Scipion Dupleix (1569–1661) both held that God could, at least by miracle, produce matter without any form whatever.[12] Dupleix, indeed, went further: such a miracle was not even strictly required, since a purely formless chaos could at least be entertained in thought, and that therefore, though one does not actually find such naked matter in nature, there is nothing absolutely repugnant to nature about the existence of such unformed matter. The possibly-independent existence of matter, sets the stage for a metaphysically austere sort of being, free from any notion of substantial form, which could form the ontological basis of a mechanistic rather than a hylomorphic physics.

In the hands of the Scotists, the distinctive individuating function of matter was also relocated: far from signate matter (i.e., matter 'signed' by quantity) being the principle of individuation, as in Aquinas, the principle of individuation for the Scotists was located in an ultimate specific difference, a 'haecceity,' or a 'this-ness,' which belonged to form rather than to matter.[13] Form and matter became less complementary, and their respective metaphysical functions were reassigned according to various desiderata. Matter became more substantial and independently-actual, while the principle of individuation shifted from matter to form. Ariew sees the transfer of the particularising function of matter to form, to be a step along the way to treating form not as the principle of existence per se of the thing, but merely a kind of individuating accident like structure or shape which particularises a lump matter which has its own pre-existing intrinsic

[12] Roger Ariew and Marjorie Grene, 'The Cartesian Destiny of Form and Matter,' in Roger Ariew, *Descartes and the Last Scholastics*, Cornell University Press, 1999, p.81.

[13] Ibid., 46.; John Duns Scotus, *Ordinatio II*, Dist. 3, quaest. 6.

nature as matter.[14] This move could also be interpreted as tending in a dualistic direction, for if form is entirely independent of matter as a principle of individuation, there is much less impetus to regard the form and matter of, say, an intelligent creature (which the Scholastic is committed to as capable of subsisting without matter at least in some limited form) as constituting a true metaphysical unity with matter. As a theory of inanimate nature, at least, the Scotist line of development (which Ariew argues was predominant in the time of Descartes and with which Descartes would have been quite familiar)[15] was friendly to the eventual Cartesian direction in many ways. Descartes, moreover, was aware of the Scotist tradition, and tended to side with it on questions where the Scotists disagreed with the Thomists.[16]

This leaves the question of whether, for all the foreshadowing of the Cartesian break with hylomorphism, Descartes was merely carrying trends inherent in scholastic hylomorphism to their conclusion. In the first place, it bears mentioning that the Scotists, however ascendant, did not have it all their own way. Defenders of a distinctly Thomistic line on the issues above such as Antoine Goudin (1639–1695), who tended to emphasise the radical incompleteness of matter in the ordinary case, continued to exist among the scholastics.[17] Moreover, for all this change, the language of hylomorphism retained its currency, and the sense that, at least in the ordinary case in relation to ordinary objects in ordinary contexts, matter and form depended upon each other and performed complementary metaphysical functions, was not fundamentally questioned. It was not then inevitable, despite the development of what were in retrospect

[14] Ariew and Grene 1999, p.83.

[15] Ibid., p.55.

[16] Roger Ariew, in *Descartes and the Last Scholastics*, Cornell University Press, 1999, p.55, lists Descartes as siding with Scotus against Aquinas on the object of the intellect, the univocity of being, on the independence of extension, on the principle of individuation, among other issues.

[17] Ibid., p.83.

certain philosophical antecedents in the late scholastics, that the hylomorphic discourse should have led to the characteristically Cartesian break with hylomorphism and its underlying metaphysical themes.

1.2.2. Suarez and the Substantiality of the Soul

Another important line of development feeding into Descartes' dualism is the change in the notion of substantial form represented by Francisco Suarez (1548–1617). Like the Scotists, Suarez's notion of matter is more independent in its existence than is found in Aquinas. For Suarez, substantial form is not, as a Thomist would have it, simply the 'intrinsic principle that gives being to a thing,' despite that Suarez himself tentatively accepts such a definition earlier in Disputation XII of his *Disputationes metaphysicae*.[18] Such a definition attributes the source of a thing's existence per se to its substantial form, which in turn implies that the matter in a thing, which is a part of that existence, derives its existence and causality from form. Instead, Suarez in Disputation XV defines substantial form as "a simple and incomplete substance which, as the act of matter, constitutes **with it** the essence of a composite substance"[19] (emphasis added). While substantial form, for Suarez, is still the principle of unity in a complete substance, it is not the sole source of existence in a substance, which in turn implies a degree of independent existence for the other components of a substance. While Suarez is happy for matter not to have any intrinsic power or quality of its own, he still holds, with the Scotists, that it could be created on its own by divine power.[20]

[18] Francisco Suarez, *Disputationes Metaphysicae*, Disputatio XII, 'De Causis Entis in Genere,' Section 3.3: 'Causa intrinseca quae dat esse rei.'

[19] Francisco Suarez, *On the Formal Cause of Substance: Disputation XV*, translated by John Kronen, Marquette University Press, 2000, p.20.

[20] Ibid., p.120.

Suarez's path to motivating the notion of substantial form is also peculiar. Unlike Aquinas, whose theories on the nature of substantial form begin with the traditional Aristotelian analysis of substantial change, Suarez takes the immortal human soul, a common theological premise with his interlocutors, as the starting point for the motivation a notion of substantial form.[21] In characterising the human soul, which must provide both for 1) the continuing subsistence of the soul after death, and 2) the real union of the soul with the body, Suarez characterises substantial form as that which is both an 'incomplete substance' capable of subsistence on its own, and also capable of acting as a metaphysical component in a substance, performing the function of conferring actuality on a thing which is composed of form itself in addition to a subject fit to be informed by form.[22] This notion of an 'incomplete substance' effectively gives the human soul two modes: The soul is, when alone, effectively a substance in and of itself, capable of an independent subsistence. What makes it an 'incomplete' substance, is that it is also capable of functioning as a metaphysical part of a complete substance, performing one of the constitutive operations of the substance (i.e., acting as its principle of unity) in conjunction with matter. Thus, Suarez can both affirm that the human soul is capable of independent subsistence, while explicitly denying that the soul is a second substance when it is joined to and moving the body. In his paradigm case, then, Suarez preserves something of the metaphysical complementarity within the living individual which characterises the Aristotelian synthesis, while at the same time allowing the relative independence of the soul which the doctrine of the immortality of the soul demands. Such a notion of the soul as a substance unto itself, at least when separated from the body (as opposed to a part of a whole), anticipates the radical substantial divide between soul and body in Descartes.

[21] Ibid., p.20.

[22] Ibid., p.21.

By analogy to the subsistent substantial form which is the human soul, Suarez posits that non-human substances likewise are constituted by a substantial form and a subject, i.e., matter, disposed to receive it: since human beings receive such a subject from their surroundings, that subject must pre-exist human beings in their environment and, absent actualisation by human substantial form, be actualised by something else.[23] The significance of this is that substantial form, modelled on the self-subsistence of the separated soul, acts as something of an internal efficient cause within inanimate substances, in addition to its more commonly accepted metaphysical role of conferring unity upon such a substance. Suarez variously lists substantial form as the probable explanation of: 1) the return of substances to their equilibrium state;[24] 2) the distinction between substantial destruction and mere alteration, the chief example of which for Suarez is the destruction of the ability of certain substances to return to their equilibrium state;[25] 3) the ability of the diminution of one accident to give rise to a reduction in another distinct accident, which requires substantial form to link them;[26] and lastly 4) by how there must be some finite quantity preserved whereby, in certain substances, increase in one respect is accompanied by diminution in another unrelated respect, which Suarez attributes to form.[27] The functions of substantial form in inanimate objects, as a kind of internal principle of elasticity and binding which unites certain collections of accidents and causes them to be responsive to one another and to return to a certain base state where it is present, opens this notion to critique by various empirical means, which offer alternative explanations of this kind of elasticity which do not require some primitive, non-directly-accessible principle of unity.

[23] Ibid., p.21.

[24] Ibid., p.22.

[25] Ibid., p.26.

[26] Ibid., p.26.

[27] Ibid., p.27.

This change of emphasis in the role of substantial form is motivated by pressure from Suarez's interlocutors on both the empirical and metaphysical fronts. On the empirical front, there were two worries: firstly that substantial forms could not be observed, since substantial forms could not be identified with the accidents which were more directly accessible to the senses: the heat and colour of a fire, for instance, were considered accidents, and accessible to the senses, but a 'fire' apart from these sensible qualities, the substantial form underlying the sensible accidents which was not identified with them, was not seemingly accessible to the senses. Secondly, the worry was that the behaviour (at least of inanimate substances) could be accounted for without such mysterious principles as substantial form. These empirical worries were addressed by Suarez's arguments above, and while perhaps consistent with the best understanding of nature at the time, put the substantial form in danger of becoming outmoded.

More fundamental sceptical worries were also advanced: for instance, it was argued that the two functions of substantial form, that it is 1) a subsistent principle which 2) informs matter or were realised in matter, were incompatible with each other.[28] That is, while it is 'informing' the form-matter composite, the substantial form (particularly in the human case, one supposes) is not self-subsistent.[29]

Suarez answered the latter sceptical worries by means of an a fortiori argument with respect to human souls: The self-subsistence of human souls is grounded in their characteristic, immaterial operations unique to the soul itself, i.e., the intellect. But there is nothing repugnant to reason that these operations (especially given the intrinsic insufficiency of matter) should also be part of a wider range of functions which includes material components like sensation. It is precisely by participating in such operations that the soul informs and imparts substantial

[28] Ibid., p.18.

[29] Ibid., p.20.

unity to the whole, but having metaphysical operations in conjunction with others does not imply one must lack an intrinsic operation by which one subsists. A fortiori, if self-subsistent human souls can act as the source of unity in a thing, then surely less subsistent souls, the function of which is even less independent and more bound up with their realisation in matter, could be coherent – their denial, indeed, would be incoherent. Moreover, for Suarez, the conceptual necessity for substantial forms is also supplied by the insufficiency of matter, which in itself is one and uniform, to account for the diversity of beings in nature. Accidents could not do so, since accidents by definition inhere in some thing, leaving form to account for the diversity of being.[30]

According to Hattab, this change in emphasis is an important development: the empirical justification of substantial form, and the reconceptualization of substantial form as something like an internal efficient cause, provides a target against which to raise a counter-hylomorphic research program, with better mechanical principles to account for the changes in substantial form.[31] The metaphysical and logical motivations for substantial form would be directly attacked as both producing an intrinsically incoherent result, and as potentially replaceable by an alternative metaphysics which did not require an obscure mediator between accidents and matter. Suarez's changes, however, for all that they would provide the target for many of Descartes' arguments against substantial form, and the foundation for some of Descartes' sense in which the soul could be called a substantial form, did not alter the fundamental scholastic sense of the complementarity of the principles of form and matter (and hence, mind and body). This is evident from Suarez's arguments for substantial form which trade precisely on the

[30] Ibid., pp.30–31.

[31] Helen Hattab, *Descartes on Forms and Mechanisms*, Cambridge University Press, 2009, p.64.

inadequacy of other causes, like accidents and matter, to account for the full existence of the diversity of nature.[32]

1.3. The Atomists

In addition to developments within Aristotelian Scholasticism, an important element in the intellectual climate leading up to Descartes was the atomist strand of anti-Aristotelianism, as exemplified in the work of Sebastian Basso (b.1573), who is an atomist rejected by Descartes by name. [33] This element is important to note, both for its affinity with Cartesian mechanical philosophy of clearing away explanatorily otiose Aristotelian metaphysical concepts, but also for the Cartesian project's distinct ontological aims, which were to reduce physical reality not merely to its smallest physical constituents, but to its quantitative aspects.

For Basso, the fundamental constituents of things are not metaphysical principles like form and matter, but discrete indivisible, indestructible and non-transmutable physical simples, impenetrable particles of a certain volume and shape, of which everything else is an arrangement.[34] Unlike the Scholastics, who understood generation to produce truly novel entities through the combination of new form with pre-existent matter, Basso's atomism allowed only what the scholastic Aristotelians would consider 'accidental' change—change which did not produce anything qualitatively new, but only varied extrinsic relations between existing substances, characterised by the position of the atoms relative to each other. Hence all apparent qualitative change could be explained by notions of condensation and rarefaction which in turn reduced all change to local motion. Although contemporary Scholasticism also

[32] Suarez 2000, p.20, 25.

[33] CSMK III 26–27.

[34] Ariew 1999, p.134.

contemplated entities analogous to atoms, e.g., 'natural minima' in inanimate substances,[35] the explanatory and descriptive pre-eminence of position, motion and shape for the existence, genesis and corruption of natural objects on atomism (to the exclusion of formal and final causes and notions of qualitative change) were a significant departure nonetheless, providing ontologically pared-down explanations of a proto-mechanistic kind which would have considerable affinities with Descartes' own project. As with Basso, for Descartes all motion can be reduced to local motion. Also like Basso, for Descartes macro phenomena may be explained by local motion in micro-phenomena: changes of quality can be explained without the need of positing anything except motion, size, shape and the arrangements of parts.[36] If mechanistic explanations (i.e., explanations in terms of size, shape and motion) had advantages in perspicuity and observability, these advantages would be common to both Cartesian mechanical philosophy and Atomist naturalism.

For all the affinities in advancing an account of change purely in terms of space and position, in the *Principles*, Descartes takes pains to distance himself from atomism.[37] Even taking into account that the *Principles* are intended to be a textbook used in the Aristotelian-dominated schools,[38] and that therefore Descartes is inclined to maximise the distance between his philosophy and that of the anti-Aristotelians, Descartes' reasons for distinguishing himself from the atomists remain instructive for illustrating the true sources of the Cartesian project. Though he shares the atomist's commitment to a corpuscularian account of the physical, Descartes opposes the idea that corpuscles are indivisible,[39] that there is a vacuum around the

[35] Ibid., p.125.

[36] CSM I 89.

[37] E.g., in CSM I 231.

[38] Ariew 1999, 124.

[39] CSM I 231.

corpuscles, that corpuscles have gravity, or that the corpuscles can be satisfactorily linked to actual observable phenomena on distinctively atomist terms.[40] Descartes, in short, disagrees with the characteristic atomist move toward explanation by indivisible particles, even as he agrees with the mechanical explanations of change to which atomism (because it is a form of corpuscularianism) lends itself. This led to further differences: for instance, Descartes thought that motion was possible even in the void (though for his own reasons, he denied that a true void was possible), as opposed to requiring mediation by particles of 'ether.'[41] Descartes was even more committed than Basso to the fundamental fungibility of material things: as comprising nothing but quantities like size and shape, Cartesian extended things are fundamentally transmutable with each other. Motion, framed in Cartesian philosophy as the variation in quantitative attributes over time, was an inherent feature of all moving substances, rather than something inherently imparted by a special 'aether' form of matter. While Descartes agrees with the atomist's negative project of undermining the scholastic idea of substantial form, the positive vision of reality which he aims to establish instead of Aristotelianism has a carefully limited affinity with Democritean atomism.

What chiefly attracts Descartes to mechanistic explanation is the explanatory power of quantitative analysis itself, rather than Basso's derivation of mechanistic explanations from his fundamental atomist convictions. It is thus Descartes' own positive reflections on first principles which lead him to conclude that he has nothing to learn from Basso.[42] While

[40] CSM I 287.

[41] Ariew 1999, p.137. For Descartes, though he strenuously denies the existence of a void (since a void must be extended, and for Descartes, extension is a 'principal attribute' of substantiality) continual motion is sustained by God's immutability in creating the natures of finite objects, rather than the movement of particles by the aether.

[42] Ariew 1999, p.127.

Cartesianism shared many sympathies with the Democritean project of clearing away theoretically unfruitful accretions such as substantial forms and final causes, and the latter can be seen in many ways as generating trends which come to their fruition in Cartesianism, the reification of quantity as the ontological basis of material reality ought to be seen as a uniquely Cartesian contribution to a mechanistic picture of nature. In advancing such a view, Descartes sees himself as appealing to universal principles that, as far as physical reality is concerned, both the Aristotelians and the anti-Aristotelian atomists would be bound to accept, which are therefore 'most ancient of all.'[43]

1.4. Aristotelian Mechanics

Central to the Cartesian vision of nature which supplanted hylomorphism was the idea of the 'mechanical,' which had an origin within the Aristotelian paradigm itself, as the name given to the 'servile' arts by which human agents could 'violently' alter nature to meet their variable ends. In order for the idea of the mechanical to rise to the level of a metaphysics describing all of reality rather than merely a set of useful principles for practical ends, considerable conceptual space had to be traversed. Understanding the development of the discipline of mechanics from an art, useful but not reflecting necessary truths about the intrinsic nature of things, to an intellectually respectable science capable of establishing general principles of nature on certain demonstrative foundations, is therefore essential in identifying the source of those conceptual innovations which would allow Descartes to unravel the scholastic Aristotelian synthesis and produce a vision of nature without the central hylomorphic themes of Aristotelianism. **Section 1.4**, drawing on the analysis of the theoretical motivations for the

[43] CSM II 391.

development of Aristotelian mechanics in Hattab (2009),[44] reconstructs a theoretical trajectory in Aristotelian mechanics which sets the stage for Cartesian mechanism, and identifies key theoretical advantages of a mechanistic view of nature, especially over the medieval scholastic synthesis. While the influence on Descartes is perhaps indirect, it is nevertheless important for situating the Cartesian turn within its intellectual context vis-à-vis hylomorphism.

1.4.1. Mechanics: from Art to Science

In the scheme of the sciences and the arts prior to the rise of the mechanical philosophy, the study of 'mechanics' did not enjoy the status of a science. Broadly, the sciences were traditionally bodies of speculative knowledge rooted in first principles (propositional or ontological; the distinction was not always clearly drawn)[45] that were susceptible of certain demonstration, and from which certain demonstrations to knowledge of necessary truths could be derived. Arts, by contrast, were bodies of knowledge which, though they were governed by principles of their own, were not concerned with deriving certain or necessary truths.[46] As Aristotle put it, 'art is concerned with what can be otherwise.'[47] Among the arts, the 'mechanical arts,' as in, e.g., the list in the *Didascalion* of Hugh of St Victor,[48] were diverse arts ordered toward merely 'useful' ends, and were mainly concerned with 'violent' movement, i.e., movement against the intrinsic natural tendencies of things toward a 'natural place,'

[44] Helen Hattab, 'The mechanical alternative to substantial form' in *Descartes on Forms and Mechanisms,* Cambridge University Press, 2009, p.85–119.

[45] Peter R. Anstey, 'Principles in early modern philosophy and science,' in D. Jalobeanu, C. T. Wolfe (Eds), *Encyclopedia of Early Modern Philosophy and the Sciences*. Cham: Springer, 2020, p.4.

[46] Peter Anstey and Alberto Vanzo, 'The Origins of Early Modern Experimental Philosophy' (2012), *Intellectual History Review,* 22(4), p.502–503.

[47] Aristotle, *Nicomachean Ethics,* 1140a1–1140a23.

[48] Hugh of St Victor, *The Didascalion of Hugh of St Victor*, translated by Jerome Taylor, Columbia University Press, 1961, p.56.

contrived according to human aims rather than from the first principles of natural movement. The 'mechanical' arts encompassed not only the principles of machines, but of practical crafts like agriculture and weaving. Such things could be very useful in practice, but did not rise to the level of an ordered *science* which, on the Aristotelian model, proceeded by demonstration from certain first principles toward necessary truths concerning the intrinsic natures of things.

With the recovery of the pseudo-Aristotelian *Questiones mechanicae* (which was regarded as a genuine Aristotelian work in the 16th and 17th centuries), which dealt with the operation of the five simple machines, the study of mechanical questions came to be treated as a discipline straddling an art and a true science: its principles and methods of explanation were mathematical, and therefore capable of logical demonstration from evident axioms according to universal rules, but these were applied to the movements of physical objects, put to human purposes. Understanding the medium in which the abstracted objects of mechanics were realised—the physical entities involved in mechanisms—was the object of physics rather than mechanics. Because mechanics dealt with violent motions realised in a natural object, a true science of the machine (which was necessarily physical) could not be derived from the principles of mechanics. Even the more technical and mathematical sort of mechanics, then, remained an 'art', unifying geometry and physics for a human (therefore variable) end, rather than a 'science' of reality in itself.

Alessandro Piccolomini in 1547 would argue for the scientific status of mechanics, first by separating out that practical discipline of which the characteristic methods of demonstration involved mathematical principles from crafts like smithing or agriculture, and arguing that this discipline is a science not because of its *end*, but because of the *methods* of demonstration proper to it (this latter move, classifying an art by its method of demonstration rather than its

end, is a significant departure from Aristotle).[49] In contemplating the mathematical principles behind machines, which inform every application of the mechanical art, the mechanist engages in a properly 'scientific' mode of contemplation and demonstration rather than 'mere' craftsmanship governed by rules of thumb. The significance of such divisions is more than an exercise in academic 'turf defence.' The main obstacle preventing the mechanical arts from being regarded as a proper 'science,' even with its employment of geometric reasoning and mechanistic explanation, derives from a fundamental Aristotelian division in nature: that between the 'natural' and the 'violent.' Natural principles were intrinsic and proper to things in themselves, defining what they were and therefore what they did universally (unless violently affected), giving rise to doctrines of natural motions and natural place. Violent motions, on the other hand, were motions which did not stem from the natures of things in themselves, but were imposed extrinsically, particularly, in the case of the machine, by human design and contrivance. As long as mechanics was a 'mixed science' between physics and geometry, it would be regarded as inevitably subordinate to the Aristotelian principles of physics and pure mathematics, and could not supplant the need to refer to the Aristotelian principles of physical things in the explanation of their nature and behaviour, and this in turn prevents the mechanist, qua mechanist, from contemplating the true causes of phenomena (i.e., to truly contemplate such causes, she would have to consider them as a physicist or as a geometer rather than as a mechanist). To reach the Cartesian notion of the *res extensa*, where the physical consists in pure extension and dispensed with the need for the crucial (to Aristotelians) notion of substantial form, it would be necessary to unite the study of the geometrical principles of physical things and physics as study of a single subject matter.

[49] Hattab 2009, p.96.

1.4.2. The Identification of the Objects of Physics and Mathematics

Hattab locates the preparations for the shift from the hylomorphism of the scholastics to the Cartesian *res extensa* in three moves by the Aristotelian mechanists: Firstly, there is the use of the image of God as a divine geometer, creating the universe according to mathematical proportions, which raises the study of these proportions to a kind of first principle. Such a move is found in such expositors of Aristotelian mechanics as Henri de Monantheuil (1536–1606), Giovanni di Guevara (commenting in 1627), and Josephus Blancanus (1566–1624).[50] [51] Secondly, mathematical modes of demonstration were applied to the behaviours of physical objects, as in Monantheuil, who explained the motions of machines by means of their geometric properties, and Guevara, who explained the motions by their relative properties and their proportions.[52] The effect of such a move is to raise mathematical demonstrations from assertions about the properties of abstract mathematical objects, to explanations of the causal powers of concrete things, in turn giving mathematical description a kind of ontological 'heft' they had previously lacked. Thirdly, physical objects could in some cases be identified as imperfect instances of geometrical objects, as Blancanus would do, which again served to bring mathematical demonstration and physics together.[53]

These tendencies would come together in helping mechanics attain the status of a science which genuinely explained the motions of both natural and artificial objects, rather than a mere

[50] Ibid., 99.

[51] These commentators were known to Descartes' friend and interlocutor, Marin Mersenne. and cited by name in his *Questions Théologiques, Physiques, Morales et Mathématiques,* composées par L. P. M. Paris: Henry Guenon, 1634, p.38, in Peter Dear, *Mersenne and the Learning of the Schools,* Cornell University Press, 1988, p. 126.

[52] Ibid.,109.

[53] Ibid.

calculating device to generate predictions: if mechanical principles are the eternal principles by which God has created the world, mechanical (i.e., applied-mathematical) explanations governed the motion of objects within the world, regardless of whether it was a machine or a natural entity. Indeed, if physical reality could itself be conceived as the imperfect instantiation of geometrical objects, then key aspects of metaphysical respectability—rootedness in eternal truths, explanatory power and ontological self-sufficiency—are secured for mechanistic explanations, which explain the motions of physical objects both natural and artificial. This in turn would make it easier for someone like Descartes to altogether eliminate the ontological distinction between violent and natural motion (grounded ultimately in the notion of substantial form) to which the artificial/natural distinction maps, in favour of a purely mechanical conception of nature. In assimilating nature to the image of the machine, which even for the Aristotelian is governed by no 'intrinsic nature' except geometric and mathematical principles governing the interaction of solids, a mechanist could unify all of physical nature, 'natural' and 'artificial,' under the same explanatory and conceptual framework, and render the distinction unimportant. From the developments in Aristotelian mechanics, the application of mathematics and in particular geometry to the understanding and explanation of the intrinsic nature of physical phenomena, promised *unification* and *simplification* of our science of the physical without loss, but rather increase, of explanatory power.

1.4.3. The Epistemic Advantages of Mechanism

The close relationship between abstract mathematical demonstrations and mechanical explanations of real physical objects and their motions, and the epistemic advantages of adopting such a close relationship, help to explain the eventual liberation of mechanical philosophy from Aristotelian metaphysics. Humanist sceptics, like Francisco Sanchez (1551–1623), attacked the Aristotelian theory of demonstration by attacking the possibility of non-nominal definition: for Sanchez, the definitions of terms are purely verbal rather than reflective

of the real essences of things, and hence cannot supply insight into the actual causal efficacy of real objects, in turn undermining the capacity of syllogistic demonstrations based on such definitions to yield certain insights into reality.[54]

For the Aristotelian proponents of mechanical science like Blancanus, the mathematical definitions which served as the first principles of mathematical demonstrations were not mere names, but contained in themselves the essence of their subject matter, the abstracted
570 mathematical object as it appears before the intellect, which is shorn of its material imperfections.[55] Mathematical objects, unlike physical substances, are known in themselves and immediately through their very definitions, and that knowledge in turn provides knowledge of the causes of the mathematical object's properties:[56] e.g., for an equilateral triangle, the definition of having three equal sides explains the property of having three internal 60 degree angles. In this way, the properties of the mathematical object follow from and are caused by the definition so understood. Mathematical definitions, then, are 'causal definitions'[57] (i.e., definitions capable of serving as an explanation for the inherence of attributes)[58] for mathematical properties, capable of being genuinely explanatory, and that in a way entirely

[54] Francisco Sanchez, *That Nothing Is Known = (Quod Nihil Scitur)*, translated by Douglas F. S. Thompson and Elaine Limbrick, Cambridge University Press, 1988, pp.174–175.

[55] Josephus Blancanus, *De Mathematicarum Natura Dissertatio*, Appendix to *Aristotelis Loca Mathematica*, translated by Gyula Klima, in Paolo Mancosu (ed.), *Philosophy of Mathematics and Mathematical practice in the Seventeenth Century*, Oxford University Press, 1996, p.180.

[56] Ibid., 181.

[57] Ibid., 183.

[58] John Loneway, 'Medieval Theories of Demonstration', The Stanford Encyclopedia of Philosophy (Spring 2021 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2021/entries/demonstration-medieval/>

transparent to the intellect. The transparent and indubitable connection between mathematical definitions and the properties of mathematical objects is thus extremely attractive epistemically, and the application of such principles to the physical, as in mechanical science, promised a science of the physical which would avoid sceptical worry.

Guevara, for his part, attempted the application of such mathematical principles to the physical world by means of a two-stage process involving Aristotelian syllogisms. In each such syllogism, the middle term between the explanans and the explanandum is supplied by geometrical or mathematical demonstration from a mathematical principle immediately grasped. In reasoning from the effect to the geometrical causes for it via such necessary principles, and then from that cause to the effect to show that the effect invariably or inevitably follows from the cause, a 'scientific' conclusion, i.e., certain knowledge, is reached.[59]

Epistemic and metaphysical issues, however, remained for the project of a mechanical science of physical reality. Firstly, if the geometric properties of physical objects are mere 'accidents,' (i.e., attributes which are when considered separately from primary substances, mere abstractions) and yet the very being and causal power of a physical object stems from its substantial form (or internal principle of real definition which is not a mathematical object) the project of a truly 'causal' demonstration according to Aristotelian standards, by reasoning toward a purely mechanistic explanation, is doomed. This problem corresponds roughly to the issue of information loss in relation to our mathematical idealisations of reality. Mechanics, in the traditional Aristotelian framing, could serve as a useful tool of calculation, but not a full description of physical reality. Even if this problem is not something which is fatal for all purposes (e.g., technological purposes, which require only a good-enough prediction of

[59] Hattab 2009, p.118.

results), it is a difficulty if one aspires to a demonstration of the true first principles of being and change in Aristotelian primary substances, the 'real' constituents of reality.

A second problem for an Aristotelian mechanist is the inductive inference from the imperfect (e.g., sphere-like) entities in physical nature to the properties of ideal solids (i.e., ideal spheres) which explain the behaviours of the physical entities. The Aristotelian mechanists could certainly, by fiat, treat physical objects as imperfect instantiations of mathematical objects. However, if induction works by abstraction from the particular, e.g., to abstract sphericity from the actual shape of the concrete object which is not actually a sphere, even if it somewhat resembles a sphere, then the ability of mathematical principles to act as real causes of the behaviour of physical objects, because of the lack of a causal connection between the mathematical object and the physical object, blocks the inductive demonstration and thus prevents 'scientific' (in the Aristotelian sense) demonstration from cause to effect. An analogue of such a problem in the modern sense, is the strangeness of the applicability of mathematics to physical reality,[60] where it is not clear why abstract mathematical reasoning should be as good a guide to physical reality as it is. For the mechanist looking to replace the Aristotelian conception of the physical, it would therefore be a valuable coup to secure the applicability of mathematical principles to physical reality by undermining the motivations for the rival Aristotelian conception of physical nature.

Both these challenges are fundamentally metaphysical in nature, as Descartes realised. It is precisely the distance between the scholastic-Aristotelian model of physical reality on the one hand, which is built on a scaffold of non-mathematical ontological principles, and a mathematical physics on the other which provided the main obstacle to reaping the great

[60] See, e.g., Eugene P. Wigner, 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences' (1960), *Communications of Pure and Applied Mathematics*, 13(1), pp.1–14.

theoretical benefits of a quantitative, geometric conception of physical reality; such a conception would make mathematics the language in which physical reality would be comprehensively described and physical phenomena explained, enabling a 'physicalising' of mathematics.[61] The chief obstacle to traversing this distance were the notions of real qualities and substantial form at the heart of the scholastic synthesis which by their very nature resisted reduction to their mathematical properties. Descartes, aware of the Aristotelian mechanists via interlocutors like Mersenne and Mydorge,[62] would have to frame a new ontology which kept the epistemic robustness and metaphysical plausibility of the mechanistic approach while rendering hylomorphic notions like substantial form explanatorily otiose.

1.5. On the Benefits and Costs of Descartes' Mechanistic Turn

Descartes' concern with finally eliminating substantial forms and thereby the hylomorphism holding the scholastic synthesis together in favour of dualism, comes mainly by way of his commitment to mechanistic explanation—that is, explanation of physical phenomena by means of the quantitative, mathematically describable features of those phenomena. A mechanistic view of non-mental nature first commends itself in Descartes' investigations of various physical phenomena, particularly in terms of the apparently superior accessibility, clarity and precision of its explanations and predictions. In response, the place of mind (paradigmatically, the intellect) in nature is confined to the human soul, which already had, due to the Scotists and Suarez, a certain independence from matter and indeed, in Suarez, the status of a substance in its own right. Subsequently, Descartes would construct a (relative to the elaborate framework

[61] John Schuster, *Descartes-Agonistes*: *Physico-Mathematics, Method and Corpuscular-Mechanism 1618–33*, Springer, 2012, p.56.

[62] Hattab 2009, p.92

of Aristotelian-Scholasticism) simplified metaphysics in which to frame the division between non-intellectual nature and intellectual nature as a fundamental division in nature.

Descartes, having in his early adulthood come under the influence of corpuscularian-mechanist Isaac Beeckman, was extremely impressed by the power of mathematical and geometric reasoning to explore with remarkable precision, transparency (or, in Cartesian terms, with 'clarity and distinctness') and therefore certainty the nature and necessary implications of shapes and solids. [63] Though the requirement that the deliverances of the sciences be deductive derivations from certain foundations was a core commitment of Descartes' scholastic Aristotelian forebears, Descartes saw the function of substantial form in the explanation of physical phenomena as obscure or 'occult' compared to the 'manifest and mathematical reasons for natural actions,'[64] and therefore as a source of the very uncertainty that a practitioner of science ought to eliminate. If a true science was to be built on nothing but secure foundations, then, these occult principles needed to be removed, for they only compounded obscurity. For Descartes, the mechanical philosophy centred on the mathematical modelling of reality, by treating its explanatory agents with rigour and certainty, and supplemented by observations and confirmed by practical implementation, is preserved from corruption and confusion which comes of treating indistinct ideas and forms of words. As Descartes put it in a letter to Plempius dated October 3, 1637,

> If my philosophy seems too 'crass' for him, because, like mechanics, it considers shapes and sizes and motions, he is condemning what seems to me its most praiseworthy

[63] Descartes, addressing physics, says that '*The only principles* [he] *will accept or require in Physics are those of Geometry or pure Mathematics; these principles explain all natural phenomena, and enable us to provide quite certain demonstrations regarding them*,' CSM I 287.

[64] CSMK III 208–209.

> feature, of which I am particularly proud. I mean that in my kind of philosophy, I use no reasoning which is not mathematical and evident, and all my conclusions are confirmed by true observational data. Whatever I concluded to be possible from the principles of my philosophy, actually happens whenever the appropriate agents are applied to the appropriate matter. I am surprised that he does not realize that the mechanics now current is nothing but a part of the true physics which, not being welcomed by supporters of the common sort of philosophy, took refuge with the mathematicians. This part of philosophy has in fact remained truer and less corrupt than the others, because it has useful and practical consequences, and so any mistakes in it result in financial loss. So if he despises my style of philosophy because it is like mechanics, it is the same to me as if he despised it for being true.[65]

By describing visible 'macro'-phenomena in terms of their quantitative aspects of shape, size and motion, and explaining them in terms of the shape, size and motion of their component parts (i.e., describing and explaining them in mechanistic terms), Descartes saw that one could elide the mysterious causal role of substantial forms altogether. This notion of substantial form as primarily a source of motion and change, in effect acting as a 'little [Cartesian] soul,'[66] as Hattab points out,[67] owes much to Suarez's preferred justifications for the existence of substantial form, which themselves under pressure from sceptical empirical worries such as those of Sanchez, treated substantial form as a kind of posited internal cause of observed

[65] CSMK III 64.

[66] Descartes advances a similar argument against 'real qualities,' i.e., non-quantitative qualities in nature, where he says, 'The first is that I do not suppose there are in nature any real qualities, which are attached to substances, like so many little souls to their bodies, and which are separable from them by divine power,' CSMK III 216.

[67] Hattab 2009, p.186.

behaviours which was in-principle replaceable by a sophisticated mechanist project. In place of mysterious and occult substantial forms, Descartes conceived a mechanistic approach, centred in the analogy of the machine and the way in which its parts, under consideration not of substantial forms but their shape, size and motion, or the 'physico-mathematical' approach, which replaced substantial forms with the analysis of physical objects as primarily constituted by their 'extended', geometrically describable attributes, such as size, shape and motion in space.[68]

Such an epistemic commitment to the explanatory superiority of mechanistic explanations and
690 their consistent usurpation of the causal role of substantial forms in the behaviour of visible phenomena, suggested the project of replacing the notion of substantial form altogether in the realm of physics. Hence in *The World* (written between 1629–1633), Descartes would endeavour to describe a world comprehensively in terms of mechanistic principles, without the intervention of anything like a substantial form or prime matter.[69] The model of a complex physical substance like a living thing was the machine without any internal substantial form, unified only by a design extrinsically imposed and operating by means of internal parts structured like mechanisms such as pipes, springs, bellows and clocks.[70] The principles by which such a machine worked, which fully accounted for the behaviours of the machine, were

[68] While 'physico-mathematics' is not necessarily an actor's category in Descartes' case, in the sense that Descartes did not explicitly apply the approach to his own work, it does arise in the work of Descartes' mentor, Beeckman, and is used here to refer to Descartes' emphasis on the use of peculiarly mathematical principles of explanation and demonstration in his physics. On the subtle differences between Descartes' use of mathematics and Beeckman's, see Roger Ariew, 'Comments on John Schuster and Frédéric de Buzon concerning Physico–Mathematics and Mathesis in Descartes' (2018), *Journal of Early Modern Studies,* 7(1), pp.175–186.

[69] CSM I 90.

[70] CSM I 100.

at least in principle physico-mathematical. Though *the World* contained no precise mathematical statement or premise, because it endeavoured primarily to replace considerations of what kinds of 'soul' might govern a physical substance with the image of a series of machines governed by considerations of size, shape, arrangement, and motion, nevertheless through its central 'machine' analogy such a world is tractable to the 'mathematical' approach to physics Descartes mentioned to Plempius. Yet such a work, while a powerful statement of the possibility of a mechanistic vision of nature, was still a relative half-measure as concerned the liberation of physics from Aristotelian metaphysics: The notion of substantial form, as long as it remained credible in the actual world, still served to obscure the rightful pre-eminence of mechanism, since if mechanism was only an aspect of the true natures of things as captured in the broader notion of substantial form, physico-mathematics could not be the centre of the 'true physics,' and Descartes indeed desired nothing less than the enshrining of a physico-mathematical vision at the heart of physics. What was needed was a metaphysics—a set of fundamental categories which motivated the privileging of mechanism and its geometric subject matter in the characterisation of physical reality and the explanation of physical behaviours. As such, joined to the Cartesian project of promoting mechanism to a true physics, was the creation of a metaphysics which put the object of mechanism—the physical object *qua* extended—at the centre of a philosopher's notion of physical existence.

One of the obvious costs of this project to Descartes, even before any metaphysical work is done towards its end, is that 'physico-mathematical' principles of the machine, while in themselves clear and distinct and with an impressive claim to modelling many phenomena, seemed on grounds as clear and distinct as the mathematical concepts themselves, to be unable to account fully for the qualitative, and particularly the *mental* aspect of reality which immediately confronts any thinker. As Descartes famously argued, even if all else that confronts the thinker is deception, the thinker at least cannot be deceived in that he thinks, and

in what he immediately is thinking (even if that thought has in turn no correspondence in external reality), and thinking in turn is (in many ways Descartes would explain) difficult to treat as part of extension. In Descartes' *Treatise on Man* (the second part of '*The World'*), Descartes already alludes to the 'rational soul' as a second component to his mechanical 'man,'[71] contrasted with and yet in charge of the 'machine.' Though the precise causal relations between the rational soul and the mechanical body in the *Treatise on Man* are lost, we can see already the germ of Descartes' later dualism (as in, e.g., the *Meditations on First Philosophy* (1641) and the *Principles of Philosophy* (1644)), showing the close connection between the project of mechanisation of nature and the production of the dualistic, radically independent 'soul.' In the *Meditations*, this stark division between the realms of extension and non-extension could thus be seen only as the divide between extension on the one hand, and the other completely indubitable foundation of experience—thought, which is also the origin of all impressions of non-extended qualities. Since Descartes is interested in affirming as his first principles only what is indubitable, clear and distinct, and in building upon these foundations only what follows from such first principles, he becomes willing to accept a place for the qualitative forms which he has eliminated from the physical world in the mind, where their existence as thoughts or ideas is indubitable. God, though he is not the focus of this thesis on dualism, is a third and perhaps more fundamental idea for Descartes, who regarded God's existence as just as self-evident as that of the mind and of mathematical truth, and the idea which, for Descartes, ultimately unifies all of reality. **Part 2** will address the particular arguments that Descartes offers for his dualism in detail. For now, however, it is sufficient to note that the project of deriving certain, clear and indubitable foundations for scientific knowledge has driven Descartes to two plausible candidates for such a foundation, each in its

[71] CSM I 101.

own way seemingly immediately evident to reason: 1) extension, to which all the behaviours of things with size, shape and motion might be reduced, and 2) thought, which grounds the very possibility of enquiry, and also (in its limitability in finite thinkers) the capacity for error and confused ideas of all sorts. In 'thinking substance,' immaterial substance is re-conceived around the central phenomenon of 'thought' (encompassing 'pure thought', imagination and sensation), and shorn of most of its hylomorphic baggage.

To the end of establishing what Descartes would call the two 'principal attributes' to which all scientific investigation of reality must refer, Descartes would establish a simpler 'substance-mode' ontology quite unlike the elaborate scholastic hierarchy of being. Being, in the scholastic hylomorphic synthesis, was capable of being qualified in as diverse ways as there were kinds of causes—there was a division between substantial being and accidental being, but also a hierarchy of substantial being distinguished according to the status of their form-matter constituents, ranging from purely formal angels at the top, mostly-material but partially-purely-formal humans, and matter-form composites such as ordinary living things at the bottom. Quantity, in the scholastic ontology, following Aristotle, was a ubiquitous accident of material things,[72] but also somewhat of a 'sideshow' in the order of being—it was an accident of substance rather than substance itself.

By contrast, Descartes' new substance-mode ontology treated only of two fundamental kinds of substance, or independently existing thing: finite substance, which was independent of everything but God, and infinite substance, which just is God. [73] Of finite substance, which was limited to have particular attributes or aspects under which they are intelligible as existing,

[72] Aristotle, 'Categories,' in Aristotle and Jonathan Barnes (ed.), *Complete Works of Aristotle, Volume 1: The Revised Oxford Translation*, Princeton University Press, 2014, pp.8–10.

[73] CSM II 210.

there were only two fundamental attributes: *extension* and *thought*, of which the various kinds of extension and thought were themselves particular manifestations. Extension and its associated modes were not a sideshow, but the fundamental basis of one of two basic categories of finite object. The advantage of this scheme is that it served as a framework which made extension and thought the two 'principal attributes' to which all description of finite reality must be referred. This legitimised the mechanistic foundations of the new physics as a science of things-in-themselves, in turn legitimising all the precision and predictive power of mechanistic methods, avoiding the baggage of scholastic metaphysical speculation and terms of disputation, while yet preserving what Descartes thought was theologically, experientially and existentially important—the immortality of the soul, the possibility of understanding, and the existence of God. The most obvious costs to Descartes of the mechanistic or physico-mathematical approach to physical nature could thus be dealt with by deriving a second realm of nature just as, if not more, accessible to the intellect—that of the intellectual nature itself. This division between thought and extension, stemming first from the desire to reify a 'physico-mathematical vision' of nature in aid of increased scientific clarity, empirical tractability and explanatory power, would be the most significant one in finite nature for Descartes, since for him, thought and extension were the starting points for conceptualising all of nature in a clearly intelligible form that left no remainder. **Part 2** shows how this deep division in nature which Descartes introduces, generates its own insuperable difficulties in the form of an interaction problem which cannot be resolved without undermining the whole framework of Descartes' dualism. As we shall see in **Part 3**, the enduring appeal of the 'physico-mathematical vision' of nature would outlast Descartes' own concerns with indubitable epistemic foundations and complete clarity and distinctness of ideas, and indeed would outlast even his commitments to the immortality of the soul which led him in a specifically *substance*-dualist direction. In the form of the modern 'Structure-and-dynamics argument from dualism,' something like the

Cartesian physico-mathematical vision continues to vex modern physicalism, and in turn exerts a conceptual pressure against the idea that everything is or can be reduced to the physical. This in turn generates a fundamental division in nature which is difficult or impossible to bridge merely with the tools used to construct the physico-mathematical vision. The division thus created is, as Descartes' own struggles with causality in substance dualism would show, difficult or impossible to square with a unified account of nature.

In order to come to a full appreciation of the significance of Descartes' moves for generating problems of dualism, it is necessary to remind us of what Descartes has lost by the removal of substantial form. Recalling **section 1.1**, The scholastic-Aristotelian synthesis, in the notion of Substantial Form (of which quantity and quality both were 'accidents'), had an *immanent* principle of nature which, in serving as a common source and site of complementary ontological functions, could unite the quantitative and qualitative aspects of reality in a way (in principle) accessible to the finite mind. Whatever the difficulty of such concepts in the face of sustained critique, the adherents of a system in which they featured could comfortably see quantitative and non-quantitative properties of systems as a complementary whole, in which the division between quantitative and non-quantitative properties, while real, is not overly ontologically significant or problematised. Quantitative and non-quantitative properties could be mounted unproblematically on a common 'metaphysical scaffold'. With the removal of such a metaphysical scaffold which is immanent to reality, the prospect of reunifying by means of human intellectual effort the halves of nature thereby created is rendered more remote and difficult. Descartes' God, as the ultimate source of both the principal attributes of thought and extension, is arguably, among other things, a remnant of the theoretical necessity for such a sense of ultimate unity and coherence—in God, at least, the two different orders of reality might have some explanation in principle for their harmony, and have some common source of their existence, even if it is a source remote and inaccessible to us finite reasoners. God, after

all, represents unlimited and unqualified being in Descartes' ontology,[74] and therefore at least a distant unifying principle for all reality to pick up the slack. For atheist intellectual descendants of Descartes, if indeed they are bound by an inherited vision of physical reality to run into similar problems of dualism, even that comfort is going to be cold indeed.

[74] CSM I 198–199.

Part 2. On Dualism and the Interaction Problem in Descartes

In **Part 1,** we examined the shift from the hylomorphism of the medieval scholastics, a metaphysics predicated on the central, complementary concepts of Matter and Form, to the dualism of Descartes. Having begun with a note about the complementary metaphysical functions which motivated the critical notions of 'matter' and 'form', in Part 1 I traced the gradual unravelling of the medieval hylomorphic synthesis and the pressures which led to that unravelling.

We have noted the theoretical attractions to Descartes of a mechanistic and 'physico-mathematical' notion of physical reality, the basic principles of which were mathematical and the central analogy of which was the machine, and how adopting such a notion of physical reality provided the theoretical impetus to decisively reject the hylomorphic synthesis (particularly in its Suarezian form) as an ontology of physical reality. Because not everything could, at least for Descartes, be plausibly assimilated to the physico-mathematical-cum-mechanistic vision, we noted that Descartes was well-motivated to produce a new dualistic synthesis, in which the difference between mind and matter is central. He addressed these motivations by the development of arguments for dualism, rooted in a new and simplified ontology which he aimed to show could encapsulate the whole of nature. We left Descartes with the question of whether he could tie together these halves of a newly-bifurcated nature together, in absence of the 'metaphysical scaffold' provided by hylomorphism.

Part 2 examines in detail Descartes' arguments for dualism and the difficulties his arguments faced in light of the rejection of the hylomorphic synthesis. In large part, the analysis of Descartes' arguments which follows will be sympathetic, showing how, given his starting points and the data available to him, much of what Descartes desires can be granted against

immediate objections, given the terms of his substance-mode ontology. However, it will be argued that even granting that Descartes has defensible avenues toward mind-body dualism from his starting points, the very success of that division, and the terms on which it is obtained and by which it fends off objections, entails the impossibility of interaction between mind and body. Part 2 will consist primarily in exegesis and evaluation of the arguments for Descartes' substance dualism provided in the primary texts.

S**ection 2.1** sets out some fundamental notions of Cartesian ontology, namely, his concepts of (principal) attribute, mode and substance, and gives particular emphasis to the 'one principal attribute' rule as the source of Descartes' most compelling dualistic results. I set forth the chief challenges to that rule which are dealt with in following sections. S**ection 2.2** shows that Descartes has ample resources to refute the objection that thought and extension might be 'modes' of each other, and hence that he has the resources to address the possibility that mind and body might not constitute independent substances. **Section 2.3** evaluates two arguments which Descartes offers for the principle that each substance can have only one principal attribute. By showing in sections 2.2 and 2.3 that Descartes' one principal rule is highly defensible given his epistemic starting points, I in turn show that the pressure toward dualism given his starting points is very strong. With a demonstration in hand that dualism does indeed plausibly follow from Descartes' starting points, **Section 2.4** turns on Descartes, showing that the very terms on which he secures his dualism make interaction between the two substances impossible, with catastrophic consequences for his philosophy. **Section 2.5,** having shown that Descartes' dualistic project is doomed to an irresolvable difficulty in the form of the interaction problem by the very conceptual moves which give rise to Descartes' dualism in the first place, reflects on how those problematic conceptual moves were themselves compelled by the loss of the 'metaphysical scaffold' provided by scholastic hylomorphism and the construction of an

inadequate one in its place. Part 2 concludes with a further reflection on the role such a 'metaphysical scaffold' plays in the occurrence of such 'problems of dualism' in Descartes.

2.1. Substances, Principal Attributes and Modes

Descartes' replacement metaphysical framework for the hylomorphic synthesis is found most clearly and systematically in the *Principles of Philosophy* (1644). In the *Principles,* Descartes divides the objects of thought and perceptions as either "things, affections of things or eternal truths which exist only in thought,"[75] by which he means substances, modes, and the common notions or axioms which undergird all reasoning. The defining characteristic of substance is ontological independence, in contrast to the intrinsic dependence of the mode. Thus for Descartes, God, who depends upon nothing, is the paradigmatic substance, whereas substances other than God, while not completely independent, are also substances in a sense, in virtue of depending only upon God's causal concurrence.[76] *Substances* are the subjects of properties (i.e., modes and attributes),[77] and lacking the kind of dependence properties have, exist independently (in the case of created substance, of everything but God). Substances are distinguished from *modes*, which depend not only upon God, but other creatures for their existence, particularly in the sense that modes, unlike the substances in which they inhere, cannot be thought of except as modifications or alterations of an underlying substance,[78] in which they thereby 'inhere.' Modes characterise substances as they are contingently or from time to time. Attributes are similar to modes, in that they are understood primarily as attributes of some existing thing allowing that existing thing to be characterised as belonging to a certain type, but are not alterations or modifications of that thing (as in the attributes of an unchanging

[75] CSM I 208.

[76] CSM I 210.

[77] CSM II 114.

[78] CSM I 211.

God, or the unalterable 'principal' attributes of created substances, as discussed below). For the purposes of contrasting secondary and principal attributes, however, Descartes often calls the latter simply 'attributes,' and the former simply 'modes,'[79] and that is the terminology that this section will adopt.

Substances, according to Descartes, are known as themselves through their *one principal attribute,* which 'constitutes its nature and essence, and to which all its other properties are referred.'[80] The principal attribute might be thought of as the 'master-attribute', of which all the various permutations of minds and bodies are instances, and through which, as providing the basis of all such permutations, the substance may be understood as a substance (i.e., as an
900 independent thing). Bereft of its principal attribute, the substance as such becomes unintelligible—as Descartes says, 'we do not have immediate knowledge of substances'.[81] Descartes holds that there are two kinds of principal attribute: 1) extension in length, breadth and depth, as pertains to *bodies;* and 2) thought, particularly abstract thought or pure intellectual content apart from imaginary representation and sensation,[82] which Descartes applies to *minds.* Because these principal attributes are exclusive, it follows that there must be two fundamentally different kinds of substance.

[79] CSM I 300.

[80] CSM I 210.

[81] CSM II 156.

[82] See, e.g., the *Second Meditation* at CSM II 22, 'I now know that even bodies are not strictly perceived by senses or the faculty of imagination but by the intellect alone, and that this perception derives not from their being touched or seen but from their being understood,' or the *Sixth Meditation* at CSM II 50–51, 'I can of course understand the figure of a pentagon, just as I can the figure of a chiliagon, without the help of imagination.'

Descartes' arguments for these two principal attributes trade on the impressive conceptual diversity of thought and extension. In the *Second Meditation* (1641), for instance, Descartes argues that introspection can show at least that the 'thinking thing' one is forced to identify with as a result of the Cogito, is not to be identified precisely with the physical body which occupies space and has a determinate shape and position, since all of those physical attributes are dubitable, while his own status as a thinking thing is not:

> I am not that structure of limbs which is called a human body, I am not even some thin vapour which permeates the limbs—a wind, fire, air, breath or whatever I depict in my imagination; for these are things which I have supposed to be nothing. (CSM II 18)

If we are careful, we will note that though Descartes' conclusion is that he is not to be identified with his human body, or any merely physical part of that body, it doesn't follow that the 'thought' with which he identifies himself is the *whole* of himself, i.e., it doesn't follow that thought exhausts all that he is and excludes all materiality whatever, or that thought is capable of constituting a distinct substance unto itself. Descartes himself notes immediately after the preceding passage,

> Perhaps it happens to be the case, however, that these very things which I suppose to be nothing, because they are unknown to me, do not in reality differ from that I that I know? I don't know, I don't dispute about this yet: I can only judge about those things that are known to me.[83]

While Descartes' arguments examining the ideas of thought and extension may be sufficient to show that such ideas are *conceptually* distinct, and may be thought of without each other, if Descartes is to arrive at substance dualism he is obliged to further supply a reason that thought

[83] Ibid.

and extension cannot be merely modes of the other, or that they cannot be co-principal attributes. This is a dialectical distance which Descartes does not explicitly traverse between the *Second* and *Sixth Meditation*; by the *Sixth Meditation,* Descartes has shifted from conceiving of thought and extension as distinct, such that one can be thought without the other, to thinking of them as two different kinds of substances which God, at least, could separate:

> Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God. The question of what kind of power is required to bring about such a separation does not affect the judgement that the two things are distinct. Thus, simply by knowing that I exist and seeing at the same time that absolutely nothing else belongs to my nature or essence except that I am a thinking thing, I can infer correctly that my essence consists solely in the fact that I am a thinking thing. It is true that I may have (or, to anticipate, that I ttribunly have) a body that is very closely joined to me. But nevertheless, on the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it. (CSM II 54)

In the argument above, Descartes' assumptions about the nature of principal attributes play an important role. In appealing to the notion that 'absolutely nothing else belongs to my nature or essence except that I am a thinking thing,' Descartes clearly relies on the idea that in reflecting on his thinking activity, he has access to the *whole essence* of the thinking thing—i.e., he has access to *thought*, understood as a principal attribute. If Descartes is correct that principal attributes exclude each other, and each unified substance can have only one such attribute, it is actually a trivial step from understanding an attribute to be a principal attribute to the

conclusion that the substances constituted by two such different attributes are in fact separable and independent, even if we don't happen to find them apart in common experience. If there are two different and exclusive principal attributes, then there must be two substances. Since neither substance could contribute anything of itself to the other, neither substance could owe its existence to the other, and they would be distinct and separable. Descartes' argument in the *Sixth Meditation* need not therefore turn on a controversial shift from mere conceivability to
960 metaphysical possibility, as in certain reconstructions of his argument.[84] Rather, the argumentative load is borne by 1) Descartes' underlying metaphysics of what it is to be a principal attribute, and what that implies about the differentiation of substances; and 2) his confidence that he has in fact correctly identified the principal attributes of the substance he is.[85]

If Descartes is to establish that they really are two separable substances corresponding to thought and extension, which do not owe their existence to each other, he is obliged to supply an argument that they really are both principal attributes, and that no substance can have more or fewer than one principal attribute. This ought to occur in two stages: 1) rule out the

[84] For a typical reconstruction which critiques Descartes' argument based on the move from conceivability to possibility, see David Armstrong, *The Mind-Body Problem: An Opinionated Introduction* (1999) Westview Press, pp.20–21.

[85] Michael Hooker, in his chapter, 'Denial of Mind-Body Identity,' in *Descartes: Critical and Interpretive Essays*, edited by Michael Hooker, The Johns Hopkins University Press, 1978, pp.171–184, disputes that it is indeed relevantly conceivable that one is a mind which exists independently of body, since for all Descartes shows in the *Second Meditation*, it remains also conceivable that one is a body, leading to the unpalatable (to Descartes) conclusion that one is a body. Reconstruction along the lines I have suggested makes Descartes' experience of thought 'as his whole essence' the key datum of the argument (from which he then derives the possibility of his independent existence via the one principal attribute rule), and is consistent with Descartes not yet having established the independence of mind 'as far as he knows' in the *Second Meditation*.

possibility that thought and extension are modes of each other or of some further attribute, and 2) show that, as principal attributes, they are exclusive of each other. These are arguments that he supplies in his other work, as will be examined below.

2.2. Thought and Extension are not Modes of Each Other or Some Further Attribute

Descartes argues against the possibility that thought and extension are modes of each other in the Fourth Set of Replies,[86] where he addresses an objection by Antoine Arnauld that 'adequate' knowledge of a thing's essence (which Descartes takes to mean 'knowledge which contains absolutely all the properties which are in the thing which is the object of knowledge'[87]) is required to establish that it is capable of independent existence. First, Descartes clarifies that 'adequate' knowledge is not required to establish the independent existence of a kind of substance. Rather, 'complete' knowledge is required—the knowledge required to understand its object as a 'complete' substance, i.e., a thing unto itself and not a mode of something else.[88] For Descartes, an attribute is known as that of a complete substance if it is that quality in virtue of which a substance possessing it may be understood as something existing in its own right. By contrast to the principal attributes of a complete substance, an attribute which is merely a particular determination of some more-fundamental attribute, cannot be that in virtue of which a substance exists in its own right.

An attribute is truly known for Descartes insofar as it is *clear* and *distinct* to the mind. By *clear* Descartes refers to a perception or idea which is 'present and accessible' to the attentive mind.[89] By 'distinct' Descartes refers to an idea which is both clear and so sharply distinguished from

[86] CSM II 155.

[87] CSM II 155.

[88] CSM II 156.

[89] CSM II 207.

all other ideas that it contains in itself only its clearly understood content.[90] Descartes takes thought and extension to be clear and distinct (i.e., well-understood) ideas, such that their entailments and dependences (if any) may be reliably examined, and any conceptual dependences reliably identified, even if there may be some degree of mystery for an insufficiently well-versed inquirer as to exactly how all dependent modes are related to the principal attributes. Recalling that 1) a mode or non-principal attribute is a mode precisely in virtue of its being implicitly dependent on some further principal attribute, and 2) combined with Descartes' convictions about the conceptual transparency of thought and extension, these two Cartesian premises yield an argument that thought and extension are indeed not modes of each other: If thought or extension were dependent in the way that modes characteristically are, an analysis of the ideas of thought and extension as clearly and distinctly present to one who considers them, would (if those ideas are indeed clearly and distinctly present) discover that dependence. Since such an analysis does not show such dependence, it must be concluded that thought and extension, as attributes of substances, cannot be modes of each other.

Arnauld raises the issue of whether the apparent principal attribute of thought, as it is occurs in human understanding, is sufficiently well-known that its being a mode of extension can be ruled out. Arnauld cites the idea of a right-angled triangle, from which it is not clear that the Pythagorean ratio of the hypotenuse to the sides applies.[91] Descartes responds that, while it is

[90] CSM I 207–208.

[91] 'Suppose someone knows for certain that the angle in a semi-circle is a right angle, and hence that the triangle formed by this angle and the diameter of the circle is right-angled. In spite of this, he may doubt, or not yet have grasped for certain, that the square on the hypotenuse is equal to the squares on the other two sides; indeed he may even deny this if he is misled by some fallacy. But now, if he uses the same argument as that proposed by our illustrious author, he may appear to have confirmation of his false belief, as follows: 'I clearly and distinctly perceive', he may say, 'that the triangle is right-angled; but I doubt that the square on the hypotenuse is equal to

possible to conceive of the right-angled triangle without understanding the ratio of its sides, it is impossible to coherently and explicitly deny that the Pythagorean ratio exists between the sides and affirm that the triangle is right-angled. Indeed, says Descartes, even if we could not see why the Pythagorean ratio could not be denied of a right-angled triangle, we could not understand the right-angled triangle without understanding its sides as having *some* ratio.[92] For Descartes, for a quality to be understood as a mode, i.e., as referrable to another, more fundamental quality, there has to be *some* indication that there is some further quality (which needn't be fully understood) without which that quality could not exist.

Descartes' test for how we can know whether some quality X, as clearly and distinctly understood, is a mode or a principal attribute, thus runs as follows: If X is a mode, then there is some further attribute Y (which need not be fully known), the existence of which cannot be explicitly denied without denying the existence of X. If X is not a mode, then X is a principal attribute. Thought and extension, by this standard, seem to be independent of each other rather than merely distinct as a mode and a principal attribute would be, since denying the existence of one does not seem to suggest the non-existence of the other, nor indeed are there any plausible candidates for more fundamental kinds of attributes which, if denied, entail the non-existence of thought or extension. Descartes, then, doesn't rely on a brute 'seeming' that thought and extension are radically distinct principal attributes—there is a test, and a plausible one given the conceptual relation of a mode to a principal attribute, which one can apply to the concepts to determine whether they are indeed principal attributes. By means of such a test, Descartes might have discovered, despite some initial impression to the contrary, that thought and extension might be related to some further attribute as the right-angled triangle is related

the squares on the other two sides; therefore it does not belong to the essence of the triangle that the square on its hypotenuse is equal to the squares on the other sides,' CSM II 141.

[92] CSM II 158.

to the ratio between its sides. However, it seems that, according to the test, they are not modes referrable to some further attribute, and therefore, if Descartes can know anything about either thought or extension, he has ample justification for believing that they are not modes of each other or anything else, and are *t*herefore principal attributes.

It is certainly open to Arnauld to insist, despite Descartes citing the evidence of the conceptual independence of thought and extension, that it is still barely possible that we don't know enough about the concept of thought or extension to know whether mind as conceptually independent from body (and vice versa) is indeed coherent. Such a sceptical conclusion would amount to a challenge to the proposition that we have such a clear and distinct idea of thought or extension as Descartes takes himself to have demonstrated to exist in, e.g., the *First* and *Second Meditations*. Descartes could well insist that Arnauld owes him a powerful argument showing some respect in which either the qualities of thought or extension must be (or even might plausibly be) referred to the other or to some further property. Barring such an argument, it seems that, resting on the impressive conceptual diversity of thought and extension, Descartes has good entitlement to his conclusion that thought and extension are not, as far as anyone can show, modes of each other or anything else, and are therefore very plausibly principal attributes. The issue of whether Descartes' defence of the idea that thought is not a mode of extension or vice versa is successful outside his immediate dialectical context will be taken up later.

2.3. Two Arguments for Thought and Extension as Exclusive Principal Attributes

Supposing Descartes to be successful (at least in his own dialectical context) in demonstrating that thought and extension are not modes of each other, it is then possible to examine the two independent arguments Descartes offers for the conclusion that thought and extension are exclusive principal attributes of substances. They are, 1) what I shall call the argument from

'non-identity' found in the *Comments on a Certain Broadsheet*,[93] and 2) the argument from divisibility and indivisibility.

2.3.1. The Argument from Non-Identity

In the *Comments on a Certain Broadsheet* (1648), Descartes explicitly entertains the problem of a single substance possessing two different principal attributes. Against the idea that thought and extension 'are not contraries, but merely different,' Descartes argues that especially when it comes to the essential or principal attributes of a substance, to be different is to be contrary as much as are 'is' and 'is not,' such that affirming two different principal attributes is to affirm a contradiction, at least in the case of simple substances,

> For, when the question concerns attributes which constitute the essence of some substances, there can be no greater opposition between them than the fact that they are different: when he acknowledges that one attribute is different from the other, this is tantamount to saying that the one attribute is not the other; but 'is' and 'is not' are contraries…
>
> …As for the attributes which constitute the natures of things, it cannot be said that those which are different, and such that the concept of the one is not contained in the concept of the other, are present together in one and the same subject; for that would be equivalent to saying that one and the same subject has two different natures- a statement that **implies a contradiction**, at least when it is a question of a simple subject rather than a composite one.[94] (CSM I 298, emphasis added)

[93] Also identified as the 'Argument from Real Distinction' in Marleen Rozemond, 'Descartes' Case for Dualism' (1995), *Journal of the History of Philosophy;* 33(1), pg. 29. I have named it differently because the present account does not focus on Descartes' theory of distinctions.

[94] CSM I 298.

At first this special feature of principal attributes as excluding what is different seems puzzling. Descartes himself accepts that different *modes* are not contraries of each other, [95] so it may be difficult to see what sets principal attributes apart from other qualities in this respect. While Descartes does not explicitly justify this special status of principal attributes within the *Comments,* the nature of the principal attribute itself perhaps supplies a clue. Recalling the Fourth Replies, in particular his reply to Arnauld, principal attributes for Descartes constitute the nature of a *complete* substance, 'an entity in its own right which is different from everything else',[96] while modes are particular specifications of or manifestations of that attribute, which refer back to that attribute. If a substance had more than one alleged principal attribute, then neither of its multiple principal attributes could be the *complete* nature of a self-subsistent entity, through which it is understood as complete: they would at best only *partially* describe the nature of such an entity. But a partial description of the nature of a substance, cannot be in itself that substance's nature, and the attribute limited to such a partial description, could not therefore be a principal attribute. The true nature of a single substance which seems to have more than one principal attribute, i.e., that nature which makes intelligible its existence as one substance and unifies its properties as belonging to one substance, would have to be some 'third attribute'. Insofar as it must completely constitute a substance's most basic nature, such a principal attribute cannot tolerate a rival non-identical attribute with an equally comprehensive claim to characterise the substance. Hence thought and extension, if they are principal attributes, must be exclusive, from which it follows that substances are indeed divided into at least these two fundamental types.

[95] Ibid.

[96] CSM II 156.

Modes, on the other hand, being by their very nature limited modalities of the substance in which they inhere, can happily coexist with other partial descriptions of a single substance. Descartes does in his *Comments* qualify his defence of the one principal attribute rule to the case of 'simple' substances, which are not composed of more-fundamental substances.[97] A non-simple substance, presumably, could have multiple principal attributes, but only in virtue of not being a true unity, or a truly singular substance unto itself.[98]

It might be objected that the substance itself could serve the function of a principal attribute—it could be its own 'complete nature,' and yet answer to no description more fundamental than the inescapably varied principal attributes applied to it. Descartes would find this result unacceptable, indeed unintelligible. The principal attribute is precisely that through which the substance is known as itself. To deny a single unifying, distinguishing attribute, is to deny any ground of unity of the substance, and hence, to deny any ground for positing a single substance which underlies the attributes at all. Multiple conceptually independent 'principal attributes' could only be attributed to a single substance for Descartes, if their status as 'principal' were only apparent, and a more fundamental attribute hypothesised. With a lack of any good candidate for such a more-fundamental attribute of which thought and extension are specifications, or any compelling reason for positing such an unknown third attribute,

[97] CSM II 298 at 350.

[98] The reconstructed Cartesian argument for the 'one principal attribute' rule has deep parallels with Thomas Aquinas' arguments for the unicity of substantial form. For Aquinas, substantial form, like the principal attribute, is that which 'makes a thing *to be simpliciter*,' unlike accidents (compare modes) which 'make a thing *to be such*' (ST I Q 76 Art 4), and hence, for a thing to have different substantial forms is for there to be two different beings simpliciter, rather than two ways for a single being to be. In this comparison, 'substance' functions analogously to prime matter, an identity-less principle lacking definition which nevertheless grounds and individuates the attributes which could not exist 'on their own.'

Descartes could feel justifiably confident in concluding that mind and body are indeed principal attributes, and *therefore* correspond to distinct substances.

One type of reason to posit some more-fundamental attribute, even if we do not know its true nature, is considerations of parsimony, considerations with which Descartes was quite familiar.[99] In one sense, positing two substances, mind and body, where some kind of mind-body unity might in principle do, seems unparsimonious. It might be theoretically better, if at all possible, to treat mind and body as aspects of some single substance. Yet parsimony cannot be used to urge ad-hoc unity between things which are otherwise to all appearances diverse. Ad-hocness is a theoretical defect, and parsimony is only relevant between explanations of equal theoretical virtue. While Descartes' substance dualism requires only those ideas which are clearly and distinctly present to the mind, and the commitment only to such substances as the attributes themselves suggest, to posit a substantial unity of mind and body via an unknown principal attribute is to simply *invent* a ground of unity where there otherwise would be no warrant for it, and much warrant for the reverse. Thought and extension, after all, appearing for all the world as different principal attributes, have less in common than most things Descartes might encounter. Being so mysterious, the explanatory power of this third attribute is not only poorly motivated and even more superfluously posited than dualism, but extremely limited in explanatory power compared to substance dualism, which by contrast easily explains the conceptual diversity and apparent irreducibility of thought and extension. Since positing a mysterious third attribute is so ad hoc and so much less explanatorily adequate, Descartes might well conclude that dualism, for all that it superficially appears less parsimonious, has the better

[99] For example, the final appeal of the *Treatise on Man* (CSM I 108), which is an attempt to render plausible a mechanistic view of reality, makes appeal to the few and invariable mechanistic principles over and above the extravagant notion of vegetative or sensitive soul.

of the principles of parsimony than the proposal of some mysterious neutral third neutral ontological principle.[100]

An anti-substance-dualist might thus be tempted to concede Descartes' thesis that positing an entirely novel, mysterious third 'neutral' principle is metaphysically and methodologically implausible. Instead, it might be contended that the third attribute which constitutes a single substance need be no more than the conjunction of thought and extension. Let us call this candidate attribute '*the conjunction of thought and extension'* (CTE). This would have the advantage of ontological parsimony (at least, relative to Descartes' substance dualism), since it avoids positing two fundamentally different kinds of substance, has some explanatory value for the co-occurrence of thought and extension in human experience, and is moreover not wholly mysterious—it is possible to see why thought and extension would be partial descriptions of this attribute.

While Descartes did not explicitly consider this avenue of attack, he is not without recourse. If CTE were a principal attribute, extension and thought in themselves would have to be, at least where CTE is instantiated, modes of CTE, just as dimensions and shapes are modes of extension, and desire, understanding and imagination are modes of thought. However, whereas the respective modes of thought and extension seem intrinsically to refer to the underlying attribute of which they are modes as a matter of conceptual necessity, no such conceptual necessity for a further underlying attribute seems to apply to thought and extension themselves.[101] That is, the concepts of thought and extension alone do not suggest that there is

[100] Dutton has a less positive assessment of Descartes' one principal attribute rule in Blake Dutton, 'Descartes' Dualism and the One Principal Attribute Rule' (2003), *British Journal for the History of Philosophy*, 11(3), pp.395–415, to which this account is something of an answer on behalf of Descartes.

some more fundamental attribute to which thought and extension may be referred, as thought and extension can be thought of not only apart from each other, but also can be thought as existing consistently with denying that anything has the CTE. It is important to note that in order to show that thought and extension are *necessarily not* modes of CTE, and thus that CTE is necessarily not instantiated, it is only necessary to show that either thought or extension *may* be instantiated without CTE. For if an attribute *can* be instantiated without reference to a more fundamental quality, then that attribute is, considered in itself, 'complete'. Since the features of an attribute considered in itself and apart from everything else, just are the features of that attribute in every possible contingent state of affairs, if an attribute is possibly 'complete,' then it is necessarily complete. If modes thus necessarily imply the existence of a principal attribute to which they refer, and neither thought nor extension imply the existence of the conjunctive attribute, then they *cannot* be modes or secondary attributes of a substance, and therefore CTE, which suggests that they are just such modes, cannot be a principal attribute. This argument may be put as follows:

1. If CTE (the conjunction of thought and extension) is a principal attribute, then thought and extension are possibly modes of CTE
2. Neither thought nor extension are possibly modes of CTE

   2.1. If some quality X is possibly a mode of Y, then X cannot be coherently affirmed while some attribute Y is denied

   2.2. Thought can be coherently affirmed while CTE is denied

   2.3. Extension can be coherently affirmed while CTE is denied

   Therefore,

   2.3. Neither thought nor extension are possibly modes of CTE.

3. Therefore, CTE is not a principal attribute.

Any refutation of this defence, at least within a Cartesian substance-mode ontology, must directly contend with the impressive apparent conceptual independence of thought and extension, and attempt to demonstrate the kind of ontological incompleteness necessary for them to be modes or secondary attributes of some further principal attribute. To unify two apparently diverse attributes which do not seem to be modes of a further attribute, as modes of a principal attribute, it is not sufficient merely to posit the conjunction of those attributes on an ad-hoc basis. It is necessary to strongly motivate a novel third attribute by means of some indication of ontological incompleteness in thought or extension. Without meeting this challenge, CTE cannot supply what the substance monist needs to meet his burden of plausibly unifying the apparently diverse attributes of mind and body.

In the argument from non-identity, then, Descartes powerfully leverages both the striking conceptual diversity of thought and extension, and the concepts of thought and extension as principal attributes in which the modes and qualities of common experience inhere, to motivate his dualism, denying the monist conceptual space in which to formulate an alternative to dualism (since all qualities a substance may have seem referrable either to extension or thought) and putting the monist in the position of having to invoke poorly understood and mysterious principles of unity to fuse mind and body together on an ad-hoc basis. While Descartes cannot completely excise the bare possibility that he is mistaken about the substantial diversity of mind and body, bare possibilities are cheap, and he can make it very difficult to give any weight to that possibility.

The strongest motivation to posit a deeper unity between mind and body than Descartes allows, of course, is the fact of their interaction. Given that they are substances of fundamentally different kinds, utterly independent of the other in their existence, it seems completely obscure

how one contributes to changes in the other. Since this difficulty is a key means of meeting Descartes' challenge to substance monism, we shall defer for now a critical examination of this consideration while we consider the argument for dualism from divisibility and indivisibility, and the objections peculiar to that argument.

2.3.2. The Argument from Divisibility and Indivisibility

In the *Sixth Meditation*, Descartes offers the following argument for the real distinction (i.e., the independent existence) of mind and body.

> The first observation I make at this point is that there is a great difference between the mind and the body inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible, for when I consider the mind, or myself insofar as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete… As for the faculties of willing, of understanding, of sensory perception and so on, these cannot be termed parts of the mind, since it is one and the same mind which wills, understands, and has sensory perceptions. By contrast, there is no corporeal or extended thing that I can think of which in my thought I cannot easily divide into parts; and this very fact makes me understand that it is divisible. (CSM II 59)

What Descartes means by 'divisibility' in the case of extension is clear enough from what he says of extended substance in the *Principles*,

> And we can also be certain that, if it exists, each and every part of it, as delimited by us in our thought, is really distinct from the other parts of the same substance. (CSM I 231)

Bearing in mind that 'real' distinction is the capability in principle of separate existence as a substance, the difference between divisible substance and indivisible substance is the capacity or lack thereof for the division of a substance into lesser, but no less independently-existing

substances. The modes of extension which pertain to divisible, extended substance can and must be realised in a substance which is a concatenation of multiple in-principle independently subsistent, non-identical (to the whole), extended parts.

The indivisibility of mind, by contrast, seems to arise from the fact that all the various modes of thought are modes of the individual mind as a whole rather than potentially-subsistent 'parts.' Adopting different modes of thought does not, from the clear and distinct idea one has of thought, seem to imply further sub-thinkers in the way that variation along a dimensional axis implies potentially- independently-subsistent, in-principle separable parts. More strongly, the modes of thought are not implicitly unified with other modes by extension among potentially-separable 'sub-thinkers' thinking alongside each other, but rather seem to require union in virtue of something simply singular, which cannot be extended, and therefore, a non-divisible substance identified with thought, *mind*, must be posited to account for their unity.

If mind is truly indivisible, then the independence in principle of mind from body is established, since extended things are typically destroyed by division into more base components, but an indivisible substance, if it goes out of existence, cannot do so on the same grounds. Having different grounds for their continuing existence, mind and body must accordingly be independently-existing substances.

The first objection which might be raised is that mind seems to admit of extension at least in respect of duration. If mind is divisible according to duration, then it seems that mind and body cannot be distinguished along the lines of indivisibility. Descartes, in responding to this issue as it was raised by Burman,[102] answers that the kind of extension constituted by duration is not the kind of extension which pertains to nature. That is, duration is not the kind of extension which seems to multiply potentially-subsistent substances, as extension does for its constitutive

[102] CMSK III 355.

parts. Extension in time is for Descartes distinguished from spatial extension precisely in virtue of the non-co-existence of temporal parts,[103] but it is precisely the property of extension which requires the simultaneous co-existence of potentially-independent, hence really distinct parts, which makes it divisible. So, extension in time cannot, for Descartes, make mind divisible.[104] Descartes thus allows change to mind while avoiding the implication of divisibility which might follow from granting mind a form of extension.

Secondly, it might be objected, as Dutton does,[105] that we may distinguish between senses of the indivisibility of the mental: 1) a weak sense, in that mind is not divisible just insofar as it is mental, and 2) a strong sense, in that mind, just insofar as it is mental, is indivisible. Descartes according to Dutton can only justifiably hold that the idea of thought is *different* from the idea of extension and its mode of divisibility, and that therefore only the *weak* sense of indivisibility is justified. If only this weak sense of the indivisibility of mind applies, then for all that mind is not divisible just insofar as it is mental, it might still be divisible in virtue of its being conjoined with an extended body. However, Descartes' case can very plausibly be put in a much stronger form than Dutton credits it—Descartes need not only be arguing that mind is distinct from extension (and thus divisibility), he could well argue that the way in which the modes of thought are unified would be *undermined* if they were really being thought by really distinct, separable and therefore divisible sub-thinkers. To affirm that there are such sub-thinkers as constituents of thought, would raise a dilemma—either there is a 'superior' thinker, over and above the sub-thinkers, in whom the sub-thinkers' thoughts are unified, in which case this is the indivisible thinker Descartes aims at, or there is no such superior thinker, in which

[103] CSM I 200.

[104] Modern concepts of perduring 4-dimensional 'worms' would, it seems, generate a problem for Descartes in this vein, but it would be beyond the scope of this section to delve into the issues raised in detail.

[105] Dutton 2003, p.413.

case it seems that one is forced to deny the Cogito. Descartes could thus quite plausibly argue that indivisibility does follow from the characteristic union of modes of thought, and thus that thought, precisely as thought, requires instantiation in indivisible substance—precisely the strong sense of indivisibility which Dutton thinks Descartes needs.

It is still possible, again, to raise the bare possibility that Descartes is mistaken. Descartes, in rooting his argument in the conceptual analysis of ideas, is always vulnerable to such bare possibilities that his analysis of ideas does not sufficiently correspond to reality. But as with the similar objections raised to the non-identity argument, it is difficult (outside of the very important objection to follow) to give these objections theoretical motivation to overcome the impressive conceptual diversity of thought and extension, and the ways in which Descartes' analysis of the ideas sets each implacably opposed to the other. Absent such a motivation, one buys deference to bare possibilities only at the price of crippling scepticism.

As Arnauld points out in the Fourth Replies, however, there remains a powerful motivation for postulating a deeper unity of mind and body than Descartes allows. Arnauld argues that being in some sense divisible can serve as a good explanation of the extinguishing of the rational faculties in madmen and the gradual awakening of such faculties in infants.[106] Mind, in these senses, seems to 'come into its own' in piecemeal fashion, and this seems inexplicable if mind did not have something in it corresponding to the characteristically piecemeal contribution of extended substance to bodily growth. The problem of how mind's modes correspond to the behaviour of matter, when its naturally indivisible, unified character might lead one to expect otherwise, is nothing other than a raising of the interaction problem, which is a powerful reason, rooted as it is in universal human experience, to suspect some unifying factor which Descartes' dualism of substances cannot allow. It is to this problem that the present section now turns.

[106] CSM II 160.

2.4. The (Mostly) Unsolvable Problem of Interaction

The difficulty raised by the interaction problem was not lost on Descartes. Princess Elisabeth of Bohemia raises the issue forcefully in her letter of 6 May 1643,[107] where she notes that the ability of physical things to affect each other is premised upon their common physicality—their ability to interact as extended things in space, and to impart impulses to each other as items moving in space:

> So I ask you please to tell me how the soul of a human being (it being only a thinking substance) can determine the bodily spirits, in order to bring about voluntary actions. For it seems that all determination of movement happens through the impulsion of the thing moved, by the manner in which it is pushed by that which moves it, or else by the particular qualities and shape of the surface of the latter. Physical contact is required for the first two conditions, extension for the third. You entirely exclude the one [extension] from the notion you have of the soul, and the other [physical contact] appears to me incompatible with an immaterial thing. This is why I ask you for a more precise definition of the soul than the one you give in your Metaphysics, that is to say, of its substance separate from its action, that is, from thought.

It seemed to Elisabeth that by denying to mind precisely those features which allow extended substances to interact with each other, Descartes obscured the sense in which mind could, as with common experience, act as a cause. Without some ground of mutual causality (except their common causation by God, whom Descartes seems reluctant to invoke to solve this issue), Descartes has a real difficulty on his hands.

[107] Princess Elisabeth of Bohemia and René Descartes, *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*, translated by Lisa Shapiro, University of Chicago Press, 2007, p.62.

Descartes' attempts to meet this difficulty within the bounds of his metaphysic are disappointing. In his first reply to Princess Elisabeth, Descartes imagines the contribution of the soul as akin to the infusion of tendency-qualities, such as *heaviness*, into matter, in an account resembling nothing so much as Aristotelian formal causation:

> For example, in supposing that heaviness is a real quality, of which we have no other knowledge but that it has the power to move a body in which it is toward the centre of the earth, we have no difficulty in conceiving how it moves the body, nor how it is joined to it; and we do not think that this happens through a real contact of one surface against another, for we experience in ourselves that we have a specific notion for conceiving that; and I think that we use this notion badly, in applying it to heaviness, which, as I hope to demonstrate in my Physics, is nothing really distinct from body. But I do think that it was given to us for conceiving the way in which the soul moves the body.[108]

While Descartes denies that this image is true in the literal sense, since his project of the mechanisation of material reality is concerned precisely with reducing all such qualities of movement, including that attributed to heaviness, to extension and its modes, he nevertheless maintains to Elisabeth that such are the notions appropriate to understanding the interaction (as far as it may be understood) of mind and body. Elisabeth is famously underwhelmed by this solution, which seems wholly mysterious, and confesses that she would find it easier to attribute extension to the soul as to enable its interaction with the body than to accept an occult process of quality-transfer.[109] Likewise, Elisabeth brings up precisely the dependence and subjection of the faculties of the immaterial soul to material interference as an example of

[108] Ibid., p.66.

[109] Ibid., p.68.

something inexplicable if the immaterial stands so far apart from matter as Descartes contends.[110] Descartes' further reply attempts to assuage Elisabeth's worries by admitting that the manner of the union of mind and body is indeed, given their substantial distinction and lack of common modes, obscure.[111] Taking up her willingness to attribute extension to mind, Descartes encourages Elisabeth to think of the conception generated by the attribution of extension to mind as nothing other than the notion of the union of body and mind, extension and thought.[112]

From what has been already noted of Descartes' substance-attribute-mode metaphysics, Descartes' recommendation to Elisabeth cannot be regarded as anything but an elision of the issue. For Elisabeth to literally attribute extension to mind (or mind to extended substance), would be to regard extension as a mode of thought. But, for reasons given above in Descartes' reply to Arnauld, this is impossible to clearly conceive as a unity given the way that Descartes understands thought: it remains of the essence of thought and extension that the one may be denied without denying the other, hence they are each, relative to each other, complete and independent and as such, not modes. The respective modes of thought and extension, moreover, are exclusive to each, since a mode is merely a determinate form that the principal attribute may take, and different principal attributes, as previously established, exclude each other. At best, for the thoroughgoing Cartesian, one may conceive the substances in close association *alongside* each other, but this would not amount in itself to a causal relation clearly conceived. Descartes, then, is finally forced to invoke a mysterious connection between diverse substances to save the appearances. Having invoked this mysterious connection, substance dualism is much less attractive as a view compared to its rivals, since the mysteries of how seemingly

[110] Ibid.

[111] Ibid., p.70.

[112] Ibid., p.71.

independent attributes may constitute one substance are no more impenetrable than how seemingly independent substances with different principal attributes may form a causal nexus.

The situation of Cartesian dualism is actually worse than that it is merely 'just as mysterious' as its rivals. The metaphysical moves by which Descartes motivates his dualism make interaction, by Descartes' own lights, positively impossible given the intrinsic limits of mind and body. Let us first recall that Descartes motivates the dualism of substances, rather than dualism of modes, by reference to the One Principal Attribute principle. That principle entails that a substance has or can have only those modes or attributes which are possible modifications or specifications of its 'principal' attributes, which unify it as a substance. Substances with each of these principal attributes cannot even *potentially* have modes of the other—modes of *thought* are not even potentially determinate forms of *extension*, and vice versa. Since Descartes thinks of causality as the production by the cause in the effect of what the cause has 'formally' or 'eminently' in itself (in turn motivated by the principle that out of nothing, nothing comes),[113] it would seem that the very driver of Descartes' dualism of substances, the complete self-sufficiency of thought and extension, also precludes causal interaction between these substances. There are simply not enough common principles left in virtue of which thought and extension could in principle be unified enough for a causal relationship.

Descartes in responding to this argument might invite us to more closely examine his definitions of 'formal' and 'eminent' containment. ***Formal*** possession of a quality is for the object of an idea to possess a quality which 'exactly corresponds' to the idea we have of it. ***Eminent*** possession of some quality is where, although the quality as it is in the object is 'does not exactly correspond to our perception,' is so great (i.e., in degree of reality) that it can 'fill

[113] CSM II 28.

the role' of that which does so correspond.[114] Mind cannot have extension as a mode (and hence, could never have extension formally), and hence cannot have extension of its substance, but it might *generate* modes of extension, and thus, in extension *of its power*, be said to 'fill the role' of the modes it causes, and thus contain extension 'eminently.' It is not incoherent to say that mind has such power, so mind can in principle 'eminently' contain material effects.

Logically, such a response runs the risk of vicious explanatory circularity: An entity is capable of causing its effect, in virtue of possessing that effect eminently, and an entity can contain its effect eminently, in virtue of being both of higher degree of reality, and able to *cause* it. To break out of this circle, Descartes might appeal to a 'reality-adequacy' principle: An entity can contain its effect eminently, just in virtue of possessing a higher degree of reality. So any substance, being of a higher degree of reality than a mode, contains 'eminently' *all possible modes*, and can therefore contribute modes to other substances. Descartes appears to contemplate just such a move in the *Third Meditation*.[115] Despite this, it is not clear that having a greater degree of reality, which simply means that it is more of an entity 'in its own right,' entails that a substance in any sense 'contains' all possible modes. If there is nothing more to 'eminent containment' of all possible modes than being a substance, then it is not clear how causation by a substance in virtue of the modes it eminently 'contains' could be distinguished from causation ex nihilo (i.e., without pre-containment at all).[116] If causation by a substance is

[114] CSM II 114.

[115] 'As for all the other elements which make up the ideas of corporeal things, namely extension, shape, position and movement, these are not formally contained in me, since I am nothing but a thinking thing; but they are merely modes of a substance, and I am a substance, it seems possible that they are contained in me eminently,' CSM II 31.

[116] Note, causation 'ex nihilo' here does not refer to the causality of God, who does have 'somewhat more' than mere ontological superiority to his effects.

not distinguished from causation ex nihilo, it is further unclear how the metaphysical principle, 'out of nothing, nothing comes,' which motivates the whole causal principle, is preserved. It does not follow, then, that simply having a 'higher' degree of reality, entails having a degree of reality 'so great that it can fill the role' of its effect. Descartes needs somewhat more than ex nihilo 'causal' efficacy and a 'higher degree' of reality if he wishes to retain his 'reality-containment' notion of causal adequacy.[117] Denying the reality-containment notion of causal adequacy would, in Descartes' mind, not only imperil the principle that out of nothing, nothing comes, but his commitment to God's existence and everything which follows from it.

There are in Descartes' metaphysics limited ontological functions which could constitute this 'somewhat more' which the cause of an effect may have. There is the pure, unlimited substantiality of God, of whom both extended and finite mental substance might be thought a limited approximation in different respects. It is God's unlimited-ness, which implies that all other substantiality is some limited approximation of his, which suggests a way in which God eminently contains his effects. It is possible that it was this kind of substantiality which Descartes in the *Third Meditation* was attributing to himself in absence of God as capable of eminently containing the elements of material reality which he conceived in his ideas. Such an option, however, is not available to finite substances.

Secondly, it is possible to consider a substance to 'eminently contain' the modes which it may take on. Between material substances, for instance, the modes which each can 'impart' to another are imparted or changed through changes in spatial contiguity, which unifies diverse extended substances into new substances, or divides singular extended substances into multiple

[117] Geoffrey Gorham, 'Descartes' Dilemma of Eminent Containment' (2003), *Dialogue: Canadian Philosophical Review/Revue Canadienne de Philosophie*, 42(1), pp.3–25.

substances, in the process combining or dividing modes, and educing new modes from extended substance. As the kind of substance possessing both (as a substance) a higher reality than the modes it generates, and a way in which it intelligibly produces such modes (i.e., being determinable to particular modes through combination and division), extended substance considered apart from its particular formal modes provides an intelligible way to describe a substance as 'eminently' containing its effects. On this model of causal adequacy, a finite substance's range of possible referrable modes maps exactly onto its capacity for eminent containment of modes, and distinguishes the production of new modes of the substance from creation ex nihilo. This, however, would imply, as argued before, that causal interaction between two different kinds of substance constituted by two different principal attributes is impossible.

There is a further class of options for the idea of eminent containment is available to Descartes. On the model of how the modes of material substance are combined, divided and altered, it might be that certain modes of mind and body are not eminently present in either mind and body considered severally, but can be educed from them considered as a 'unity' or a 'totality.' This would provide a sense in which at least some modes of mind and body could arise out of their interaction. Given Descartes' ontology, however, they can only be 'united' in a few ways, grouped according to the three degrees of reality which Descartes allows.

Firstly, they might be united by means of common modes, which 'straddle' two substances.[118] This is ruled out by Descartes' metaphysics of modes: If a mode is always a mode-of-a-substance, and if there is *no 'third substance'* corresponding to a third principal attribute

[118] Paul Hoffman, 'Union and the Interaction of Mind and Body,' in *A Companion to Descartes*, edited by Janet Broughton and John Carriero, John Wiley & Sons, 2007.

bridging mind and body of which thought and extensions are modes,[119] the only substances of which there could be modes are the two substances severally, and hence, there can only ever be two modes-of-a-substance, not one straddling mode. This, in turn, would leave the two substances causally isolated.

Secondly, mind and body might be united as a single finite substance (of which thought and extension, if unified in such a substance, would have to be modes), which Descartes' doctrine of the one principal attribute prohibits, as argued in Section 2.3.1.

Thirdly, they might be connected by infinite substance: thought and extension might share the common factor of being created by God, and their apparent interaction is due either to God's direct intervention in each instance, or to a pre-established harmony which God ordains. The difficulty with this option is precisely that it fails the principle of causal pre-containment, hence it cannot be said that mind or body or even the two together eminently contain the effect: it is only God who does so, and only God who is the causal actor at the occasions where mind and body interact. The prospects are dim for any real unity of mind and body which might be said to eminently contain effects attributable to diverse principal attributes interacting.

A final alternative we might consider on Descartes' behalf would be to consider a model of occasional causation (not to be confused with occasion**alist** causation, mentioned above), as

[119] Even for Cartesian 'trialists' like John Cottingham in 'Cartesian Trialism' (1985), *Mind*, 94(374), pp.218–230, the third principal attribute of the 'union of mind and body' does not give rise to a third substance, and therefore must be a kind of mode, but this is ruled out because modes are always modes *of* some principal attribute, and principal attributes are possessed exclusively by substances. If we were to go for a 'strong' trialism, and posit three substances- *res extensa*, immaterial *res sensus* and immaterial *res cogitans*-we would only intensify the interaction problem.

suggested by Nadler.[120] Hitherto, we have been considering models of causation in which the cause in a sense communicates some feature it has either formally or eminently to its effect (which may be an internal effect, e.g., thoughts produced by the mind), which has that feature formally—Nadler calls this 'efficient' causation (which when exercised by one substance on another is 'transeunt' causation), which he contrasts with 'occasional' causation. On an occasional-causal account, changes in mind or body act as 'occasions' upon which the other is
1460 'induced' to act upon its own intrinsic efficient-causal powers.[121] The occasional cause A of some effect B does not exercise efficient causation, through which A communicates something of itself to B, but A is nonetheless, for Descartes, a kind of real cause. This seems to have been a real feature of Descartes' mature views on the interaction of mind and body.[122]

Occasional causes are similar to Humean 'constant conjunctions' in the sense that the occasional cause and its effect are not linked by any intrinsic feature of the cause in virtue of which it produces the effect (and are to that extent 'accidental'), but are distinguished from Humean conjunctions in that there is a genuine explanatory principle for the lawlike conjunction, i.e., in a divine creative decree. Occasional causes are distinguished from occasionalist causation in that the efficient causal powers 'induced' to act by the occasion of a

[120] Steven Nadler, 'Descartes and Occasional Causation,' in *Occasionalism: Causation among Cartesian*, Oxford University Press, 2010, pp.29-47.

[121] Ibid. p.33.

[122] Descartes allows that occasional causes are genuine, albeit 'remote and accidental' causes, "Something can be said to derive its being from something else for two different reasons: either the other thing is its proximate and primary cause, without which it cannot exist, or it is a remote and merely accidental cause, which gives the primary cause occasion to produce its effect at one moment rather than another. Thus, workers are the primary and proximate causes of their work, whereas those who give them orders to do the work, or promise to pay for it, are accidental and remote causes, for the workers might not do the work without instructions." CSM II 304.

physical or mental change are not in God, but in the counterpart substance (hence, as Nadler says, occasionalism is a species of occasional cause). For an occasional-causal relation to obtain between mind and body, God merely has to decree a lawlike correlation between changes in some material body and changes in mind and vice versa, rather than be induced to intervene to produce effects in every instance (as in occasionalism). By 'outsourcing' the coordination of cause and effect from the intrinsic nature of mind and body to an external divine decree, an occasional-causal account provides, arguably, a kind of 'causal' nexus between mind and body that does not require 'efficient' causality or indeed any kind of 'real' union between mind and body (in that, there is no sense in which mind and body are united by any immanent feature). The question that remains, then, is whether this account of 'interaction' ameliorates any internal or external costs of dualism on Descartes.

Internally, Descartes requires an account of interaction which is capable of explaining the existence of ideas of sensory perception which are not a product of the intellect acting alone, such that God is not a deceiver,[123] and which is also capable of grounding a sense of special union with the body, as betrayed in, e.g., sensations of hunger and pain, also in the *Sixth Meditation*.[124] Occasional causation and the extrinsic principles of coordination it relies on are

[123] "But since God is not a deceiver, it is quite clear that he does not transmit the ideas to me either directly from himself, or indirectly, via some creature which contains the objective reality of the ideas not formally but only eminently. For God has given me no faculty at all for recognizing any such source for these ideas; on the contrary, he has given me a great propensity to believe that they are produced by corporeal things. *So I do not see how God could be understood to be anything but a deceiver if the ideas were transmitted from a source other than corporeal things*. [emphasis added]" (CSM II 55)

[124] Nature also teaches me, by these sensations of pain, hunger, thirst and so on, that I am not merely present in my body as a sailor is present in a ship, but that I am very closely joined and, as it were, intermingled with it, so that I and the body form a unit. If this were not so, I, who am nothing but a thinking thing, would not feel pain

ill-equipped to provide the grounds of such explanation. Because, on an occasional-causal account, mind and body do not form a unity in virtue of any intrinsic features of body or mind, Descartes must give up the notion that the unity of mind and body through their interaction is anything more than that of the "sailor piloting the ship." If it is conceded that they are not united in a third substance,[125] nor by a straddling mode, nor even by transeunt causation, the impression must be false that the thinking substance is in any kind of intrinsic union with the body. God (at least to the Descartes of the Sixth Meditation) would therefore be a deceiver. In an occasional-causal account of the interaction between mind and body, then, though we may concede that Descartes is able to assert a relatively weak form of interaction, it would not assuage Elisabeth's worries nor vindicate concerns with God's truthfulness raised by failure of efficient-causal interaction.

To step away somewhat from the stance of internal critique, the invocation of the "deus ex machina"[126] of occasional causation underpinned by divine decree is a confession that given Descartes' sharp division between the mental and the intellectual, there is no immanent principle 'within' finite nature which coordinates the material occasional cause and the mental effect (and vice versa). Descartes may have been comfortable with this state of affairs, seeing

---

when the body was hurt, but would perceive the damage purely by the intellect, just as a sailor perceives by sight if anything in his ship is broken. Similarly, when the body needed food or drink, I should have an explicit understanding of the fact, instead of having confused sensations of hunger and thirst. For these sensations of hunger, thirst, pain and so on are nothing but confused modes of thinking which arise from the union and, as it were, intermingling of the mind with the body. (CSM II 56)

[125] See, e.g., Paul Hoffman, 'the Unity of Descartes' Man' (1986), *The Philosophical Review* 95(3), pp. 339-370, who argues that Descartes does see human beings as a substantial unity, and argues, contrary to our analysis of the impossibility of this on Descartes' theory of principal attributes and modes, that mind and body are united as a single substance with actual, separable sub-components.

[126] Nadler 2010, p.44.

it as a kind of genuine (if "remote") 'causal' relationship,[127] but the account still constitutes a definite explanatory extravagance relative to the alternative (that mind and body are somehow deeply and naturally unified as seems evident in common experience), since it posits (on an ad-hoc basis) an extra coordinating factor of indeterminate ontological status, the divine decree, which is independent of the active divine will (to distinguish it from occasionalism) of indeterminate ontological status and external to body and mind. What Descartes has put asunder, it seems only God could hold together.

The failure of union of mind and body as to form a nexus of causal interaction which eminently or formally contains its effects is thus not merely 'difficult to understand.' It is, in most respects, ruled out by precisely the ontological moves which Descartes introduces to support his dualism in the first place, particularly the substance-mode doctrine as it interacts with the one-principal-attribute rule, Descartes' epistemology and Descartes' own causal doctrines. Substance-mode ontology helps Descartes establish the fundamental kinds of relations between attributes and modes: modes depend upon, and are indeed intrinsically particular specifications of, more-general attributes, requiring most-general 'principal attributes' which terminate the regress of dependence and are not themselves specifications of any other. From this it follows that any given substance could only have one 'principal attribute' which is exclusive of the other, since the only thing which could ground the unity of a substance is unity in its principal attribute. Granting Descartes' epistemology, which posits that the implications of an idea may be thoroughly studied by an inspection of a 'clear and distinct idea,' we achieve near-certainty that thought and extension truly are distinct principal attributes. Thus isolated, Descartes achieves a dualism not only of properties, but of substances. However, this isolation of principal attributes from each other, intrinsic in the very notion of a principal attribute, prevents the production of anything in one substance constituted by a particular principal attribute by

[127] CSM I 304.

another with a different principal attribute. Interaction between mind and body is only somewhat viable on the limited terms of occasional causation, which 'papers over' the causal gap with divine fiat but is not sufficient to establish the kind of real substantial unity required by experience. Insofar as Descartes argues that the union of mind and body is suggested by the direct and indubitable experience of certain modes of thought, he only makes the unsolvable problem more acute.

2.5. The 'Metaphysical Scaffold' and the Significance of Descartes' Interaction Problem

As shown above, the most immediate costs to the Cartesian project lie in the explanation of those parts of mental experience which Descartes acknowledges cannot exist but for the unity of mind and body. Descartes' philosophy predicts a phenomenology that requires genuine union between mind and body: if the mind could not really be united to the body, then the mind would be united to body only as the pilot of a ship, and its apprehension of its unity to the body would be via purely intellectual acts of understanding, rather than being bound up with imagination and sensation such that it identifies with the movements, passions and sensations of the body. The union of body and mind enters necessarily into the explanation of how thought may be confused and distanced from pure understanding. Though we may note with Rozemond that sensations are still modes of thought (and therefore mind) rather than some mind-body composite substance,[128] it remains true that the notion of union, if it be not merely an obfuscatory gesture, requires some real sense in which mind and body are one, which Descartes' substance-attribute-mode ontology, as argued above in section 2.4, does not allow.

A largely intractable interaction problem, then, rooted in the impossibility of the production of modes in one substance by another with a different principal attribute, contradicts what is,

[128] Marleen Rozemond, *Descartes' Dualism*, Harvard University Press, 1998, p.212.

even for Descartes, a foundational datum of human experience: the appearance of sensation and imagination. Even granting the slight relief provided by an occasional-causal model of mind-body interaction, the interaction problem for Descartes represents a fundamental failure of mind and body to form a unified reality, giving weight to the sceptic's doubts that Descartes' analysis of extension and thought is correct. The only remaining unifying principle is God, who must be invoked, on something like occasionalist or parallelist terms, in order to salvage anything like interaction at all.[129] Given the manifest appearance of interaction in the quality of sensory and imaginative experience, and phenomena (like the gradual improvement of mental faculties in conjunction with bodily development, as noted by Arnauld) which suggest some real and not merely apparent basis of unity between mind and body, the resort in an ad-hoc fashion to the mysterious intentions and decrees of God to explain the interaction of mind and body at least puts Cartesian dualism on an explanatory par with monisms which posit some mysterious ground of unity between the apparently diverse attributes of thought and extension. Indeed, insofar as the latter contemplate an immanent as opposed to wholly transcendent ground of the harmony of thought and extension, the latter has the theoretical advantage of in-principle tractability.

The interaction problem manifests a deeper issue than peculiar infelicities in Descartes' ontology, epistemology and metaphysics of causation. It is important to recall at this point that Descartes' ontology is motivated in the first place by the desire to re-found metaphysics on the (to Descartes) intellectually transparent and seemingly all-encompassing notions of thought

[129] Indeed, it is doubtful that divine intervention would avoid further problems— as raised already in connection with occasional causality, if God creates the impressions of the unity of mind and body despite this being substantially impossible, is might be necessary to conclude that God is indeed a deceiver, which would be catastrophic for Cartesian epistemology.

and extension, and thereby to displace the broadly hylomorphic scholastic synthesis. Between his doctrines of thought and extension, it is the attribute of extension and its role in physical reality which is the greater departure from hylomorphism and which needed greater shoring-up with Descartes' new substance-attribute-mode ontology.[130] The great advantage that the substance-attribute-mode ontology conferred for the Cartesian project over the hylomorphic framework is that it allowed extension, hitherto an 'accident' which was ontologically secondary next to a primary 'substance' constituted chiefly by a substantial form, to serve instead as the 'principal attribute' (i.e., the 'whole nature and essence') of extended, material things, and therefore supply real explanatory insights into physics. This deliberate metaphysical choice came at the price of the notion of 'substantial forms' and the neutral 'metaphysical scaffold' they provided in which qualitative and non-qualitative properties could play complementary roles. In Descartes' ontology, the primacy of extension to physical reality is so comprehensive and absolute that there is no way for non-extensive properties to exist except as an entirely different kind of substance altogether. Even if Descartes can attempt to motivate his dualism by means of his a priori arguments such as he attempts in the *Meditations,* it remains the case that Descartes' conception of the physical itself exerts a conceptual pressure in favour of dualism. If one initially considered only Cartesian extended substance, the occurrence of mental substance is maximally surprising and utterly incommensurable, most demanding of an explanation (in virtue of its unusualness) and yet least accommodating of explanation. This 'surprisingness' of the mental in a physical world is a major theoretical disadvantage to any theory of mind and body, inviting either the elimination of the 'surprising'

[130] The doctrine of the immaterial soul, by contrast, which Descartes referred to as the 'substantial form of the body' and the characteristic operations of which greatly resembled the intellectual soul of the scholastics which had an 'immaterial,' purely formal, potentially-subsistent existence, had 'pure thinking', which greatly resembled abstraction in its freedom from particular mental imagery, as its characteristic operation.

phenomenon (which, in the case of mental properties, carries its own worries) or the explanation of that phenomenon in terms of some common principle with the physical. It is this sense of 'surprise' which is the fundamental problem of dualism.

To the degree that one relaxes one's commitment to a notion of physical substance which consists entirely of extension, the mental properties excluded by extension as the essence of the physical become more cohesive with the physical. For example, property dualism seems to have a great advantage over substance dualism in that, as merely different properties of a single substance, one avoids the necessity of explaining whence an utterly unforeseen second order of basic substance originates. And yet, as shall be seen in Part 3, property dualism and the various forms of 'neutral monist' models it helps motivate are not without their own analogues of the interaction problem. As long as there remains a sense in which extension or something like it is our ontological 'default' for characterising the physical, then, something like a problem of dualism will recur.

Of course, given how far empirical science and understanding have come without the use of such a vantage point, and how relatively inconsequential to the vast majority of such projects, even in cognitive science, the recovery of such a vantage point is, the resolution of the problem of 'surprise' cannot simply be a matter of returning to a pre-Cartesian hylomorphism. Part 3 of this thesis will examine the degree to which something like Descartes' notion of the material continues to inform our conception of the physical in the form of the structure-and-dynamics conception of the physical, whether it contributes to modern problems of dualism, and what reasonable steps may be taken toward a solution.

Part 3: Modern Dualism and the Ghost of Descartes

Part 3 explores the parallels between the conceptual pressures leading Descartes to his dualism and its associated problems on the one hand, and the continuing conceptual pressure in favour of dualism in the modern discourse on the other. Part 3 articulates the sense in which certain notions of the physical which are attractive even today in turn make dualism attractive. In particular, it focuses on the current discourse regarding 'structural-dynamic' argument against materialism, which most closely parallels the pressure toward dualism faced by the Cartesian project of mechanisation and physico-mathematisation. While this Part is a broad survey of a complex field and will actively defend a particular slant on the structure-and-dynamics argument and its significance, it is not the purpose of this Part to provide exhaustive rejoinders to all rival positions or establish unassailable conclusions. Rather, the aim of this Part is primarily to illustrate how the conceptual pressure toward dualism endures in the present and fruitfully interacts with contemporary approaches to situating mind in nature, and this is best accomplished by engaging these approaches in argument.

Section 3.1. surveys some of the differences between Descartes' dialectical context and the modern one, and note the resilience of dualism in the face of the physicalistic challenge. **Section 3.2.** explains the 'structure-and-dynamics' argument against materialism. **Section 3.3.** explains the significance of the structure-and-dynamics argument, in that it both justifies and explains the resilience of dualism, presents a key source of the 'hardness' of the 'hard problem' of consciousness, and closely parallels the anti-materialistic pressure generated by Descartes' physico-mathematical vision. **Section 3.4.** evaluates the objections to the structure-and-dynamics argument, and argues that the anti-materialist pressure which the argument exerts survives most obvious challenges very well, with minimal modification. **Section 3.5.** evaluates several ways of responding to the force of the structure-and-dynamics argument, including

physicalistic, dualistic and neutral-monistic strategies. **Section 3.6.** suggests that addressing the problems raised by the structure-and-dynamics argument requires a deep re-examination of our foundational metaphysical choices in the characterisation of nature. By reference to what was lost in the transition from scholasticism to Cartesian mechanism, the costs of which endure in a new form even today, some suggestions for relieving the dualistic pressure will be considered.

I conclude this thesis by reflecting on the degree to which the discourse on dualism has, for all its increased sophistication, remained subject in remarkable ways to conceptual pressures which have endured since the days of Descartes.

3.1. The Continuing Challenge of Dualism

It is a great understatement to say that there has been considerable change in our understanding of the physical (i.e., non-mental) world since Descartes, and we do well to briefly note the general direction of these changes. There are, for instance, differences in epistemological ambition with respect to the physical world. Whereas one of Descartes' chief aims in accepting only the principles of geometry and pure mathematics in physics was to generate 'certain demonstrations explaining physical phenomena,'[131] modern science has more moderate aspirations to empirical and explanatory adequacy—our understanding of the structure and laws of the physical world is commonly taken to be mediated by fallible models and theories subject to future revision.[132] There are also differences in the metaphysical commitments of the

[131] CSM I 147.

[132] For an anti realist like Van Fraasssen (1980), '*Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate*,' and realists also tend to be fallibilist, as noted by Chakravartty (2017), *'Realists are generally fallibilists, holding that realism is appropriate in connection with our best theories even though they likely cannot be proven with absolute certainty.'*

modern naturalist. Where a Cartesian mechanistic view of physical nature allowed only causal interaction between physical particles and extended 'substances,' modelling their interaction on the way that the parts of a machine determine the motions of the whole,[133] the modern model of physical reality now allows the influence of things like forces and fields as physically interactive causes. Unlike Descartes, modern naturalism is committed to the causal closure of the physical based on conservation laws and further developments in the kinds of entity which can have a causal effect.[134] While Descartes took his substance-attribute-mode ontology to exhaust the underlying metaphysical structure in nature, modern naturalism tends to prescind from 'deep' metaphysical commitments based on an a priori analysis of fundamental ontological categories which go beyond the kinds of categories involved in the course of empirical scientific work. Without the kind of substance-attribute-mode ontology Descartes invoked both to support his mechanisation of physics and the reification of mind, there is considerably less pressure in modern naturalism to adopt Descartes' brand of substance dualism.

Instead of the 'extended substance' which is suggested to Descartes by a 'clear and distinct idea' of an exhaustive 'principal attribute,' which for Descartes provides the common root of the basic ideas in physics and mathematics, a characteristically modern naturalistic approach to metaphysics posits only entities of the kind required by our best empirical science, and among the sciences, physics, while prescinding from a priori metaphysical commitments concerning the intrinsic natures of substances and ideas. While there has been some dispute as

[133] In Part Four of his *Principles* Descartes writes, '[of the visible universe] Up till now I have described this earth and indeed the whole visible universe as if it were a machine: I have considered only the various shapes and movements of its parts,' CSM I 279.

[134] Justin Tiehen, 'Explaining Causal Closure' (2015), *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 172(9), 2405.

to whether ontological naturalism (the view that we ought posit only the kinds of entities posited by our best science) entails physicalism (the view that only the kinds of entities posited by our best *physics* are metaphysically respectable), and whether physicalism entails materialism (which might be distinguished from physicalism as the more specific view that the only physical substances are something like inanimate, senseless subjects of extension, figure and motion), it will be assumed in this part that naturalism entails physicalism (since it is physical theses like the causal closure of the physical which make ontological naturalism plausible), and that the more restrictive forms of materialism are alternative words for 'physicalism' so-conceived (the more restrictive forms of materialism, in other words, will not be taken to exhaust the meaning of 'materialism,' in light of the use of materialism as an equivalent to physicalism in the literature).[135]

Despite all the differences, however, there is still much in contemporary physicalism which vindicates Descartes' broadly 'physico-mathematical'[136] approach to physics. Even if the image of a machine composed entirely of extended, divisible substances in motion interacting and exerting forces upon each other by direct contact has faded in favour of a richer universe of fields and forces in addition to microphysical particles, the modern practice of modelling nature by means of identifying its fundamental quantities and expressing the relations between them in the language of mathematics seems to be more of a development from, than a fundamental rejection of, Descartes' basic approach to physics. The precision and predictability of mathematics and geometry, which impressed Descartes so much and which, when reified

[135] Daniel Stoljar, 'Physicalism', *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/physicalism/>.

[136] It bears noting that while this was not, in Descartes' case, an actor's category, it still connotes the distinctiveness of Descartes' emphasis on mathematical categories and relations in the explanation and demonstration of things concerning physics.

via Descartes' metaphysics, promised direct knowledge of the workings of nature, is little less impressive in its precision, clarity, predictive power and technological applicability today. Even without Descartes' precise ontological underpinnings, there remains a strong theoretical impetus to encompass all of nature within a science which employs such impressive foundations.

The remarkable fact which attracts our attention despite these developments is that mental phenomena have proven a site of peculiar resistance to assimilation to a purely physicalist ontology. There have been modern and sophisticated recurrences of dualism, beginning with David Chalmers' articulation of the 'hard problem of consciousness.'[137] The 'hard problem of consciousness' refers to the problem raised by certain properties of conscious experience, particularly 'qualia,' the 'what-it-is-like'-ness of subjective experience, which are thought to be in-principle difficult to metaphysically assimilate to a physicalist view of the world.[138] While there are contemporary defenders of a more or less traditional Cartesian substance dualism,[139] Chalmers has favoured a 'property dualism,' the idea that there are two objectively different kinds of properties in the world—mental and physical properties. It is a main contention of this part that this recurrence of dualism is no accident. While the pressures toward dualism are not quite the same as in Descartes' day, e.g., there is less religious pressure to secure a demonstration of the immortality of the soul, and the paradigm cases of immaterial

[137] David Chalmers, 'Facing Up to the Problem of Consciousness' (1995), *Journal of Consciousness Studies*, 2, pp.200–19.

[138] David Chalmers, 'Facing Up to the Problem of Consciousness.' In *The Character of Consciousness*, Oxford University Press, 2010, p.5.

[139] Richard Swinburne, 'What makes me, me?' in Andrea Lavazza and Howard Robinson (eds), *Contemporary Dualism: A Defense* (1st ed.). Routledge 2013.

properties on modern arguments for dualism focus on phenomenal qualities as opposed to abstract intellection, there remains even for an atheist like Chalmers something uniquely 'queer' about at least some of the properties of consciousness. That is, there is, intuitively, some sort of 'remainder' in consciousness which goes uncaptured by even the most complete physical description of the world, which are remnants of that which is excluded by a crypto-Cartesian notion of the physical. Famous thought experiments like Jackson's 'What Mary didn't know'[140] and Chalmers' own 'zombie' thought experiments,[141] which draw upon our intuitions about what would be entailed by a complete physical description of the world, purport to imply that facts about the intrinsic nature of consciousness—the 'what-it-is-to-be-like-ness' or phenomenal qualities of conscious properties, could not be derived from such complete physical descriptions of the world.[142] Such arguments raise a 'hard' problem which, so dualists assert, cannot be resolved merely by continuation of the physicalist programme as currently understood.[143]

[140] Frank Jackson, 'Epiphenomenal Qualia' (1982), *The Philosophical Quarterly*, 32(127), pp. 127–136.

[141] David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, 1996, p.94.

[142] It bears noting that the paradigmatic instance of Descartes' principal attribute of 'thought' is abstract understanding, e.g., the concept of the chiliagon apart from its instances, as opposed to the phenomenal qualities of particular instances which impress modern dualists such as Chalmers. Descartes would have, for his own reasons, regarded the phenomenal qualities which impress Chalmers as the product of the interaction of thought with the physical. Even if Chalmers and Descartes are impressed by different things as the 'remainder' which cannot be assimilated to the physical, the commonality between them is that the idea of the physical, as far as we can grasp it, implies a kind of ontological 'remainder' which cannot be assimilated.

[143] That dualism remains a formidable challenge is reflected in even physicalist commentators such as Lycan, in 'Giving Dualism its Due' (2009), *Australasian Journal of Philosophy*, 87(4), pp.551–561.

Certainly, such intuitive arguments for dualism face formidable obstacles: Our physical understanding of neurological systems is still in its infancy, and it might therefore be premature to say what a completed physicalist programme would and would not be able to explain (i.e., there is the hope that a future completed physics would bridge the apparent nomological and ontological gaps between a physical description of the world and the phenomenal properties which occur to us in consciousness). There are also extremely undesirable theoretical consequences of dualism: modern property dualism may not give rise to the same kind of interaction problem as substance dualism did for Descartes, but the problem of the relationship of such mental properties to the physical is deeply analogous, giving rise specifically to problems of epiphenomenalism and associated epistemological issues which will be treated more fully below. In general, any form of dualistic property or principle of consciousness generates an issue in how its interaction or co-occurrence with the more familiar physical properties or substances are to be articulated, predicted, understood and explained. Wherever the physical and the non-physical seem to lack any sense of ontological unity, their co-occurrence seems inexplicable, and this is a deep source of theoretical inelegance—this division in nature between the physical and the non-physical would seem to be, as we noted in the conclusion of Part 2, maximally surprising from the point of view of the study of physical properties, maximally demanding of explanation (because of its unexpected nature), and minimally accommodating of explanation.

If it can be shown that there is a plausible root in the modern conception of the physical which generates such intuitions, even if dualism is not on the whole to be preferred to physicalistic monism, that will suffice to show that there is a deep theoretical difficulty in the modern concept of the physical which bears solving. If in turn that problem should bear similarities to Descartes' conceptual pressure toward dualism, which owes its existence in significant part to an overly-reductive concept of the physical, that might in turn suggest what would need to be

done to recover a theoretically satisfying metaphysical synthesis. It is my overall objective in this part to explore the parallels between the Cartesian and modern conceptual pressures in favour of dualism, analyse some of the options for dealing with this pressure, and suggest some promising aspects of any resolution to this problem.

3.2. The Structure-and-Dynamics Argument against Materialism

The conceptual pressure in favour of dualism from Descartes stems from his intuition that the 'principal attributes' of thought and extension are independent and mutually exclusive and this exclusivity stems from his rethinking of the idea of the physical in terms of pure extension. A promising parallel to this conceptual pressure in the modern discourse lies with what has been called the 'structure- and-dynamics' argument against materialism, as advanced by David Chalmers.[144] This sub-section will outline the structure-and-dynamics argument against materialism and its rejoinders, and assess its implications for the conceptual pressure toward dualism.

The structure-and-dynamics argument proceeds from the notion that a completely physical description of the world describes the world solely in terms of 'structure and dynamics.' Such an account maintains that the complete physical truth about the world consists in truths about its microphysical structure and the truths about the dynamic evolution of that structure over time, and all the further truths which can be deduced from these fundamental structural-dynamic truths. As Chalmers explains,

> *A microphysical description of the world specifies a distribution of particles, fields, and waves in space and time. These basic systems are characterized by their spatiotemporal properties, and properties such as mass, charge, and quantum wavefunction state. These*

[144] Chalmers 2010, pp.120–123.

latter properties are ultimately defined in terms of spaces of states that have a certain abstract structure (e.g., the space of continuously varying real quantities, or of Hilbert space states), such that the states play a certain causal role with respect to other states. We can subsume spatiotemporal descriptions and descriptions in terms of properties in these formal spaces under the rubric of structural descriptions. The state of these systems can change over time in accord with dynamic principles defined over the relevant properties. The result is a description of the world in terms of its underlying spatiotemporal and formal
1780 *structure, and dynamic evolution over this structure.*[145]

More precisely, structural dynamic truths (henceforth, 'structural truths') are truths about the world which can be captured in formal (i.e., logical and mathematical), spatiotemporal, and nomic (lawlike descriptions of its evolution over time) terms.[146] [147] With this notion in hand, the argument can be stated in brief as follows:

1. All physical truths are purely structural truths.

[145] Chalmers 2010, p.120.

[146] Chalmers 2010, p.140, 'In formal terms, a structural-dynamic description is one that is equivalent to a Ramsey sentence whose O-terms include at most spatiotemporal expressions, nomic expressions, and logical and mathematical expressions'.

[147] There is certainly an issue of distinguishing between the notion of a physical truth as referring to a particular kind of subject matter, or employing certain kinds of terms (causal, nomic, spatiotemporal, etc) to describe its referent. For our purposes of showing that there is a 'remainder' after the complete physical truth is accounted for, it does not seem to matter—if a description employing a restricted range of terms is an incomplete description of its referent, it is reasonable to infer that, while it successfully refers to some aspects of its referent, there is some other real aspect of its referent which is not captured in such a description. Stoljar notes this issue in Daniel Stoljar, 'Russellian monism or Nagelian Monism?' in *Consciousness in the Physical World: Perspectives on Russellian Monism*, edited by Torin Alter, and Yujin Nagasawa, Oxford University Press, 2015., p.338.

2. Only further structural truths can be deduced a priori from a body of purely structural truths.
3. At least some truths about consciousness are not truths about structure and dynamics.
4. Therefore, there are some truths about the world which cannot be deduced from a complete physical description of the world.[148]

The argument's first and second premises attempt to establish a conception of the physical under which there is a pressure toward dualism—if physical truths are, insofar as physical truths, purely structural truths or deducible from purely structural truths, then any complete physical description of the world can consist only in purely structural truths. If this is so, the only way to have a complete understanding of mind be part of a complete physical understanding of the world would be for truths about consciousness to themselves be purely structural truths. However, by the third premise, it seems that on introspection at least some truths about consciousness, in particular, truths about the phenomenal qualities of consciousness, or 'phenomenal truths,' (such as what Mary learns when she leaves the colourless room) are not purely structural truths, and *therefore*, by the second premise, not a priori deducible from any structural truths. The conclusion of the argument is that there is an in-principle *epistemic* gap between the complete physical truth about the world and at least some truths about consciousness.

The first premise is motivated by the nature of scientific explanation. According to Chalmers,[149] scientific progress is made precisely insofar as apparently puzzling behaviours or structures (Chalmers gives the example of the *elan vital* of the vitalists) can be fully accounted

[148] Torin Alter, 'The Structure and Dynamics Argument against Dualism' (2016), *NOUS*, 50(4), pp.794–815, p.795.

[149] Chalmers 2010, p.16.

for in terms of the behaviour and structure of their physical components, and ultimately, their microphysical components as defined in terms of structure and dynamic terms. The core output of the scientific endeavour, therefore, is the exposure of the structural underpinnings of phenomena.

The second premise shares similar motivations to the first: the physical truths successfully explained by derivation from more fundamental structural truths are themselves rendered susceptible of explanation in this way by analysis in terms of structure and dynamics, which are subsequently shown to be deducible from the behaviour of the more fundamental structural truths.[150]

The third premise is motivated by the introspective character of conscious experience: that something appears to be left over even when one offers an account of the functional role that consciousness plays in the physical system in which it is embedded. There are certainly issues with the introspective basis of the third premise, but more shall be said in **section 3.4.**

The result of the structure-and-dynamics argument, that there is an in-principle epistemic gap between the complete physical truth about the world and at least some truths about phenomenal consciousness, is a key component of the case for any sort of dualism: it sets the epistemic gap between the physical and phenomenal truths on a firm explanatory foundation, pointing out a plausible fundamental feature of the physical truth which leads to the explanatory gap, which in turn serves as the foundation for the further inference that there is a *modal* gap between the complete physical truth about the physical world and certain phenomenal truths (i.e., it is possible that the complete physical truth is instantiated without the associated phenomenal

[150] Chalmers 2010, p.16.

truths), and subsequently, that there is a genuine ontological gap between the complete physical truth and certain phenomenal truths.

3.3. The Significance of the Structure-and-Dynamics Argument against Materialism

In order to help us further understand the significance of the structure-and-dynamic argument, we might summarise the various positions on the assimilability of phenomenal truths to the physical truth as follows, according to their position with regard to the epistemic, modal and ontological gaps of the anti-materialist and how to solve them:

Type A materialists: there is no epistemic gap between physical and mental properties. (i.e., reductionism or eliminativism with respect to qualitative, phenomenal truths).

Type B materialists: there is an epistemic gap, but no modal gap. Ignorance undermines the inference from the epistemic to the modal gap. We might bridge this gap with a posteriori identity relations (even on an ideal physics, the gap would be bridged by empirically discovered a posteriori relations rather than a priori deductions).

Type C materialists: There is an apparent epistemic gap for now, but on a completed physics, there would be no epistemic gap (hence, on type C materialism, unlike type B, at least in a completed physics one would be able to a priori deduce all physical truths from the fundamental physical truth, whatever that might turn out to be).

Type D dualists: there is an ontological gap and mental and physical properties are both fundamental, and there is interaction (interactionism)

Type E dualists: there is an ontological gap, and there is no interaction. (epiphenomenalism)

Type F monists: 'Russellian monism,'[151] the position that the physical and the mental both capture aspects of an underlying unity which is neither wholly physical nor mental in nature. In effect, this denies that both the physical and the mental are fundamental properties in favour of a more-fundamental 'neutral' ontological status which ontologically unifies the mental and the physical.[152]

The primary result of the structure-and-dynamics argument against materialism is not, in itself, an argument for dualism. By providing a principled argument for the epistemic gap from the fundamental nature of physical truths, the structure-and-dynamics argument could, by motivating a further modal and ontological gaps, motivate a dualism about properties. Given the theoretical difficulties dualism presents, however, a theoretical context where dualism becomes attractive is also one where a kind of neutral monism which explains and unifies the two different kinds of attributes and promises to ameliorate some of the theoretical disadvantages of dualism becomes attractive.[153]

In terms of the range of positions summarised above, the structure-and-dynamics argument, if successful, provides a direct argument against type A materialism (in positing an in-principle epistemic gap at all) and type C materialism (since the epistemic gap is one which applies in-principle if the physical qua physical deals purely in structural-dynamic truths, even an ideal and completed physics would not avoid the epistemic gap). The argument also provides the

[151] Bertrand Russell, *The Analysis of Matter*, Routledge, 1927, p.382.

[152] This list is taken from Alter 2016, p.796.

[153] Daniel Stoljar (2015) differentiates between a 'Russellian' monism in which the distinction between physical and non-physical, even if explained by the prior neutral substrate, remains a major distinction, from a 'Nagelian' monism which, while critical of current physics, does not think that the distinction will continue to be a major one in the final theoretical context. Stoljar's Nagelian monism, according to Stoljar, does not differ too drastically from type B materialism as articulated by Chalmers.

first step of an argument against type B materialism, according to which the identity relations between mental and physical properties are discovered a posteriori, despite an a priori epistemic gap. Chalmers can then move from that in-principle epistemic gap, via arguments like his 2-dimensional semantics argument against materialism, to a modal gap.[154] [155]

1870 According to Alter, a major significance of the structure-and-dynamics argument against materialism is that it pinpoints the sense of the physical which, in conjunction with certain aspects of experience, exerts the strongest pressure toward dualism.[156] That is, even if the argument is not ultimately successful as an argument against materialism, it explains why the intuition that physicalism 'leaves something out' has the force that it does. Moreover, if successful the argument explains why that the gap is not just a feature of our temporary ignorance: it is a feature of the physicalist project *as such* that it cannot account for phenomenal properties, *because* the physicalist project is per se the project of identifying all truths about the physical world with structural truths, but no amount of structural truth could bridge the epistemic gap between physical and phenomenal qualities.

1880 There is a more than superficial resemblance here between the trajectory of Descartes' pressure toward dualism and that pointed out by the structure-and-dynamics argument. Chalmers'

[154] In terms of Chalmers' 2D argument against materialism, the structure-and-dynamics argument provides a reason to think that the intensionality of the physical truth doesn't vary with whatever world is actual, and if phenomenal truths don't vary either, any identity between the phenomenal truth and the physical truth cannot be necessary a posteriori. If, as on the structure and dynamics argument, phenomenal and physical truth cannot be a priori identified, that leaves only a contingent relation between the phenomenal and the physical, confirming dualism.

[155] David J. Chalmers., 'The Two-Dimensional Argument Against Materialism' in *The Oxford Handbook of Philosophy of Mind,* edited by Ansgar Beckermann and Brian P. McLaughlin, Clarendon Press, 2009.

[156] Alter 2016, pp.799–801.

notion of the identification of physical truth with 'structural truth' is clearly analogous to the Cartesian notion of extension as a principal attribute: both extension and structural truths offer powerfully attractive candidates for the reduction of physical objects of common experience, precisely in virtue of their mathematisability and predictive power, and yet seem to conflict with our intuitions about the full range of the kinds of truths there are. Indeed, Cartesian extension seems to be precisely a kind of structural truth reified as a thing-in-itself and overlaid upon the barest non-structural ontological skeleton (i.e., Descartes' substance-attribute-mode metaphysics). Phenomenal experience seems to confront us with an apparently primitive richness of non-structural, non-extensive reality which seemingly cannot be reduced to merely geometric-cum-dynamic (Descartes) or structural abstractions. This 'un-assimilable' remnant, is then left out of our picture of reality as an uncomfortably fundamental and surprising (from the point of view of our physical science and its implicit metaphysics) ontological anomaly.

The significance of the structure-and-dynamics argument can endure even if its identified 'essence' of physicalism-the identification of physical truth with structural truths- is either not decisive (because, perhaps, there are rival views which turn out to be more attractive) or untenable only for indirect reasons (i.e., if it leads to dualism, and dualism has its own peculiar troubles). As long as structural truth remains an attractive *option* for what it is to be a physical truth, there will be a problem of situating structural truths without generating a problem of dualism. The issues with accommodating the practical emphasis on discovering structural truths about the physical world without creating a pressure toward dualism will be considered more detail in sub-**section 3.6**., after direct challenges to the structure-and-dynamics argument against materialism are considered.

3.4. Evaluating Objections to the Structure-and-Dynamics Argument Against Materialism

The physicalist who objects to the structure-and-dynamics argument against materialism has many options for resisting the force of the argument. This section aims to give a broad survey of the kinds of issues the structure-and-dynamics argument raises and in turn give a sense of how the argument captures an enduring issue with materialism.

3.4.1. Structural Truths Do Not Exhaust Physical Truths

First, it might be objected that the central notion of what a 'physical truth' is—that it is equivalent to structural truths—does not plausibly exhaust the content of an ideal physics. There are many approaches to this general point, and this sub-section will survey the most salient ones. Put in the most general terms, given the kinds of paradigm shifts that physics undergoes, it seems probable that we know too little about the final content of physicalism to make a judgement that structural truths exhaust the physical truth. The immediate difficulty with this objection is that it threatens to vacate physicalism of any content—if we know so little about the future content of an ideal physics that just any metaphysical arrangement might (for all we know) count as a 'physicalism,' it seems that physicalism ceases to be a metaphysical thesis with any particular content at all. This is one branch of Hempel's dilemma as applied to the philosophy of mind:[157] either we must restrict the content of 'physics' to what we presently know, in which case an epistemic gap is quite likely, or we must leave the future content so open that physicalism might, for all we know, cease to be even a monistic thesis, and hence difficult to hold as against dualism.

[157] Carl G. Hempel, 'Reduction: Ontological and Linguistic Facets,' in Sidney Morgenbesser, Morton White, Patrick Suppes and Ernest Nagel (eds), *Philosophy, Science and Method: Essays in Honor of Ernest Nagel,* St Martin's Press, 1969, pp.186–187.

A way to strengthen this objection is suggested by Stoljar where he points out examples where the structural truth doesn't seem to exhaust even the currently known physical truth.[158] Stoljar invites us to consider physical truths such as a 'volcano is erupting,' under conditions of complete structural omniscience—i.e., knowledge of the complete structural truth of fundamental physical quantities, the structural, nomic and dynamic relations between them. Stoljar contends that knowing the structural truth would still leave us unable to identify any particular set of such structural truths as constituting the truth that 'this volcano is erupting.' Therefore, it is not plausible that, even in the case of such relatively well-understood phenomena as volcanoes, where nobody asserts a problem of dualism, the structural truth does not capture everything there is to be known physically about the volcano as it is in itself.[159] This serves as a refutation to the first premise of the argument: there are perfectly ordinary physical truths which are not captured by structural truths.

I do not think that such a strategy is successful. Even if we could not tell in what sense the truths encompassed by the complete structural truth (i.e., the complete truth about the structural properties of the world) map on to our present grasp of the truth that 'the volcano is erupting,' it might still be the case that the truth that the volcano is erupting fully reduces to some set of truths within the complete structural truth—i.e., the truth that a volcano is erupting would be

[158] Daniel Stoljar, 'Russellian Monism or Nagellian Monism?' in Yujin Nagasawa and Torin Alter (eds), *Consciousness and the Physical World*, Oxford University Press, 2015, p.341.

[159] Stoljar (2015) raises this objection in the context of forcing a tension between Chalmers' phenomenal functionalism about spatiotemporal properties, the doctrine that spatial properties are picked out by their role in generating a certain manifold of phenomenal experience. Since I am not committed to Chalmers' phenomenal functionalism, this context is not relevant for the argument at hand, but I think the objection, though not as strong as when Stoljar uses it to force a tension with Chalmers' view, can be adapted to provide a general objection to the structure-and-dynamics argument which better serves our purposes.

'nothing over and above' some set of structural truths about the underlying microphysical properties and the causal, nomic and spatiotemporal relations which give rise, ultimately, to the structure and dynamics of the erupting volcano. If this is so, then it turns out that the truth that the volcano is erupting does fully reduce to structural truth, and hence, even granting that 'the volcano is erupting' is a physical truth which we would have trouble connecting to the complete structural truth in our understanding, it would not entail that, in the truth that 'a volcano is erupting,' there is any physical truth which is not a structural truth or (under the aspect of its corresponding truths within the complete structural truth) deducible from a structural truth.

Another strategy for denying that physical truths are equivalent to the structural truth might be to argue that structural truths by themselves imply further truths which are not structural, which the physicalist qua physicalist ought not have trouble accepting. For example, structural truths, if we conceive of them as fundamentally 'extrinsic' relational truths about physical entities, might imply further truths about 'intrinsic,' non-relational properties.[160] Relations between the points on an object's surface imply certain intrinsic properties of shape, which depending on our understanding of structural truth, might not be 'structural.'[161] Relatedly, it has been argued in the emerging integrated information theory (IIT) space, that certain kinds of at least mind-adjacent 'intrinsic' properties such as a subjective 'point of view' may be objectively derived from structural truths about the biological system considered as 'for itself'.[162] Without going

[160] Derk Pereboom draws the relevant distinctions between extrinsic, 'comparatively intrinsic' and absolutely intrinsic properties in relation to a structural argument against materialism in Derk Pereboom, 'Russellian Monism and Absolutely Intrinsic properties,' in Derk Pereboom, 'Russellian Monism and Absolutely Intrinsic Properties,' *Current Controversies in Philosophy of Mind*. Routledge 2014, pp.40–69.

[161] Stoljar 2015, p.803.

[162] Garrett Mindt, 'Not All Structure and Dynamics Are Equal' (2021), *Entropy*, 23, p.1226.

too deeply into the details of such a view, the basic idea is that there is an objectively quantifiable sense in which certain kinds of self-sustaining systems consisting in interdependent subsystems generate an implicit 'internal point of view' by the fact of their self-sustaining activity and the associated functional hierarchy among subordinate mechanisms. Such functions can be captured in recognisably structural-dynamic terms and may be said to imply certain facts about an intrinsic 'point of view.' In both these cases, the countervailing force of the objection derives from the fact that at least the existence of non-structural phenomenal truths about an 'intrinsic point of view' can be a priori derived from certain structural facts about the system, undermining the notion that some truths about consciousness, in virtue of those truths not being structural, could not be derived from the complete physical truth considered in structural-dynamic terms.

I agree with Alter (2016) that this sort of counterargument has some force, and requires a certain modification of the structure-and-dynamics argument.[163] Alter contemplates the following modification to the Structure-and-dynamics argument. Rather than physical truths being identical only to structural truths, they might be identical only to structural truths or what Alter calls 'weaker-than-structural truths,' that is, truths which are entailed by structural truths which do not themselves entail further structural truths. Under this modification, the argumentative load shifts to the third premise, that *certain phenomenal truths are neither structural truths nor weaker-than-structural truths*. Such a modified third premise appears well-suited to bear its new load. Even if we granted to the optimistic proponent of IIT that *some* phenomenal truth or other was guaranteed by certain structure-and-dynamical facts about a given conscious system, in the sense that perhaps some system's visual apparatus might a priori imply that it has *some* intrinsic point of view on the visual sense-data it receives, it is far from

[163] Alter 2016, p.801.

obvious that such structural information alone could a priori imply the peculiar phenomenal quality of 'blueness' rather than 'redness' attaching to particular kinds of physical stimuli within the intrinsic point of view of a physically-realised visual system. The phenomenal quality itself, which is underdetermined by the physical truth, would be neither a structural truth, nor a weaker-than-structural truth, unlike the truth that there is some intrinsic perspective-or-other, which would be a weaker-than-structural truth. If such phenomenal qualities were impressively resistant to assimilation to structural truth before Alter's modification, they will be equally impressive after, and the force of the argument overall would be undiminished.

One further way we might entertain for showing that physicalist truths are not limited to structural truths is to argue that structural truths make the physical truth too abstract to plausibly encompass the physical truth about the world.[164] To possess the complete structural truth might not entail all the particular truths about the world. For example, it might be that an idealist world, where the underlying nature of the world is that it is an idea in the mind of God, would correspond to the same kinds of structural truths with which the structuralist identifies physical truths. Given that such idealism is paradigmatically opposed to physicalism, it can't be true that all physicalist truths are structural truths—there must be something to the physicalist truth that distinguishes it from idealism.

This strategy doesn't seem to reduce the dialectical force of the structure-and-dynamics argument too much, especially if physicalism includes a tendency to prescind from deep metaphysical commitments not strictly warranted by the needs of our best physics. If the metaphysical nature of the intrinsic substrate about which the structural truths are true are minimally informative for the purposes of characterising them in formal, spatiotemporal and nomic terms (we might have to grant that *something* possesses the structure, is quantified and

[164] Alter 2016, p.801.

related to other things by means of the kinds of relations pointed out by the structural truth, but this 'something' will plausibly be, for the purposes of physics, minimally significant), then it is difficult to see, given the physicalist's privileging of the practices of empirical physics, why the underlying 'deep' ontological facts which don't make an observable physical difference should matter. It is therefore remains quite plausible given the motivations of physicalism that any world which corresponds to the complete structural truth of this world would be, in any way that matters, physically identical to it.

3.4.2. Phenomenal Truths Might Be Structural Truths.

Another angle of objection to the structure-and-dynamics argument is the crucial premise 3—that at least some truths about consciousness are not deducible from structural truths (or weaker-than-structural-truths). The evidence of this premise is chiefly introspection. As Stoljar (2015) puts it, however, insights into metaphysics are 'not so easy.'[165] The worry might be put in terms of a question-begging worry: if the structure-and-dynamics argument is supposed to provide additional reasons to support the epistemic gap introduced by the common intuitive problems for materialism (e.g., colour scientist, inverted spectrum thought experiments, etc), and a crucial premise of the structure-and-dynamics argument themselves involve an appeal to precisely the intuitions at issue, then there is a circularity here which renders the structure-and-dynamics argument an unpersuasive one to the materialist who doesn't share these intuitions.

The first thing worth emphasising in response to this strategy is the dialectical context. The kind of materialist at whom the structure-and-dynamics argument is chiefly aimed is the kind who admits that there is at least the strong semblance of an in-principle epistemic gap, but that this semblance can be undermined by appeal to the contents of a completed future physics—namely, the type C materialist. One who accepts the appearance of such an epistemic gap does

[165] Stoljar 2015, p.336.

not base their resistance to that gap on sharing a different seeming than the advocate of premise 3, but on the hope that a completed physics will introduce some new kind of truth which can render the seeming unsurprising. Thus, it seems like the structure-and-dynamics argument and the 'standard arguments' in favour of the epistemic gap at the very least draw upon the same strength of intuitions in support of premise 3, and the intuitions for premise 3 are therefore at least not weakened.

To face the objection directly, however, even if it doesn't decisively establish the intuitions leading to the epistemic gap, the structure-and-dynamics argument can still lend some extra weight to the intuitions behind the epistemic gap. To attend to perceptual experience and attend only to its structural aspects (e.g., the shape of the red object), enables one to more precisely attend to that which is subject to that structure (i.e., the redness itself). By attending to the notion of structural truth, one can better observe that one is not merely dealing with the mathematizable structural *aspect* of phenomenal experience, but the 'intrinsic' properties which 'fill out' the structure. Hence, plausibly, applying the insights of the structure and dynamic account of the physical to conscious experience helps one more clearly perceive a positive sense of the 'remainder' of phenomenal properties which cannot be deduced merely from the structural truth.

3.4.3. The Impact of These Objections upon the Plausibility of the Structure-And-Dynamics Argument

The above objections to the structure-and-dynamics argument and the responses to them are not meant to be decisive nor comprehensive. Rather they are set out to give an idea of the ongoing dialectic and the seriousness of the problem that the structure-and-dynamics argument presents for even the critics of its anti-materialist implications. An interesting general feature of the criticisms marshalled against the argument is that few critical strategies propose an alternative *positive* conception of the nature of the physical which can in principle embrace

mental properties. Criticisms like Stoljar's focus on undermining the inference from particular conceptions of structure and dynamics to some body of truths about the world which are epistemically inaccessible from the complete physical truth, but he is content to admit that on many of the notions of structure (considered as relation, for example, or following Pereboom,[166] as non-intrinsic or comparatively-intrinsic properties) leave something out. Part of the dialectical thrust of Stoljar's criticism of the argument is that the differences between type C materialism and type F neutral monism (at least as far as the structure-and-dynamics argument can show) are minimal,[167] provided the meaning of the 'physical' in a complete physics is left open enough. To make this point is precisely to allow that there is little in our present understanding of the physical, or of what the physical-qua-physical is, which could satisfactorily account for the kinds of properties which impress the anti-materialists.

On the other hand, the attractions of something in the vicinity of 'structural' terms as the common thread running through any peculiarly 'physicalist' world-picture are not diminished—the modelling of nature in formal, nomic, relational and spatiotemporal terms and the reduction of puzzling phenomena to truths formulated in such terms remains arguably the driving engine of an advancing physicalist understanding of mind, since the space of truths expressible in such terms encompasses a huge field of possible theoretical development, is informative, and congruent with actual scientific practice. Certainly, theoretical progress in areas such as integrated information theory take themselves to be advancing just such a structural programme into the realm of the intrinsic perspective of complex self-sustaining

[166] Stoljar 2015, p.332.

[167] Stoljar 2015, p.344.

systems, but also acknowledge limits to such a programme, and that a kind of neutral monism remains attractive.[168]

Even if not explicitly acknowledged, the structure-and-dynamics argument shows us why dualism is attractive, and why it remains attractive as long as the physicalist programme has an implicitly structuralist character. Even on a complete physical (structural) description of the world, something 'breathes fire,' as Hawking put it,[169] into any such unified (implicitly structuralist) theory. If one thus acknowledges that 'structure is not enough,' one might be inclined at the least toward something like type C materialism—hoping that a future physics of some unknown content will synthesise the structural truths which empirical science so efficiently discovers with the extra unknown property. Insofar as one treats one's phenomenal experience or some other mental property as some positive thing which is not plausibly assimilable to such a view, one will be inclined toward some kind of dualism or Russellian neutral monism.

3.5. The Hard Problem, Neutral Monism and Panpsychism

The structure-and-dynamics argument thus poses an enduring problem for physicalism, and it is the explanation of why the epistemic gap is an in-principle one for physicalism (i.e., a 'hard' problem), rather than a mere accident of the incompleteness of our current physicalist programme. This sub-section outlines and briefly evaluates some available options for dealing with the kind of 'hard' problem for any variety of physicalism which grants the first premise

[168] Mindt 2021, p.1243.

[169] Stephen Hawking, Ron Miller & Carl Sagan, *A Brief History of Time: From the Big Bang to Black Holes,* Bantam, 1988, p.174.

of the structure-and-dynamics argument—what we might call a 'structure-and-dynamics physicalism, or 'S&D physicalism.'

3.5.1. Option 1: Reject Structure-and-Dynamics Physicalism

One broad class of responses to a hard problem which stems from structure-and-dynamics physicalism, is to reject S&D physicalism either as 1) unwarranted in light of our ignorance of the complete physical truth or 2) as unnecessary in light of some alternative account of a physicalist ontology (i.e., as a positive or negative characterisation), even if one also accepts that structure-and-dynamics physicalism is indeed an enduringly attractive view in some respects.

A problem with option 1), which we might call the mysterian option,[170] is that it effectively vacates the idea of the physical of any substantive content (and, potentially, physicalism's ability to rule out inappropriate content). If, in this absence of an in-principle account of the physical, we still acknowledge that S&D physicalism (despite the problems for physicalism that it generates) remains attractive, we might end up *calling* ourselves physicalists while still confronting a problem between the tractable questions of structural truth and intractable questions of non-structural truth, in-practice. The peace of mind we secure with such a move would be more in line with an aspiration or act of faith than a principled way of containing the difficulties S&D physicalism presents.

Option 2), the presentation of an alternative concept of the physical, can come in two flavours, which we might term 'positive' or 'negative.' An alternative *positive* conception of the physical which exceeded the attractions of S&D physicalism (and thereby made the problems of S&D physicalism less pressing) would have to exceed S&D physicalism in respect of the range of possible theoretical developments it encompasses, informativeness, and congruence with

[170] After Colin McGinn, 'Can We Solve the Mind–Body Problem?' (1989), *Mind*, 98(391), pp.349–366.

scientific practice—all traits which we should want out of any positive conception of the physical. The problem is that there are few such candidates on offer. An explanation of physicalism in terms of fundamental kinds of entities—i.e., atoms, particles, fields, strings, etc., would be too constraining of the theoretical possibilities, since the question of what fundamental entities are is presently in flux and, at any rate, so remote would such entities be from experience that they could only be characterised in terms of the structural truths we can express about them. In absence of an alternative positive conception of the physical, then, the positive alternative approach seems unpromising.

A negative concept of the physical would describe the physical by what it is not, e.g., Wilson's (2006) negative characterisation of the physical as the notion that mental entities are simply not fundamental definition of the physical.[171] Such a move directly challenges the intuitions that there is something fundamental about at least some mental properties (i.e., phenomenal ones), and also decisively rejects certain unintuitive consequences of S&D physicalism, e.g., the notion that the 'intrinsic' properties which 'fill out' a structural description of the world may turn out to be fundamentally mental, as in, to take a limit case, theistic idealism.

Such a 'negative' concept of the physical, compared to S&D physicalism, may have certain advantages. However, as a theory of the physical, it doesn't have as much explanatory or predictive value for the nature of scientific discovery since it doesn't predict or characterise the methods that the physical sciences deliver, and is therefore not as informative (even if it is not completely uninformative). If a more positive characterisation were possible, it would be preferred. Moreover, Wilson's negative concept of the physical seems compatible with the truth of notions like panprotopsychism or 'panqualityism,' which are, intuitively, incompatible with materialism because they posit mysterious 'non-mental' properties which do not seem

[171] Jessica Wilson, 'On Characterising the Physical' (2006), *Philosophical Studies* 131(1), pp.61–99.

penetrable by standard scientific methods of study—i.e., 'qualities' or 'proto-mental' properties. Since S&D physicalism can be framed as leading precisely toward panprotopsychism or panqualityism, as we shall see below, it is not clear that the via negativa provides an effective means of resistance to S&D's anti-physicalist pressure.

3.5.2. Option 2: Accept Dualism

If ignoring or denying S&D physicalism doesn't seem tenable as a means of resisting the pressure it exerts, one might be tempted to cave in to that pressure and simply accept dualism. While there are arguments in favour of a robust substance dualism in the literature (e.g., Swinburne) which trace trajectories similar to property dualism as advocated by Chalmers (e.g., moving from an epistemic to a modal to an ontological to a substantial-independence gap), this subsection will focus mostly on the difficulties (if they are not necessarily decisive difficulties) with property dualism, the notion that there are two fundamental kinds of properties: physical properties and non-physical mental properties.

As with Descartes, the interaction problem and in turn the problem of epiphenomenalism and its associated epistemological difficulties remains a compelling source of distaste for dualism: the physicalist monist purports, via conservation laws discovered a posteriori, to exclude the causal intervention of any exotic properties. It might be contended, for instance, that a complete structural reconstruction of our neural hardware would leave no space for the alleged non-physical properties to make any causal difference. At the very least, there is no reason to suppose that the complete structural truth would not yield such a completely closed account, and reason to believe that our present observations do not so far indicate any exception to the physical rules. Given the success of modern science in exposing and unravelling the connections between biological behaviours and the underlying physical and chemical phenomena, and the dearth of any plausible account of how non-physical properties might exert causal influence, it might be thought at the very least unpromising (if not a decisive reason, if

one has independent reasons to suspect such interactions) to expect that any 'exotic' causal interactions will be discovered or elucidated. There seems to be no causal framework which could explain and derive such apparent interactions even if they were discovered—granting something like S&D physicalism, there would be no progress to be made in understanding such interaction beyond observing the fact of certain unusual 'constant conjunctions.' While Descartes is more directly forced into a robust substance dualism by his underlying ontology, and restricted from interaction by his notion of causality, a modern advocate of substance dualism is less forced by fundamental ontology into substance dualism, and but interaction is also less categorically prohibited by the same.[172] That is, for the contemporary naturalist, causal closure has more of an empirical than an a priori character.

The interaction problem remains acute in the case of property dualism, where the central intuitions driving thought experiments such as the 'zombie' argument which presupposes an exact physical duplicate of our world, and the 'colour scientist' argument which again grants that the colour scientist acquires a complete physical description of the visual faculty down to the physical behaviour of all relevant components, generally grant that the non-physical properties in question make no causal difference. If we treat non-mental properties as strictly epiphenomenal, apart from the theoretical inelegance of having a second aspect of nature beyond what the physical truth strictly requires, such epiphenomenal qualities would generate epistemological difficulties, especially if our knowledge is supposed to be causally related in some way to the object of our knowledge. Indeed, if some sort of causal theory of knowledge were true, we wouldn't even have epistemic access to the fact that we are conscious—while our phenomenal qualities themselves might be caused by the underlying physical processes,

[172] William Lycan, in 'Giving Dualism its Due' (2009), *Australasian Journal of Philosophy*, 87(4), pp.551–563, though not a dualist, regards the possibility of interaction, at least if we grant that Cartesian mental substances, is at least an open question given the failure of physical determinism.

our knowledge of those qualia (or, at any rate, speech acts seemingly motivated by those qualities) could not, on epiphenomenalism, be caused by those phenomenal qualities. Compared to Descartes, modern property dualism is less inclined to embrace fundamental diversity of substance, wanting to allow as much causal self-sufficiency as possible to the accessible physical side of nature and in turn minimise the disturbance to our picture of nature contributed by non-physical properties. This might be thought of as simply a manifestation of the same incompatibility between the mental and the physical which stymies any Cartesian solution to the interaction problem: the non-physical has nothing to contribute to the physical, and so can't contribute anything, which makes the coordination of mental and physical change in turn inexplicable.

Perhaps the deepest sort of problem of dualism is one of explanation and bruteness—if the occurrence and variation of mental properties are not deducible from the physical truth, it becomes necessarily obscure what kind of explanatory nexus coordinates the associations and interactions between the non-physical properties and the physical ones. Denying that causality needs a causal nexus, and that there simply happens to be some kind of co-variation in the occurrence of mental and physical change, in effect makes the problem of explanation more acute. It becomes maximally surprising that such changes should occur, because there is minimal possible explanation for them. While Descartes had God as a metaphysical and explanatory nexus of last resort to generate the coordination between the mental and the physical, the modern dualist who balks at the extravagance of theism also has no such extravagant explanatory resources—the retreat to occasionalism or parallelism is not possible. The modern dualist ought if at all possible locate such an explanation at least in principle within the immanent, in-principle humanly-discoverable world, but if S&D physicalism is true and our best empirical science provides nothing but structural truths, then this project, for all our best methods can tell us even in principle, has dim prospects.

The modern dualistic discourse, then, recognisably inherits from Descartes analogous commitments and conceptual tensions. We have inherited the tensions between the sufficiency and empirical success of the structural (or extended), the insufficiency of the structural (or extended) to account for the primitive facts of experience, and the tension between this insufficiency and the need to account (perhaps in terms of 'deep' ontological commitments beyond what the physical sciences in themselves supply) for the unity and interaction of the mental with the physical. Dualism, even one as seemingly modest as property dualism, has the familiar Cartesian problem of deeply undermining the unity of nature and the elegance of our
2220 theories about nature. While it might be that nature is under no obligation to be 'elegant,' it is still correct to say that this deep inelegance is a strong impetus to investigate an alternative approach which, while doing justice to the difference between the (S&D)-physical and non-physical phenomenal qualities, is more focused on unifying them and can at least in principle begin to explain their co-occurrence.

3.5.3. Option 3: Neutral Monism and its Variants

Neutral monism, descending ultimately from Russell's view as put forward in *The Analysis of Matter,*[173] has enjoyed a resurgence of late as the view of nature which emerges if one grants that there is something deficient in standard physicalism and embraces the property dualism that results. Such a position has attracted interest from, e.g., Chalmers,[174] Alter and

[173] Russell 1927.

[174] Chalmers 2010.

Nagasawa,[175] and Sam Coleman.[176] The basic principle, as Chalmers puts it, is that the 'intrinsic properties' (i.e., non-structural, non-relational properties) of the fundamental physical constituents of nature are precisely the kinds of qualities which are directly accessible to us in phenomenal experience.[177] Such a neutral monism accepts that mental (or at any rate 'proto-mental' non-structural properties) are built into nature at a fundamental level alongside the familiar 'physical' properties: the single substance constituted by such properties is neither, properly speaking, 'physical' nor 'mental,' but 'neutral.' What makes such qualities 'proto' mental is that they are considered to be non-structural properties even apart from the context of familiar mental operations like sensation, imagination or understanding—the characteristic functions of a fully active mind. The non-physical properties of the panprotopsychist are not, therefore, properly mental in nature, even if they are most accessible through the phenomenal experience of minds.

The benefits of neutral monism are that in maintaining monism, the neutral monist retains the theoretical elegance of materialism, i.e., an immanent source of ontological unity. Neutral monism also perhaps mitigates some of the trouble of dualism, in that it diminishes somewhat the opposition between causally active, ontologically significant material properties on the one hand and causally effete, epiphenomenal mental properties on the other: on neutral monism, neither physical properties nor mental properties are causally active in themselves, but only insofar as they are instantiated in their common substance. This restores a certain kind of

[175] Torin Alter and Yujin Nagasawa (eds), *Consciousness in the Physical World: Perspectives on Russellian Monism*, Oxford University Press, 2015, p.1.

[176] Sam Coleman, 'Panpsychism and Neutral Monism: How to Make Up One's Mind', in Godehard Brüntrup and Ludwig Jaskolla (eds), *Panpsychism: Contemporary Perspectives*, Oxford University Press, 2016, pp.249–282.

[177] Chalmers 2010.

metaphysical complementarity between the physical and the qualitative, giving a role to each in a unified ontology. By locating the ontological precedent for phenomenal quality at the fundamental level, the neutral monist avoids the anthropocentric embarrassment of the non-physical emerging at or around the level of reality (i.e., that of the conscious agent) which most concerns ourselves. The move from physicalism to neutral monism is analogous to a reversal of the Cartesian innovation of reifying extension as the principal attribute of physical substance and having thereby to posit an extra mental substance to account for the undeniable facts of mental experience. Rather than structural truths being treated as primary and quality being treated as an inexplicable theoretical baggage, neutral monism weakens but does not abolish the constitutive role of structural features of substances while allowing for the introduction of non-structural features in a complementary (though still somewhat mysterious) role.

Despite the theoretical attractions of neutral monism, there are also obvious and formidable obstacles to the development of neutral monism as a viable and vital metaphysical alternative to physicalism and dualism. Part of the theoretical promise of neutral monism is that it can make phenomenal qualities less 'surprising' by locating antecedents in the proto-psychic properties of the microphysical constituents of nature. However, this gives rise to the well-known 'combination problem'[178]: given that phenomenal qualities seem to have a certain qualitative structure and to emerge and correspond to physical change in an orderly way, it seems that there should exist principles for the production of familiar phenomenal qualities from proto-mental qualities. Yet no such principles are forthcoming, nor does it seem obvious how to begin constructing such principles.[179] It might be thought that phenomenal quality could

[178] William E. Seager, 'Consciousness, Information, and Panpsychism' (1995), *Journal of Consciousness Studies*, 2(3), pp.272–288.

[179] Chalmers 2010, p.136.

be treated as a non-physical property which emerges strictly at the mental level, but that would undermine one of the key attractions of neutral monism. It is not obvious that neutral monism, as much as it is a more attractive 'final vision' of the world, actually helps to make the emergence and behaviour of phenomenal qualities and their relation to physical properties any more scientifically or rationally tractable.

A second difficulty, which might be said to come from the dualist side, might be the charge that neutral monism, in making non-physical properties ubiquitous in nature for the sake of forcing a satisfying theoretical unity onto nature, are imposing a purely stipulative extravagance onto nature in the name of a poorly-motivated expectation that nature 'ought' to be unified, which is in the end no less extravagant than dualism. This difficulty corresponds to the general form of the Cartesian rejoinders examined in **section 2.3.1** to the attempt to unify the 'principal attributes' of thought and extension in a further, more fundamental (but also fundamentally unknown) attribute. The worry might be put this way: if it seems that the physical and phenomenal are diverse and the means of correlating the emergence of one to the other are obscure, why not simply treat them as diverse rather than posit unobservable 'proto-mental' properties which indeed have little prospect of ever being observed? Even if the structure-and-dynamics argument seems to imply that structure needs something else to 'fill it out', why think that phenomenal qualities or their precursors are that 'filling'? To an extent, the rejoinders to this kind of objection will turn on whether one finds the ultimate bruteness of the diversity between physical and the phenomenal on dualism to be more distasteful than the unobservable proto-mental qualities posited by panprotopsychic or pan-qualityist versions of neutral monism.

3.6. On the Way Forward: Some Suggestions from the Medieval Synthesis

While it is not the purpose of this thesis to attempt to solve these challenges to neutral monism, I do venture some (admittedly vague and sketchy) general suggestions drawn from reflection

on the Cartesian flight from scholasticism as to how the neutral monist might fortify his position and make progress toward satisfyingly situating mind in nature.

In **section 1.5** we examined Descartes' abandonment of the scholastic 'metaphysical scaffold' in favour of Cartesian mechanism. Scholastic-Aristotelian hylomorphism was a powerfully attractive metaphysical synthesis because it created a 'metaphysical scaffold' out of complementary ontological notions and categories like substance and accident, universal and particular, actuality and potency, which could be used to frame both the quantitative and non-quantitative aspects of reality as incomplete elements which contributed to a coherent whole. As already noted, one of the great virtues of neutral monism is that it begins to re-establish something like such a complementarity between structural and non-structural properties. However, the sense of participation in complementary metaphysical operations remains 'thin:' the complementarity between the physical and the (proto-)mental is only that of structural to non-structural properties, but our paradigms of combination and emergence are often modelled on the way in which higher-order structural properties emerge from fundamental structural entities (e.g., atoms combining into molecules, molecules into macroscopic accretions, etc). There might well be a mismatch between the kinds of phenomena we are trying to explain, and the kinds of explanation we are willing to accept.

Qualities may not be like microphysical structures which form higher-order structures by means of combination and accretion. The scholastic Aristotelian synthesis, which had a rich menagerie of complementary ontological operations, also had a diverse vocabulary in which to express and understand emergence: not only did the scholastic synthesis contemplate the familiar means of generating the large by combining the small in space (what the scholastics would have called 'accidental' change), but it contemplated qualitative and substantial change: the 'eduction' of entirely new forms of being according to the objective dispositions, formal qualities, and potentials involved. Again, this is not to suggest that the way forward is to adopt

the scholastic Aristotelian synthesis wholesale, but to point out that 'deep' metaphysical reflection on the general ontological operations which transcend the quantitative-qualitative divide and provide the ontological context for them to coexist, can in a non-ad-hoc manner generate new categories which we can then use to articulate the principles of change and causality which hold between the physical and the non-physical.

Likewise, the force of the accusation that the unity the neutral monist seeks is merely stipulative, could be mitigated by the investigation and enrichment of our 'deep' ontological categories—i.e., ontological categories formed at a deeper or more general level than that of ordinary scientific practice. That there is such a 'deep' level of ontological theorising is indeed one of the important implications of the structure-and-dynamics argument against materialism: to come to a full-orbed understanding of reality, it is not enough to understand the world merely in the terms of the S&D physicalist. Given a robust enough (and, ideally, independently motivated) general ontological framework which comfortably[180] situates the structural and non-structural aspects of reality, the idea that the unity of the physical and non-physical is merely stipulative may well lose what bite it has.

The suggestion that 'deep' metaphysical reflection has a role to play in overcoming the present obstacles to the neutral monist programme, itself suggests its own difficulties: the focus on the mechanistic, physico-mathematical and structural aspects of reality was and continues to be important in advancing scientific understanding, and neglecting the role of scientific investigation in supplying us with the data to ontologically interpret could lead to a reversion from a Cartesian aporia about mind to a medieval obsession with abstract categories which don't make a productive empirical difference. But perhaps that is inevitable, and the whole

[180] I.e., has sufficient ontological tools to reduce one's 'surprise' at the existence of non-physical properties as well as explain their occurrence as part of the same underlying kind of reality.

lesson of the excursus into the causes and consequences of dualism might be that any complete and coherent interpretation of the world cannot avoid eventually involving exercises in 'deep' ontology, no matter how long it puts off such an exercise.

Conclusion

The structure-and-dynamics argument against materialism, as is clear from the foregoing survey, represents an ongoing challenge to materialism which in many ways remains bound to the conceptual space Descartes first attempted to navigate, where unqualified improvements are difficult to make and advances are obtained at the price of accumulating costs. To draw together the most pertinent of the common threads, we might state them as follows:

In common with Descartes, the modern idea of the 'physical,' which is very plausibly a 'structure-and-dynamic' view of the physical, is precisely a development of the idea of the Cartesian *res extensa*, with similar theoretical attractions. Because of the great precision and clarity of structure-and-dynamics terms, the programme of seeking out the structural features of the physical world and describing or modelling the world in structural terms promises to deliver insights of great explanatory power, scope, precision and empirical tractability. By de-emphasising the importance of the ontological categories and functions (if any) which underlie the structural features of the physical world or which might lie beyond it, S&D physicalism bypasses possibly intractable metaphysical disputes (as Descartes bypassed those of his scholastic forebears) while delivering useful and productive insights. S&D physicalism can thus serve as a compelling (if usually implicit) 'working metaphysic' of the physical, even as it need not purport to be a full account of fundamental ontology. Yet for all its unifying power, the structure-and-dynamics argument against materialism reveals the limitations of this working metaphysic.

For Descartes, the redescription of the world in terms of the *res extensa* had an obvious theoretical cost: there are aspects of immediate experience which especially resist redescription in 'physical' terms. That 'remainder' was (due to Descartes' peculiar metaphysical commitments) itself reified as a different kind of substance, at the cost (as we saw in Part 2) of causal interaction and the resort to a transcendent and mysterious explanation of such a deeply divided metaphysical order. The S&D physicalist has an analogous 'remainder,' and this thesis has focused on the case of *phenomenal qualities,* or qualia, which at the very least present a class of non-physical properties, if not full-blown substances, which have equally mysterious connections to physical properties. Similarly to some of Descartes' interlocutors, like Arnauld and Princess Elisabeth, the physicalist finds dualism unappealing and the emergence of radically novel non-physical properties at the level of conscious minds inexplicable, and attempts instead to retain the theoretical benefits of metaphysical monism and a single, immanent physical world by pinning her hope on the prospect of future reduction or a posteriori identification with some physical property. If the structure-and-dynamics argument against materialism is successful, however, such a project is doomed to failure, precisely because of the intrinsic limitations of structure-and-dynamic terms.

The space of intermediate possibilities between physicalism and dualism—e.g., the possibility of a 'neutral monism' which combines the mental (or proto-mental) and the physical, promise a degree of resolution: monism promises to harmoniously explain the union of the mental and the physical in terms of a single immanent reality which is fundamentally both mental (or at least proto-mental) and physical, reducing the 'surprise' that we should encounter both structural and non-structural properties in conscious experience. If the difference between 'structural' and 'intrinsic' properties suggested by the structure-and-dynamics argument holds true, neutral monism mitigates the causal separation of the mental and the physical by positing a sense in which they perform complementary ontological functions in physical objects, such

that each owes something to the other. If sustainable, this would be an important and fundamental advance over Descartes, since it is precisely their lack of metaphysical complementarity which makes Cartesian conceptions of mind and matter so difficult to reconcile in a theoretically satisfying manner.

However, such proposals face similar rejoinders to similar proposals made to Descartes: just as Descartes might object, in light of the impressive diversity of thought and extension, that imposing unity between the mental and the physical can come only by ad-hoc (and therefore completely mysterious) fiat, so too the modern neutral monist might be thought to achieve unity (and the promise of a reduction in mystery) only at the price of an equally mysterious fiat. The 'combination problem' for panpsychist or panqualityist versions of neutral monism is analogous to a remaining interaction problem—we have, as yet, no models of what kinds of interactions between the posited 'fundamental,' 'proto-mental' qualities produce the qualities observed in phenomenal experience, and how these interactions between qualities map onto the structural interactions we can observe by means of the physical sciences.

Progress in answering these objections, if we accept the force of the structure-and-dynamics argument implying that a new synthesis which goes beyond materialism is necessary, ought to involve the recovery of a 'deep' metaphysical perspective, which seeks out underlying ontological complementarities which go beyond what can be provided by an empirical science developed to study reality chiefly in structural-dynamic terms. By developing a framework which not only allows fundamental non-structural ontological posits, but clarifies the debts our fundamental ontological posits owe each other, we might avoid an ill-motivated fiat unity of the 'physical' and the 'non-physical,' and develop language appropriate to describe the emergence of familiar qualities from their 'proto-mental' precursors. Descartes buys his physicalist project, which gives us a world entirely penetrable by physico-mathematical methods, precisely by rejecting or minimising the debt that the 'extended' aspects of reality

owe to others. If we, as his successors in that project, are to attempt to restore the place of that debt in our understanding of reality and set to rest the difficulties of an inadequate ontology of the physical—to exorcise the 'ghost of the machine'—we must prepare a space in our synthesis of reality for the 'deep' and indeed ancient metaphysical disputes Descartes wanted to leave behind. Whatever the challenges in reconciling such a task with prevailing scientific practice and naturalistic inclinations, such a restoration seems indispensable if we are to have our surprise at mental properties diminished.

Bibliography

1. Anstey, Peter and Alberto Vanzo, 'The Origins of Early Modern Experimental Philosophy' (2012), *Intellectual History Review,* 22(4), pp.499–518.
2. Aquinas, Thomas, *Commentary on Aristotle's De Anima,* translated by Kenelm Foster and Silvester Humphries, Dumb Ox Books, 1999.
3. Aquinas, Thomas, *Commentary on Aristotle's Physics,* translated by Richard J. Blackwell, Richard J. Spath, and W. Edmund Thirlkel, Dumb Ox Books, 1999.
4. Aquinas, Thomas, *Summa Theologiae,* translated by Thomas Gilby, Image Books, 1969.
5. Ariew, Roger, 'Comments on John Schuster and Frédéric de Buzon concerning Physico-Mathematics and Mathesis in Descartes' (2018), *Journal of Early Modern Studies,* 7(1), pp.175–186.
6. Ariew, Roger, *Descartes and the Last Scholastics,* Cornell University Press, 1999.
7. Aristotle, *The Complete Works of Aristotle, The Revised Oxford Translation*, 2 vols, edited by Jonathan Barnes, Princeton University Press, 2014.
8. Aristotle, *De Anima*, edited by Hugh Lawson-Tancred, Penguin Books, 1986.
9. Armstrong, D. M., *The Mind-Body Problem: An Opinionated Introduction,* Westview Press, 1999.
10. Blancanus, Josephus, 'De Mathematicarum Natura Dissertatio,' *Appendix to Aristotelis Loca Mathematica*, translated by Gyula Klima, in Paolo Mancosu (ed.), *Philosophy of Mathematics and Mathematical practice in the Seventeenth Century,* Oxford University Press, 1996.
11. Bruntrup, Godehard, and Ludwig Jaskolla (eds), *Panpsychism: Contemporary Perspectives*, Oxford University Press, 2016.

12. Chakravartty, Anjan, 'Scientific Realism', *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.),
13. Chalmers, David J., 'Facing Up to the Problem of Consciousness,' in *The Character of Consciousness,* Oxford University Press, 2010.
14. Chalmers, David J., 'Facing Up to the Problem of Consciousness' (1995), *Journal of Consciousness Studies,* 2(3), pp.200–19.
15. Chalmers, David J., 'The Two-Dimensional Argument Against Materialism' in Ansgar Beckermann and Brian P. McLaughlin (eds), *The Oxford Handbook of Philosophy of Mind,* Clarendon Press, 2009.
16. Chalmers, David J., *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, 1996.
17. Cottingham, John, 'Cartesian Trialism' (1985), *Mind,* 94(374), pp.218–230.
18. Dear, Peter, *Mersenne and the Learning of the Schools,* Cornell University Press, 1988.
19. Descartes, René, *The Philosophical Writings of Descartes,* 2 vols, translated by John Cottingham, Robert Stoothoff and Dugald Murdoch, Cambridge University Press, 1984.
20. Descartes, René, *The Philosophical Writings of Descartes, Volume III,* translated by John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny, Cambridge University Press, 1984.
21. Dutton, Blake, 'Descartes' Dualism and the One Principal Attribute Rule' (2003), *British Journal for the History of Philosophy,* 11(3), pp.395–415.
22. Elisabeth of Bohemia (Princess), and René Descartes, *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*, translated by Lisa Shapiro, University of Chicago Press, 2007.
23. Fraassen, Bas C. van, *The Scientific Image,* Oxford University Press, 1980.

24. Gorham, Geoffrey, ‘Descartes’ Dilemma of Eminent Containment,’ (2003) *Dialogue: Canadian Philosophical Review/Revue Canadienne de Philosophie*, 42(1), pp.3–25.
25. Hattab, Helen, *Descartes on Forms and Mechanisms,* Cambridge University Press, 2009.
26. Hawking, Stephen, Ron Miller and Carl Sagan, *A Brief History of Time: From the Big Bang to Black Holes,* Bantam, 1988.
27. Hempel, Carl G., ‘Reduction: Ontological and Linguistic Facets,’ in Sidney Morgenbesser, Morton White, Patrick Suppes and Ernest Nagel (eds), *Philosophy, Science and Method: Essays in Honor of Ernest Nagel,* St Martin’s Press, 1969, pp.179–199.
28. Hoffman, Paul, ‘The Unity of Descartes’ Man’ (1986), *The Philosophical Review* 95(3), pp. 339-370.
29. Hoffman, Paul, ‘Union and the Interaction of Mind and Body,’ in Janet Broughton and John Carriero (eds), *A Companion to Descartes*, John Wiley & Sons, 2007.
30. Hooker, Michael (ed.), *Descartes: Critical and Interpretive Essays*, The Johns Hopkins University Press, 1978.
31. Hugh of St Victor, *The Didascalion of Hugh of St Victor*, translated by Jerome Taylor, Columbia University Press, 1961.
32. Jalobeanu, D., and C. T. Wolfe (Eds), *Encyclopedia of Early Modern Philosophy and the Sciences*, Cham: Springer, 2020.
33. Lavazza, A., and Robinson, H. (eds), *Contemporary Dualism: A Defense* (1st ed.), Routledge, 2013.
34. Longeway, John, ‘Medieval Theories of Demonstration’, The Stanford Encyclopedia of Philosophy (Spring 2021 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2021/entries/demonstration-medieval/>.

35. Lycan, William, ‘Giving Dualism its Due’ (2009), *Australasian Journal of Philosophy*, 87(4), pp.551–561.

36. Manning, Gideon, ‘Three Biased Reminders about Hylomorphism in Early Modern Science and Philosophy’, in Gideon Manning (ed.), *Matter and Form in Early Modern Philosophy*, Brill, 2012.

37. McGinn, Colin, ‘Can We Solve the Mind–Body Problem?’ (1989), *Mind*, 98(391), pp.349–366.

38. Mersenne, Marin, *Questions Théologiques, Physiques, Morales Et Mathématiques*, composées par L. P. M. Paris: Henry Guenon, 1634.

39. Mindt, Garrett, ‘Not All Structure and Dynamics Are Equal’ (2021), *Entropy*, 23(9), p.1226.

40. Nadler, Steven, ‘Descartes and Occasional Causation,’ in *Occasionalism: Causation among Cartesian*, Oxford University Press, 2010, pp.29-47.

41. Nagasawa, Yujin and Torin Alter (eds), *Consciousness and the Physical World*, Oxford University Press, 2015.

42. Pereboom, Derk, ‘Russellian Monism and Absolutely Intrinsic Properties,’ *Current Controversies in Philosophy of Mind*, Routledge, 2014, pp.40–69.

43. Rozemond, Marleen, ‘Descartes’ Case for Dualism’ (1995), *Journal of the History of Philosophy*, 33(1), pp.29–63.

44. Rozemond, Marleen, *Descartes’ Dualism,* Harvard University Press, 1998.

45. Rozemond, Marleen, ‘Descartes on Mind-Body Interaction: What’s the Problem?’ (1999), *Journal of the History of Philosophy*, 37(3), pp.435-467.

46. Russell, Bertrand, *The Analysis of Matter*, Routledge, 1927.

47. Sánchez, Francisco, *That Nothing Is Known = (Quod Nihil Scitur)*. Cambridge University Press, 1988.

48. Scotus, John Duns, *Ordinatio II*, 2012, translated by Peter L.P. Simpson, <https://www.aristotelophile.com/Books/Translations/Scotus%20Ordinatio%202%20dd.4-44.pdf>.

49. Seager, William E., 'Consciousness, information, and panpsychism' (1995), *Journal of Consciousness Studies,* 2(3), pp.272–288.

50. Stoljar, Daniel, 'Physicalism', *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/physicalism/>.

51. Suarez, Francisco, *On the Formal Cause of Substance: Disputation XV,* translated by John Kronen, Marquette University Press, 2000.

52. Suarez, Francisco, *Opera Omnia vol 25: Disputationes metaphysicae,* InteLex Corporation 2017.
URL = <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>.

53. Wigner, Eugene P., 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences' (1960), *Communications of Pure and Applied Mathematics,* 13(1), pp.1–14.

54. Wilson, Jessica, 'On Characterizing the Physical,' (2006), *Philosophical Studies,* 131(1), pp.61–99.