CONTENTS

CHAPTER 1

Cartesian Explanation

One of the defining features of the scientific revolution in the seventeenth century was its radical reform of the concept of explanation. Natural philosophers of this period, including Descartes, criticized trenchantly and consistently the style of explanation that was widely attributed to the scholastics and proposed instead a new ideal of what is often called, somewhat misleadingly, mechanical explanation. The success of this critique was such that, by the close of the century, the concept of explanation has been effectively redefined. In this chapter I review some of the arguments that supported this fundamental change in our understanding of what constitutes an explanation, with particular reference to their implications for a Cartesian theory of mind. In doing so I shall not quote extensively from the many writers who enthusiastically embraced the new ideal during this period. They are both too numerous and too well known. However, to avoid giving the impression that Descartes was particularly reductionist or that his methodological assumptions depended significantly on an atypical metaphysics, I shall refer briefly to comparable views that were espoused with equal conviction by Robert Boyle (1626–91). Boyle evidently belonged to a slightly later generation of natural philosophers. Although familiar with Descartes's work, he was an independent natural philosopher whose primary intellectual circle was the Royal Society in England. I refer to Boyle's parallel and similar critique of scholasticism to illustrate the extent to which this new model of explanation was widely shared by natural philosophers in the seventeenth century. In fact, if one allows for slight variations in emphasis or expression, this is almost the unanimous view about explanation held by all the principal proponents of the scientific revolution from Galileo (1564–1642) to Newton (1642–1727).

Despite this near unanimity, however, it is usually assumed that the scope of the new model of explanation was restricted to natural phenomena, and that phenomena associated with the mind—such as thinking, remembering, or consciousness—are an exception, in principle, to its otherwise universal application. In the case of Descartes, the reason usually offered for this differential approach to nature and the human mind is that he first established a dualistic

ontology, and that the substance dualism for which he is renowned in the history of philosophy decides the question about how to construct a theory, and especially how not to construct a theory, of the human mind. If mind and body are completely different kinds of substance, then the style of explanation that is appropriate to the latter may be completely inappropriate in the case of the mind.

There are two good reasons for not assuming this division of labour a priori. One is that Descartes and his contemporaries were particularly convinced of the negative evaluation of scholastic explanations, even if they continued to debate the competing merits of alternative proposals. They were convinced that the objections against scholastic 'explanations' were so fundamental that they were never genuinely explanatory. Such explanations did not work in the case of natural phenomena, and the problems that rendered them useless there did not disappear if one simply changed the phenomena to which they were applied. If anything, the debilitating flaws of scholastic explanations became more apparent when applied to mental phenomena, which, then as now, constituted a particularly intractable challenge to our attempts at explanation. The second reason was that the plausibility of substance dualism depends on the success or otherwise of applying the new scientific methods to the human mind. Unless Descartes is understood as inheriting substance dualism uncritically from his predecessors, he needs convincing arguments to persuade readers that the mental life of some biological machines is not amenable to the new style of explanation and, consequently, that we should think of human beings as combinations of two radically different types of reality. Substance dualism should follow from, rather than precede, a thorough examination of the feasibility of explaining the human mind scientifically. We should therefore expect that, in constructing a theory of mind, Descartes would not revert spontaneously to the failed concept of explanation against which he had argued so successfully, and that he would at least test the applicability of the novel explanatory strategies that he proposed with almost evangelical zeal from his earliest writings.

When this approach is tried and reaches its limits—and one would expect those limits to emerge rather quickly, given the minimal knowledge of the human brain with which Descartes was working—Descartes has a number of options available (although the following suggestions are not meant to exhaust the list). One is simply to acknowledge the boundaries of his success to date and to talk about the mind as what remains unknown. A modified version of the same response is to describe the mind as what is partly known and understood, and to signal the limits of his success in explaining mental phenomena by relying on dispositional explanations whose apparent circularity and other

limitations (discussed below) are acknowledged. A third option is to argue for a much stronger conclusion—namely, that the failure of one's explanatory efforts vis-à-vis the human mind implies a compelling, a posteriori, justification for ontological dualism.

The most plausible interpretation of the Cartesian solution can be decided only after considering the relevant texts and arguments. In this chapter, I outline the very strong reasons that should prevent any committed natural philosopher of the seventeenth century from reverting to empty scholastic explanations. If followed to their logical conclusion, these criticisms require Descartes to look for an alternative, less unsatisfactory, response to the genuine problems encountered in explaining mental phenomena. At the conclusion of this chapter, I indicate one 'way out' adopted by Digby, and I defer to later chapters any final interpretation of Descartes's solution.

CRITIQUE OF SCHOLASTIC EXPLANATIONS

The most fundamental issue on which most natural philosophers in the seventeenth century agreed was their rejection of the style of explanation that was widely shared in the schools. Some of the principal objections to 'substantial forms' and 'real qualities' in an explanatory context were articulated at length by Boyle, sixteen years after Descartes's death, in *The Origin of Forms and Qualities* (1666):

> *First*, That I see no necessity of admitting in Natural things any such substantial Forms, Matter and the Accidents of Matter being sufficient to explicate as much of the Phaenomena of Nature, as we either do or are like to understand. *The next*, That I see not what use this Puzling Doctrine of substantial Forms is of in Natural Philosophy . . . the *third*, which is, That I cannot conceive, neither how Forms can be generated . . . nor how the things, they ascribe to them, are consistent with the Principles of true Philosophy . . . (Boyle 1999–2000: v. 340)[1]

The second objection, that forms make no contribution to our understanding of any phenomenon, follows from the alleged incomprehensibility of the forms in

[1] Rozemond (1998: 102–11) points out that many commentators today conflate substantial forms and real qualities, although they were different theoretical entities for the scholastics. Substantial forms were introduced to explain substantial change, as when an object of a certain kind either comes into existence or ceases to exist. Since, according to this account, a substance is a combination of a substantial form and matter, a substance comes into existence by the arrival into undifferentiated matter of the relevant form. Descartes understood qualities as features of things that are not themselves things or substances, and he claimed that scholastic philosophers introduced so-called real qualities because they believed that, 'without them, sense perception could not be explained' (vii. 435).

question, and from the claim that we know nothing more about forms than what is already known about the phenomena to be explained.

> I do not remember, that either *Aristotle* himself . . . or any of his Followers, has given a solid and intelligible solution of any one Phaenomenon of Nature by the help of substantial Forms: which you need not think it strange I should say, since the greatest Patrons of Forms acknowledg their Nature to be unknown to Us, to explain any Effect by a substantial Form, must be to declare (as they speak) *ignotum per ignotius* [i.e. what is known by what is unknown], or at least *per aeque ignotum* [i.e. by what is equally unknown]. And indeed, to explicate a Phaenomenon, being to deduce it from something else in Nature more known to Us, then the thing to be explain'd by It, how can the imploying of Incomprehensible (or at least Uncomprehended) substantial Forms help Us to explain intelligibly This or That particular Phaenomenon? For to say, that such an Effect proceeds . . . from its substantial Form, is to take an easie way to resolve all difficulties in general, without rightly resolving any one in particular. (Boyle 1999–2000: v. 351–2)

These three objections, concerning: (1) the redundancy of forms, (2) their lack of explanatory value, and (3) the way in which their proponents fail to account for their origin, are typical of many corpuscularian philosophers of this period. One is not surprised to find that Descartes and many Cartesians expressed the same objections in almost identical terms.

Descartes raised the question about the redundancy of forms, as theoretical entities, in Chapter 2 of *The World* (1632). In the course of elaborating his own preferred explanation of what happens when a fire burns a piece of wood, he wrote:

> When it [a fire] burns wood or other similar material we can see with our eyes that it moves the small parts of the wood, separating them from one another . . . Someone else may if he wishes imagine the 'form' of fire, the 'quality' of heat, and the 'action' of burning to be very different things in the wood. For my own part, I am afraid of going astray if I suppose there to be in the wood anything more than what I see must necessarily be there . . . For you can posit 'fire' and 'heat' in the wood and make it burn as much as you please . . . provided only that you grant me that there is some power that violently removes its more subtle parts and separates them from the grosser parts, I consider that this alone will be able to bring about all those changes that we observe when the wood burns. (x. 7–8; *W*. 6–7)

Thus, in contrast with the style of explanation that prevailed among scholastics, Descartes limits his conceptual repertoire in this context to the size, shape, and motion of small parts of matter. If he can construct a viable explanation of natural phenomena by using only those properties, he argues, then the various forms and qualities to which scholastic philosophers appeal are redundant. In

fact, not only are such forms and qualities redundant; they themselves also need to be explained.

> If you find it strange that, in explaining these elements, I do not use the qualities called 'heat', 'cold', 'moistness', and 'dryness', as the Philosophers do, I shall say that these qualities appear to me to be themselves in need of explanation. Indeed, unless I am mistaken, not only these four qualities but all others as well, including even the forms of inanimate bodies, can be explained without the need to suppose anything in their matter other than [the] motion, size, shape and the arrangement of its parts. (x. 25–6; *W*. 18)[2]

The same opinion, slightly qualified by the irenic motives of its author with respect to his Dutch correspondent, is expressed ten years later in a letter to Regius in 1642. On this occasion Descartes is counselling Regius about how best to avoid needless controversy with scholastic theologians at the University of Utrecht. He advises his protagonist that it is best not to reject explicitly the real qualities and substantial forms to which Reformed theologians appeal, and even to retain them in name while offering new arguments that imply their redundancy. This was the strategy that he had adopted, he claims, in *The Meteors*: 'Do you not remember that, in the *Meteorology* (page 164), I warned in the most explicit terms that I did not reject or deny them in any way, but only that I did not need them to explain my theories? If you had followed the same plan, no one among your listeners would have failed to reject them as soon as they saw that there was no use for them' (iii. 492).[3] Even more explicitly, Descartes recommends that Regius write to Voetius, the quarrelsome rector of the University of Utrecht, that the 'harmless entities' in dispute should not rashly be rejected and that Cartesians merely claim that 'we do not need them in order to provide the causes of natural things' (iii. 500).

It should be acknowledged that, despite this apparently conciliatory advice to Regius, Descartes denies that substantial forms have any genuine function when explaining natural phenomena. However, he also makes an exception of the human soul in the very same context in which he urges the elimination of all substantial forms from natural philosophy. This coincides with the explicit

[2] Cf. Descartes to Morin (13 July 1638), where Descartes appeals to the relatively few theoretical entities he needs compared with the almost unlimited range of forms and qualities required by scholastic philosophers (ii. 199–200; *D*. 76–7).

[3] Cf. *The Meteors*, First Discourse (vi. 239): 'Know also that, to maintain the peace with the philosophers, I do not at all wish to deny what they imagine in bodies over and above what I have claimed, such as their substantial forms, their real qualities, and similar things; for it seems to me that my explanations ought to be approved so much more readily in so far as I make them depend on fewer things.' See also Descartes to Mersenne (28 Jan. 1641), in which Descartes claims to have included all the foundations of his physics in the *Meditations,* but asks his correspondent not to mention this fact to supporters of Aristotle. Instead he hopes that readers may become accustomed to Cartesian principles and recognize their truth 'before noticing that they destroy the principles of Aristotle' (iii. 298).

exception made for the human mind or soul by Boyle (1999–2000: v. 300): 'When ever I shall speake indefinitely of Substantial forms, I would always be understood to except the Reasonable Soule, that is said to inform the humane Body; which Declaration I here desire may be taken notice of, once for all.'[4] I return to the justifiability and implications of this exception below.

The second reason for excluding substantial forms from explanations of natural phenomena is that they are obscure realities that are poorly understood even by their proponents, and are so specifically tailor-made or ad hoc that they provide a facile pseudo-explanation of every conceivable phenomenon without making any progress in genuine understanding.

> proponents [of substantial forms] admit that they are occult and that they do not understand them. If they say that some action results from a substantial form, that is the same as saying that it results from something that they do not understand; which explains nothing. . . . In order to provide explanations easily of everything (if indeed an explanation of anything is provided when what is obscure is explained by what is more obscure), they have invented substantial forms and real qualities; in this enterprise their ignorance is not at all learned, but ought to be described instead as vain and pedantic. (iii. 506, 507)[5]

The core of this objection was not simply that the terms used by the scholastics appeared obscure to others who worked outside that philosophical tradition; that type of objection could be made by an unsympathetic critic of any theory. The objection was much more fundamental: that one cannot *explain* any phenomenon merely by attributing a quality or form to it which is named after the effect to be explained.[6] Otherwise, we could equally well 'explain' why houses

[4] Cf. Boyle 1999–2000: v. 340: 'But the summe of the Controversy betwixt Us and the Schools is this, whether or no the Forms of Natural things, (the Souls of Men always excepted) be in Generation educed, as they speak, *out of the power of the Matter*, and whether these Forms be true *substantial Entities*, distinct from the other substantial Principle of Natural Bodies, namely Matter.'

[5] Cf. Letter to Father Dinet (vii. 592), and *A Description of the Human Body* (xi. 243; *W*. 181): 'Now supposing that the heart moves in the way that Harvey describes, not only must we imagine some faculty which causes the movement, the nature of which is much more difficult to conceive than what it is invoked to explain; we must also suppose the existence of yet other faculties that alter the qualities of the blood while it is in the heart.'

[6] Cf. Locke (1975: II. xxi. 20) for the same objection to attributing faculties to either the mind or the body. 'But the fault has been, that Faculties have been spoken of, and represented, as so many distinct Agents. For it being asked, what it was that digested the meat in our Stomachs? It was a ready, and very satisfactory Answer, to say, That it was the *digestive Faculty*. What was it that made any thing come out of the Body? The *expulsive Faculty*. What moved? the *motive Faculty*: And so in the Mind, the *intellectual Faculty*, or the Understanding, understood; and the *elective Faculty*, or the will, willed or commanded . . . Which ways of speaking, when put into more intelligible Words, will, I think, amount to thus much: That Digestion is performed by something that is able to digest; Motion by something able to move; and Understanding by something able to understand. And in truth it would be very strange, if it should be otherwise; as strange as it would be for a Man to be free without being able to be free.'

with structural defects collapse simply by saying that they have a 'collapsing form', or why successful therapies cure people by saying that they have a 'therapeutic form'. In the caricature borrowed from Cartesian philosophy and made famous by Molière, we would explain why sleeping pills have their desired effect by saying that they have a 'dormitive power'.[7] Thus forms and qualities are obscure entities that are equally in need of explanation as the phenomena they are designed to explain.

Malebranche suggested that the willingness of scholastics to endorse 'occult qualities or imaginary faculties' was due to a lack of confidence in their powers of observation and a failure to understand the need to hypothesize unobservable causes. When natural philosophers fail to see, by the unaided eye, the unobservable particles that, according to Malebranche, are the real causes of natural phenomena, they tend to have recourse to faculties and powers that explain nothing: 'They resort to qualities in the moon, rather than to the pressure of the air surrounding the earth in order to explain the tides, and to forces of attraction in the sun, rather than to the impulses caused by the particles of subtle matter it continuously diffuses, in order to explain the rising of vapours' (Malebranche 1997: 30). Here the Oratorian's objection assumes the viability of some kind of corpuscular theory, and contrasts its superior explanatory resourcefulness with the alleged deficiencies of the scholastic alternative. However, the central Cartesian objection could have been expressed just as easily without seeming to beg the question of which theoretical framework is most likely to succeed. Even if the new corpuscular philosophy were to fail completely, there could be no justification for reverting to scholastic explanations in terms of real qualities and substantial forms.[8]

It is clear that 'occult', in this context, does not mean simply 'hidden' or 'unobservable'. Neither Descartes nor other corpuscularians of the seventeenth century objected to postulating the existence of extremely small parts of matter that, because of their size, were unobservable and therefore hidden or occult. The theoretical entities of the scholastic tradition were rejected as occult in a very different sense. From the perspective of their critics, they were not too small to be observable, but too obscure to be intelligible.

Finally, Descartes argues that, even if one granted the existence of what scholastics called real qualities and substantial forms, one would still not under-

[7] In the closing scene of *Le Malade imaginaire*, Doctor Bachelierus explains how opium makes one fall asleep by saying that it has a 'virtus dormitiva, cujus est natura sensus assoupire' (Molière 1971: ii. 1173).

[8] Rozemond (1998: 116) suggests that 'the rejection of real qualities and substantial forms depends significantly on the virtues of Descartes's mechanistic explanations'. I argue that the objections to scholastic explanations were independently valid.

stand how such entities interact with other features of physical bodies that we can understand unproblematically, such as their size, shape, or motion.

> However, we cannot in any way understand how something that is completely different from them in nature is produced by these same things (namely, by size, shape, and motion), such as those substantial forms and real qualities that many people assume in things; nor how subsequently these same qualities or forms would be able to cause local motion in other bodies. (viii-1. 322)

This conceptual or explanatory chasm between substantial forms and the kind of qualities that are described and explained in Cartesian natural philosophy—such as the size or motion of bodies—does not diminish over time, nor does it lose its significance when the explanandum in question is mind–body interaction.

Thus Descartes's consistent attitude throughout his mature philosophy was to object to the explanatory framework to which many scholastic philosophers of the period appealed. Granted, the objection was articulated from the perspective of an alternative explanatory framework, one in which only parts of matter and a very limited number of their properties were admitted. The explanatory superiority of the latter and the obviousness of its appeal were so taken for granted that, towards the conclusion of the seventeenth century, one finds John Locke giving an otherwise surprising answer to a question about how external objects cause ideas in the human mind: 'The next thing to be consider'd, is how bodies produce Ideas in us.' He answers: 'manifestly by impulse, the only way which we can conceive bodies operate in' (Locke 1975: II. viii. 11). This attitude to the very conceivability of the ways in which material bodies interact was matched by correspondingly exaggerated claims, on the part of proponents, about the degree of success already achieved by the new corpuscular philosophy. The rhetoric of success camouflaged the need to engage in a genuine evaluation of the limited progress in scientific explanation that had been made to date, and especially of the possible need to expand the range of variables required for a successful scientific research programme. In this respect, Descartes shared the unwarranted confidence of his contemporaries in the new sciences and in the fruitfulness of the conceptual framework within which they were developed.

However, even if all these qualifications are conceded, the fundamental objection to scholastic explanations remains valid. This objection hinges on the minimal value of appealing to forms and qualities when what is needed is an explanation of some phenomenon. What was at issue at this juncture was the concept of explanation itself.

STRUCTURAL EXPLANATION

Cartesian natural philosophy represented a fundamental challenge to the assumption that any phenomenon is explained by inventing a corresponding quality or form. The task of articulating what precisely was defective about scholastic explanations was almost overlooked in the unanimity with which they were rejected. Boyle, for example, takes on this task by appealing to an implicit standard of *structural* explanation.

> to explicate a phenomenon, it is not enough to ascribe it to one general efficient [cause], but we must intelligibly show the particular manner how that general cause produces the proposed effect. He must be a very dull inquirer who, demanding an account of the phenomena of the watch, shall rest satisfied with being told that it is an engine made by a watchmaker, though nothing be thereby declared of the structure and coaptation of the spring, wheels, balance and other parts of the engine, and the manner how they act on one another, so as to co-operate to make the needle point out the true hour of the day. (Boyle 1996: 150)

The effectiveness of this objection depends, to some extent, on how little the listener or reader already knows about watchmakers. If all they know is that the word 'watchmaker' means 'someone who makes watches', then the emptiness of the proposed account is doubly debilitating. It tells us nothing about (1) the specific skills of a watchmaker, or about (2) how the various parts of a watch work together to move the hands. In an obvious way, it merely names the relevant cause as 'the kind of cause that can give rise to the effect to be explained', without telling us anything about either the cause itself or how it works.

Boyle's reference to a watchmaker makes explicit a central assumption of Cartesian explanation, in contrast with the deductive–nomological model of explanation that has been widely discussed in philosophy of science in the twentieth century. According to that model, the occurrence of a particular event or phenomenon is explained by deducing a description of a single event from a universal law and relevant descriptions of initial conditions. Thus a phenomenon is said to be explained simply by showing how it is a particular instance of a more general pattern in nature. One is hesitant to claim that such an account has absolutely no explanatory power, but its limits are evident. There is little progress made in explaining why a piece of metal conducts electricity if we are told simply that all pieces of copper conduct electricity (and that this metal is copper). In fact, the explanatory limitations involved here are very similar to those to which Cartesians objected.[9]

[9] The contrast between Cartesian and deductive-nomological explanations is made explicitly by Baker and Morris (1996: 149 ff.).

In the rush to expose the vacuity of explanations in terms of forms and, in the case of natural phenomena, to substitute structural explanations, even Cartesians had to acknowledge that there might be a limited role for an appeal to fundamental dispositional qualities whose description involves the kind of circularity that had been alleged against forms. Antoine Arnauld, who, in this respect at least, was an orthodox Cartesian, assumed the task of articulating when forms or qualities were acceptable, and when they were not. In *On True and False Ideas* (1683), he distinguished between their correct and incorrect usage.

> But why do the Cartesian gentlemen have such an aversion to the general terms 'nature' and 'faculty' when the Peripatetics use them? . . . these are words which can be used correctly or incorrectly. The word 'faculty' is used incorrectly when one understands by it something distinct from the thing to which one attributes the faculty . . . when one claims to have given an explanation of an effect that is known . . . as when one says that the magnet attracts iron because it has this faculty . . . the abuse of the word consists principally in this: before knowing what is involved in iron being attracted to a magnet . . . one is satisfied with saying that the magnet [has an attractive force]. (Arnauld 1990: 153)

The objection to such ad-hoc forms was that they involved a premature short-circuiting of the work required to construct an explanation. Without knowing or even guessing the cause of a given phenomenon, one says simply that it has the kind of form that can cause the effect in question. This is trivially true rather than false, and if the triviality of the claim were recognized there could be little objection to it apart from the waste of time involved. But many proponents of forms and qualities seem to have thought of them as real entities of some kind that were distinct from the phenomenon to be explained.[10]

Isaac Newton argued, for similar reasons, against what he described as occult qualities while recognizing their limited role in scientific explanation. In Query 31, at the conclusion of the *Opticks*, Newton was anxious to distinguish the mathematical and experimental character of his own work from what he took to be the unsubstantiated speculation of the Cartesian tradition. In particular, he wished to explain that inertial force and 'certain active principles, such as is that of Gravity', were not occult in any objectionable sense of the term.

> Such occult Qualities put a stop to the Improvement of natural Philosophy, and therefore of late Years have been rejected. To tell us that every Species of Things is endow'd

[10] A similar analysis is offered by Boyle (1996: 32–3). He had no objection, he said, to using the phrase 'concocting faculty' if it were used 'compendiously [to] express several things together by one name' (namely, all the bodily functions involved in eating and digesting food), on condition that it did not imply the existence of a distinct 'real existent being'.

with an occult specifick Quality by which it acts and produces manifest Effects, is to tell us nothing: But to derive two or three general principles of Motion from Phaenomena, and afterwards to tell us how the Properties and Actions of all corporeal Things follow from those manifest Principles, would be a very great step in Philosophy, though the Causes of those Principles were not yet discover'd. (Newton 1952: 401–2)

The accuracy or otherwise of Newton's description of his own scientific method is a complicated issue that need not be pursued here. But one part of his objection in the *Opticks* is clear enough. To explain every specific natural phenomenon in terms of a form that is tailor-made for that particular case is to make no progress in explanation, 'to tell us nothing'.[11] In contrast, he argues, it is acceptable to identify a range of phenomena that may be explained by reference to a single force or active principle, even if the nature of that active principle remains unexplained. That would represent some progress in explanation in the following sense: we would have discovered that a single principle or force is active in a number of disparate phenomena, even if the principle in question is described in terms of the phenomena to be explained. At the same time, one might also anticipate that, in time, the provisionally adopted principle would be further explained.

Arnauld suggests that forms or faculties may have a legitimate if limited use in natural philosophy, similar to Newton's defence of what he called 'principles', but only a provisional function:

But if, after having explained, as Descartes does . . . what the attraction of iron by a magnet is and what the magnet contributes to it, one then asked how it comes about that . . . the magnet has screw-shaped pores, then it would be perfectly acceptable to reply by saying that it is because such is the nature of the bodies that we call . . . magnets . . . if one asks in general terms why matter is able to move, it is perfectly proper to reply by saying that this is its nature, and that God, in creating it, has given to its parts this faculty by which one of them can be moved closer to, or further from, another. (Arnauld 1990: 153–4)

In other words, an appeal to forms or qualities signals an impasse, more or less temporary, in constructing an explanation. In the case of Newton, he assumed that he could eventually discover more fundamental theoretical entities than the forces to which he appealed in the *Principia* and *Opticks*.[12] In the case of

[11] This central objection is also found, for example, in Newton's unpublished manuscripts: 'what certainty can there be in a Philosophy which consists in as many Hypotheses as there are Phaenomena to be explained' (quoted in Westfall 1980: 643).

[12] Newton's various hypotheses about what those principles might be, and whether they would ultimately be mechanical, spiritual in some sense, or discovered from his researches into alchemy, were the object of his continued research over a long time. See McGuire (1968), McMullin (1978*b*), and Westfall (1980: 298–307, 638–48).

Arnauld, he wished to acknowledge that our explanations must come to a stop at some point and, wherever that is, it is not objectionable to say that phenomena occur as they do because of some ultimate dispositions in matter. It may be that these dispositions, in turn, are identifiable only by reference to the observable properties that they cause, and in this respect they exhibit a circularity that resembles scholastic explanations.

Thus there is no absolute ban among natural philosophers of the seventeenth century on an appeal to ultimate features or dispositions of matter, even if such qualities are described in terms of the effects that they explain. However, this solution cannot be adopted, on a case-by-case basis, for every specific phenomenon for which an explanation is sought. Likewise an appeal to 'nature' may legitimately signal the limit—or the limits to date—of an explanatory programme, but one cannot explain every specific natural phenomenon simply by a reference to its nature. If these qualifications are not observed, one's explanations in terms of occult faculties, powers, or natures are trivially circular and uninformative.

Another assumption built into the theory of explanation adopted by Cartesians is that, if theoretical entities are introduced, they should be understood at least as well as the phenomenon to be explained. Once this is accepted as a general principle, it is then a contingent matter of fact to determine which realities are understood at a particular point in time, and which are still beyond our comprehension. This requirement applies both to the kind of realities to which one appeals in an explanation, and to the manner in which they act as causes (or as part of a viable explanation) of whatever phenomenon one wishes to explain. The objection to substantial forms and real qualities in this context was that, as theoretical entities, they fail on both counts. Cartesians claimed that they could neither understand what substantial forms were nor, even if their existence were provisionally conceded, how they could be put to work to cause the relevant phenomena.

In contrast, the kinds of theoretical entity that seventeenth-century natural philosophers were willing to endorse were a function of their limited understanding, at that time, of the range of variables required for an adequate natural philosophy. Without being able to stipulate in advance what would be unacceptable, they could offer paradigm examples of what fell unambiguously within the scope of their understanding. For example, Descartes argued that everyone knows that one body in motion can cause another body to move as a result of impact between the two. This does not presuppose any theoretical understanding of the forces involved, or of the mathematical calculations required to quantify the results of a particular impact. It is simply an everyday experience that, when two relatively hard bodies collide, unless one body

collapses or disintegrates on impact, they either spring back from each other or continue their motions with a modified speed or direction.[13] Since this is a phenomenon with which we are already familiar from our ordinary experience, we could equally well understand if something like this were to occur at the level of extremely small, unobservable pieces of matter. As Descartes writes in reply to Morin, in 1638: 'In the analogies I use, I compare only some movements with others, or some shapes with others, etc.; that is, I compare those things that because of their small size are not accessible to our senses with those that are, and that do not differ from the former more than a large circle differs from a small one' (iii. 67–8).[14] There is no justification in this argument for claiming that nothing apart from the size, shape, and movements of small parts of matter could possibly be accepted as variables in a theory. However, the way in which the concept of explanation developed during the scientific revolution was constrained by the haunting spectre of regressing to the occult qualities of the scholastic tradition, so that any theoretical entity that was either less familiar or less well understood than the impact of colliding pieces of matter, or that had worrying connotations of occultness, was pre-emptively excluded from consideration as part of a viable theory. The fate of forces in the history of dynamics in this period is sufficiently well documented to illustrate the point. Even 'the incomparable Mr Newton', as Locke described him, was sufficiently concerned about the taint of occultness that he searched for some acceptable explanation of gravitational effects and preferred not to publish his more speculative hypotheses about active principles and the ether.

One of the most immediate consequences of this concept of explanation was a significant reduction in the complexity and sophistication of the concept of matter with which natural philosophers of the seventeenth century worked. The matter included in theories from Descartes to Locke was one that could be described adequately in terms of very few fundamental properties, all of which were familiar to observers from their everyday experience. The relatively meagre explanatory success of this concept of matter was a function of the conceptual limitations adopted by its proponents. It would make no sense to transpose such a concept from the seventeenth century to the twenty-first, and to assume that the term 'matter' had not changed meaning during the very significant developments in scientific theory that have occurred in the intervening cen-

[13] As is well known, some of Descartes's impact rules seemed to contradict our everyday experience. The only relevant point here is that we know from experience that bodies bounce off each other on impact, whatever the details of the subsequent redistribution of their original motions.

[14] This suggestion was first made in the *Cogitationes privatae*: 'Human knowledge of natural things is acquired only by analogy with those things that fall within the scope of the senses. Indeed, we consider that the one who has philosophized best is the person who has most successfully assimilated what is sought to what is known by sensation' (x. 218–19).

turies.[15] Therefore, when Descartes speaks about what matter can or cannot do, or what may or may not be explained by reference to the powers of matter, he must be understood as talking about the matter of Cartesian theory, or what might be called *matter*$_c$.

Descartes does not claim that matter$_c$ lacks all active qualities, or that matter$_c$ has no ultimate dispositions by reference to which its phenomenal properties may be explained. The predominant limiting feature of his theory is that such qualities or dispositions of matter may not be thought of as something other than *modes*, and that the relevant dispositions required for explanations in natural philosophy must be modes of a *material object*. Thus he advises Regius:

> However, we do not deny active qualities, but we deny that any degree of reality greater than that of modes should be attributed to them, since that cannot be done without conceiving of them as substances. Nor do we deny dispositions, but we understand them as being of two kinds. Some are purely material, and depend only on the configuration or other arrangement of the parts. Others, however, are immaterial or spiritual, such as the theologians' dispositions of faith, grace, etc., which do not depend on matter, but are spiritual modes that exist in a mind in the same way that motion or shape are corporeal modes that exist in a body. (iii. 503)

This still leaves open the possibility, as indicated above in the case of Boyle, that there are two distinct types of reality—one 'purely material' and the other 'immaterial or spiritual'—and that explanations in natural philosophy are irrelevant to the task of explaining mental operations such as thinking or remembering. Before considering that option, it may be helpful to clarify what Descartes meant by the term 'real', in the phrase 'real qualities and substantial forms', and to mention his independent reasons for rejecting real qualities.

REAL QUALITIES AND SUBSTANTIAL FORMS

Descartes's concept of explanation was motivated partly, as indicated above, by the models of explanation or intelligibility that he borrowed from everyday experience, and partly by widely shared objections to the vacuity of scholastic explanations. One feature of the latter that attracted independent objections was their appeal to so-called real qualities. As Stephen Menn has shown, Descartes was not claiming that qualities such as the colour or shape of something are unreal, but that they are not 'real' qualities. 'A quality can really belong to something, and be really a quality, without being a real quality' (Menn 1995: 184). The linguistic clue to understanding what was meant by 'real qualities' is

[15] Cf. McMullin (1978*b*).

the Latin word *res* (a thing). To ask whether a quality is real is to ask whether it has the status of a thing, and this in turn is equivalent to asking whether a quality could be thought of as existing apart from the reality of which it is predicated. Although this interpretation of real qualities may not have been widespread among scholastics, it was adopted at least by some of them, including Suarez (to whom Descartes often looked as a source of metaphysical concepts). Descartes, therefore, understood 'real qualities' as entities that had the same status as substances or things. It was for this reason, he explains in the Sixth Replies to objections, that he had rejected the scholastic account of heaviness. 'When, for example, I conceived of gravity as if it were a real quality of some kind which inhered in solid bodies, although I called it a "quality", in so far as I attributed it to the bodies in which it inhered, I also added, nevertheless, that it was "real" and thus I was thinking of it as a genuine substance' (vii. 441).[16] The same point is made two years later, in a letter to Mersenne (26 April 1643), where Descartes emphasizes the distinction between denying the reality of motion, which he is not remotely tempted to do, and denying that motion is a real quality of moving bodies.

> I do not assume that there are any *real qualities* in nature, which would be added to a substance, like little souls to their bodies, and which could be separated from it by God's power; and thus I do not attribute to motion, or to all the other variations of substance which are called qualities, any more reality than philosophers usually attribute to shape, which they do not call a real quality [*qualitatem realem*] but only a mode [*modum*]. (iii. 648–9)

The motion of a body is just as real as its shape, or as any of the other features that philosophers usually call a mode. But it should not be thought of by analogy with a soul that is added to bodies and would therefore be separable from matter, if only by divine power.

One of the objections to understanding qualities as real is that the concept of a real quality involves a contradiction in terms. 'It is completely contradictory to claim that there are real accidents, because whatever is real can exist apart from any other subject and whatever is capable of existing separately in this way is a substance rather than an accident' (vii. 434). The principal reason, however, for rejecting so-called real qualities was that they are both unintelligible and redundant in the explanation of natural phenomena. Returning to a claim already mentioned above and frequently repeated in the texts, Descartes writes to Mersenne:

[16] Cf. Descartes to Elizabeth (21 May 1643): 'various qualities . . . that we imagined as real, that is, as having an existence that is distinct from that of the body and, consequently, as being substances even though we called them mere qualities' (iii. 667).

> The principal reason that makes me reject these real qualities is that I do not see that the human mind has any notion in itself or any particular idea with which to conceive of them; thus in naming them and in claiming that they exist, one believes in something that is not conceived and that one does not understand oneself. The second reason is that philosophers have assumed these real qualities only because they believed that they could not otherwise explain all the phenomena of nature; for my part, I find on the contrary that one can explain natural phenomena much better without assuming them. (iii. 649)

The objection to so-called substantial forms is exactly the same, and for this reason the two categories are often linked together in a single phrase as the target of Descartes's objections: 'real qualities and substantial forms'. Accidental forms, however, do not raise comparable difficulties. Scholastic writers used the term 'accidental form' to describe a feature of something that could change without any fundamental change in the thing itself. In this sense, the colour of a door or the weight of a person is said to be an accidental form, and a new coat of paint on the door or a change in someone's weight will not convert the former into a non-door, nor turn the latter into a non-person or a different person. Descartes adverts to such changeable qualities in the discussion of the piece of wax in the Second Meditation. 'It loses what remains of its taste, its smell is lost, the colour changes, it loses its shape, increases in size, becomes a liquid, becomes hot and can barely be touched. Nor does it emit a sound if tapped. But does the same wax remain? It must be agreed that it does; no one denies that, no one thinks otherwise' (vii. 30; *M*. 27). A piece of wax can present itself to us in 'different modes' without ceasing to be wax, and the traditional Latin term for a non-essential quality, *forma*, was used by Descartes in this context without any ontological commitments that might subsequently embarrass him. If the need arises, he could distinguish the wax itself 'from its external forms [*formis*]' and could think about it 'as if it were bare and without its clothes on'. However, in contrast with such accidental forms, it is impossible to accommodate the language of substantial forms within Cartesian natural philosophy without making ontological commitments that would compromise the coherence of the conceptual framework used.

Descartes understands a substantial form as something that satisfies the conditions necessary for being a substance and therefore, even if joined with something else, would have the status of an independent thing and would be capable of existing without being predicated of something else. In a letter to Regius (January 1642) he makes this explicit: 'Lest there be any ambiguity in terms, it should be noted here that, when we reject substantial forms, the term should be understood as some kind of substance that is joined with matter . . . which is a genuine substance, or something that exists of itself . . .' (iii. 502). Thus

substantial forms suffered from a fundamental objection similar to that raised against real qualities: they involved a contradiction of being and not being independent entities in their own right.

Descartes may have other metaphysical reasons for rejecting the traditional scholastic concept of a substantial form. If so, the appropriate place to discuss them is in Chapter 8 below, which reviews his more general reservations about the concept of a substance. For present purposes, it is enough to acknowledge that the Cartesian model of explanation requires that one explain some features of a given phenomenon (the *explanandum*) by reference to other features that are better understood (the *explanans*). This would be impossible if he were to classify some characteristics of a given phenomenon as 'real qualities or substantial forms', because Descartes understood both of these expressions as implying that the features in question are distinct substances. The implications of this insight emerged, even more explicitly, when addressing the conceptual disparity between the mental and the physical.

The extent to which mental phenomena constitute a significant and possibly insurmountable challenge to the Cartesian model of explanation is best approached by degrees. The first step is to consider how Descartes proposed to explain living things. Then, among living things, conscious or thinking beings raise even more challenging difficulties. In the case of living things, Descartes's commitment to structural explanation holds firm. When confronted by conscious experiences, however, one has to ask whether he admits defeat and reverts to ontological dualism, or whether he merely acknowledges the limits, in the 1640s, of the success of his explanatory programme.

EXPLAINING A LIVING BODY

The style of explanation proposed by seventeenth-century corpuscularians was applied to living creatures with a degree of rigour that must have appeared to many contemporaries as unwarranted dogmatism. Descartes was no more unique in this respect than he was in proposing a new concept of explanation. Traditional explanations of living creatures were borrowed from commentaries on Aristotle's *De anima*. According to Aristotle, 'the soul is, so to speak, the first principle of living things' (1986: 402a). While all living things had souls, human beings were thought to have a complex soul that was divisible into three 'parts' or basic functions, the vegetative, the sensitive, and the intellective. Souls, in turn, were understood as special cases of the more general metaphysical category of forms. Therefore those who maintained that forms, substantial or

otherwise, and natures were non-explanatory were required to look elsewhere for an alternative account of living things.

Boyle was typical of those who made the direct inference from abandoning forms and qualities to forging a new account of living things. In *A Free Enquiry*, he addressed the plausibility of the traditional argument that the spontaneous response of the human body to 'fever, pleurisies, etc.' is a good reason to believe in the independent efficacy of 'nature'.

> In order to this [i.e. to address this objection], I desire it may be kept in mind that I do not only acknowledge, but teach, that the body of a man is an incomparable engine, which the most wise author of things has so skilfully framed for lasting very many years, that if there were in it an intelligible principle of self-preservation (as the naturists suppose there is) things would not in most cases be better or otherwise managed for the conservation of the animal's life than they generally are. So that the question is not, whether there is a great deal of providence and wisdom exercised in the crises of diseases, but upon what account it is that these apposite things are performed. (Boyle 1996: 92–3)

Many physicians wished to explain the phenomenon of spontaneous life-preserving reactions in our bodies—accepted as a reality by all concerned—by reference to an 'intelligent principle they call nature'. In contrast, Boyle preferred to invoke 'the wisdom and ordinary providence of God, exerting itself by the mechanism, partly of that great machine of the world, and partly of that smaller engine the human body' (Boyle 1996: 93). This might seem initially like a mere dispute about terminology, especially since Boyle had already conceded, earlier in the same text, that animals are 'furnished with faculties or powers and other requisites to enable them to preserve themselves and procure what is necessary for their own welfare' (Boyle 1996: 74). However, such faculties and powers had to be explained without understanding them as distinct theoretical entities, apart from 'what regularly, or what most usually, happens to beings of that species'. In other words, God's providence is the ultimate principle, for Boyle, by reference to which we can understand how bodies are self-preserving. But the detailed explanation of how God's design is implemented must be provided in terms of the chemical and physical mechanisms that are found in animals themselves and in the wider environment in which they survive.[17]

Despite his spirited defence of the 'engine' analogy, Boyle was also anxious to acknowledge its limitations. He was not therefore proposing that the human body is like a watch because it is much more flexible and organic than the term

[17] Boyle acknowledges the scruples that readers may have about explaining 'the bodies of animals, though not the rational souls of men' by mechanical laws, but he reminds them that an earlier generation of readers was equally reluctant about rejecting the intelligences that allegedly moved the 'celestial orbs' (Boyle 1996: 98–9).

'engine' might suggest: 'And here I desire to have it taken notice of . . . that I look not on a human body as on a watch or a hand mill—i.e. as a machine made up only of solid or at least consistent parts—but as a hydraulical, or rather hydraulo-pneumatical, engine, that consists not only of solid and stable parts, but of fluids and those in organical motion' (Boyle 1996: 127–8). As usual, Boyle adds a second qualification to the effect that the human body also has a rational mind to guide it.[18] Since this is absent in the case of other animals, the significance of the mechanical model is obvious, and its disanalogies with the *explanandum* are accepted. In summary, Boyle's machine analogy is a methodological proposal to the effect that we should try to construct explanations of living things without postulating souls in them, because such explanations are vacuous.

Cartesians were equally enamoured of animal machines for the same methodological reasons. The best way to understand their reliance on a machine analogy is from the perspective of a theory of explanation. From this perspective, the negative motivation for proposing that model is clearer than its positive content. Fundamentally, the analogy between living bodies and machines represented a rejection of animal souls as non-explanatory, rather than a limitation imposed a priori on the content of explanations in biology or embryology to what is analogous to clocks. Thus the project outlined in Descartes's early works, including the posthumously published *World*, included a programmatic statement of how to construct an explanation of the human body. As is well known, Descartes borrowed the theoretical model used in *A Treatise on Man* from machines that had been developed in the sixteenth and seventeenth centuries and had aroused expressions of wonder in observers. He refers, for example, to the hydraulically activated machines in the royal gardens at Saint-Germain-en-Laye: 'Similarly, you may have observed in the grottoes and fountains in the royal gardens that the force that drives the water from its source is all that is needed to move various machines, and even to make them play certain instruments or pronounce certain words, depending on the particular arrangements of the pipes through which the water is conducted' (xi. 130; *W.* 107). As visitors walk through the gardens they may unwittingly trigger a reaction in one or more of the statues, directly as a result of the sensitivity of the statues to objects in their environment. 'For they cannot enter without stepping on certain tiles which are arranged in such a way that, for example, if they approach a Diana bathing they will cause her to hide in the reeds, and if they

[18] Boyle also relied on the metaphor of a pilot in a ship, in the case of human bodies: 'a man is not like a watch or an empty boat . . . but like a manned boat, where, besides the machinal part (if I may so speak), there is an intelligent being that takes care of it, and both steers it or otherwise guides it . . .' (1996: 135). I discuss Descartes's use of this metaphor in Chapter 5.

move forward to pursue her they will cause a Neptune to advance and threaten them with his trident . . . and other such things depending on the whim of the engineers who constructed them' (xi. 131; *W*. 107). In *A Treatise on Man*, Descartes proposed that non-human animals be explained as complex biological machines, with the qualification that they are vastly more complex than human artefacts such as clocks, because they were created by God: 'as I am supposing that this machine is made by God, I think you will agree that it is capable of a greater variety of movements than I could possibly imagine in it, and that it exhibits a greater ingenuity than I could possibly ascribe to it' (xi. 120; *W*. 99). In making the machine of the body, God provides it with 'all the parts needed to make it walk, eat, breathe, and imitate all those functions we have which can be imagined to proceed from matter and to depend solely on the disposition of the organs' (xi. 120; *W*. 99).

One could read this as a radically misconceived proposal that disingenuously omits from its list of explananda precisely those features of genuine, living animals that are most difficult to explain, such as their sensations, memories, and so on.[19] That would amount to reducing the complexity of the explanandum to the conceptual limitations of one's explanatory model. But the way in which Descartes develops the suggestion implies something quite different. The project is motivated, not by a priori limits on the kinds of theoretical entities that it may include, but by a rejection of the scholastic style of explanation that it is designed to replace. The Cartesian approach to explaining living creatures depends on the same principle of ontological parsimony that excluded forms and qualities, and on the same ideal of explanation that implied their redundancy. In this context, the methodological proposal was to omit all souls in so far as they are neither necessary nor useful in constructing a genuine explanation, and to push to its limits the resourcefulness of the new model of structural explanation. The extent to which this project was successful can be decided only by subsequent historical developments both in scientific theory and in empirical methodology. Whatever the degree of success achieved at any particular time, it was a firm conviction of the Cartesian view that one should never revert to vacuous explanations in terms of souls.

Why then does Descartes do precisely this when he discusses the powers and faculties usually associated with the human mind? Hobbes articulates this challenge sharply, in his Objections to the *Meditations*: 'If Descartes were to show

[19] The assumed necessary link between 'life' and 'soul' has led some commentators to describe Cartesian animal-machines as not being alive. See Des Chene (2000: 12): 'Descartes's animal-machine, perhaps the most influential image of the living in the new science, has no life.' A similar claim is made in the introduction to Des Chene (2001: 2–3): 'The science of life is henceforward to be, not the science of a special part of nature consisting in those things that live, and that therefore have souls, but an extension of physics.'

that the agent who understands is identical with the understanding, we would return to the scholastic way of speaking: the understanding understands, vision sees, the will wills, and, according to the best analogy, walking—or at least the faculty of walking—walks' (vii. 177). Even if one prescinds from Hobbes's commitment to materialism, this point is well taken. What progress in explanation or understanding of the relevant phenomena is made by attributing to human beings various faculties, such as the understanding and the will, that are described precisely and exclusively as powers to perform the kind of actions that require an explanation?

It is appropriate to recall that one of Descartes's objections to the use of forms and qualities in natural philosophy was that they were introduced as soul-like or spiritual realities in a context where something physical or material was needed and that, once introduced, it seemed impossible to explain how they could interact with the kinds of physical properties for which they were intended as an explanation. Some commentators have assumed that this objection is irrelevant to constructing a theory of mind, on the assumption that the mind is spiritual.[20] That misses the point, however, about the alleged inconceivability of a link between the mental and the physical. Descartes was correctly concerned about the viability of any explanation that relies on theoretical entities for the interaction of which we lack any model or even an elementary understanding. However, the more fundamental objection to forms and qualities was that they were non-explanatory, and this objection remains valid even when explaining mental phenomena.[21] Descartes implicitly acknowledges this in the *Meditations*, in a context where he describes our thinking and our decision making as resulting from two distinct faculties or powers, the intellect and the will. But even here he was mindful of Occam's razor. Although human beings characteristically make mistakes, one could not explain this reality by reference to a 'mistake-making faculty [*facultas ad errandum*]' (vii. 54).

Thus there is a general, principled objection to invoking faculties or powers in order to explain anything, and there is a further specific objection that we do

[20] Paul Hoffman (1986: 350) argues that one of Descartes's principal reasons for rejecting substantial forms and real qualities is that such explanations are anthropomorphic: 'Obviously, this objection that explanations appealing to substantial forms are anthropomorphic does not apply to an explanation which takes the human soul to be a substantial form.' M. Rozemond (1998: 152) also argues that the charge of anthropomorphism 'obviously does not apply to the human soul'.

[21] Paul Hoffman (1986: 351) argues that the soul or mind is not a theoretical entity at all. 'The human soul, in contrast, is not a mere theoretical entity. That he exists and that he is a thinking thing are the first two propositions Descartes claims to know with certainty in the *Second Meditation*.' Evidently, the fact that Descartes exists and thinks does not imply that he is equally certain about having an immaterial mind. Unless the term 'mind' merely redescribes the reality it is intended to explain, it must be treated as a theoretical entity that is introduced to explain those features of our experience about which even Descartes expressed no doubts, such as the fact that we think.

not understand how such faculties, if conceived as real entities of some kind, could interact with well-known features of bodies, such as their size, shape, and so on. The question, then, is whether a theory of mind is a general exception to these fundamental objections, or whether the admission of a human mind, as a spiritual substantial form, exemplifies all the objections outlined earlier in this chapter. Descartes confronts this issue most explicitly in his correspondence with Princess Elizabeth in the 1640s.

EXPLAINING BODY–MIND INTERACTION

When Descartes published the first two editions of *Meditations* in 1641 and 1642 (which included the lengthy sets of objections and replies from critics), the titles included clear references to the immortality of the human soul or its distinction from the body, the implications of which were not lost on his readers. Among the first to challenge explicitly whether this new way of talking about the mind and the body could provide an explanation of our mental experiences was Princess Elizabeth of Bohemia. Her correspondence with Descartes provided confirmation of the problems that arise when one deviates from the criteria for what counts as a viable explanation and reverts to the discredited language of the schools. Elizabeth's first attempt to clarify the issue (16 May 1643) was expressed as follows:

> How can the human soul, which is only a thinking substance, determine the movement of the animal spirits in order to perform a voluntary action?[22] It seems as if every determination of movement results from the following three factors: the pushing of the thing that is moved, the manner in which it is pushed by the body that moves it, and the quality and shape of the latter's surface. The first two presuppose that the bodies touch, while the third presupposes extension. You exclude extension completely from your concept of the soul and, it seems to me, it is incompatible with being an immaterial thing. That is why I am asking for a more specific definition of the soul than what is provided in your *Metaphysics*, that is, of the substance of the soul when it is separated from its action of thinking. (iii. 661; *M*. 148)

Descartes's reply relies on a claim that we have a limited number of fundamental concepts by reference to which we can explain successfully phenomena of different categories. He suggests (21 May 1643) that there are two different features of the soul that require explanation: 'one is that it thinks and the other is that, since it is united with the body, it can act and be acted on in conjunction

[22] The term 'determine' is a technical term in Cartesian physics, which refers to modifying the direction and velocity of a moving body.

with the body' (iii. 664; *M*. 148). His excuse for failing to address the second question in the *Meditations* was that it was less relevant to his primary objective in that book—namely, to demonstrate the immortality of the soul. But, once his royal correspondent had raised the question of mind–body interaction, Descartes accepts the challenge. It is in this context that he offers a threefold classification of explanatory categories, or what he calls 'primitive notions':

> I think that there are certain primitive notions in us which are like originals, on the model of which we construct all our other knowledge. There are very few such notions. For apart from the most general notions of being, number, duration, etc. which apply to everything that we can conceive, we have only the notion of extension that is specifically for the body, and from that follow the notions of shape and movement; and for the soul on its own we have only the concept of thought, which includes perceptions of the understanding and inclinations of the will. Finally, for the soul and body together, we have only the concept of their union, on which depends the notion of the soul's power to move the body and the body's power to act on the soul by causing its sensations and passions. (iii. 665; *M*. 149)

It is not clear what Descartes meant by this claim.[23] One implication, made explicit by Descartes himself, is that the clusters of concepts that depend on these primitive notions do not overlap and that it is impossible to explain one primitive notion in terms of another. No argument is offered on this occasion to support that claim; the kind of argument assumed will be considered in Chapter 9. It follows that it would be a mistake to attempt to explain any phenomenon by applying an inappropriate set of concepts, and it would be equally misguided to attempt to explain one primitive notion in terms of another. 'For if we try to resolve a particular difficulty by using some notion that does not apply to it, we cannot fail to go wrong. The same thing happens if we try to explain one of these notions by reference to another; since they are primitive notions, each of them can be understood only through itself' (iii. 665–6; *M*. 149). When applied to the mind–body problem, this implies that we have different primitive notions available for thinking of the soul on its own, and for thinking of it as united and interacting with a body. Descartes illustrates his point by referring back to a reply to one of the Sixth Objections. We are pre-theoretically aware of the fact that the soul can move the body, and we also know from experience that heavy

[23] For an alternative reading, see Garber (2001: 168–88). Garber argues 'that Descartes is just not *entitled* to the answer he gives Elizabeth' (2001: 169). Since God is involved as the ultimate cause of the agency of occasional causes, the explanation of all motion in Cartesian physics presupposes the intelligibility of the interaction of God (a pure spirit) with matter. Thus mind–body causal connections, rather than the concept of extension, provide the primitive notion for understanding all motion, and Descartes therefore gave Elizabeth the wrong answer to her question by assuming that body–body causal relations are intelligible and that mind–body connections are problematic. I prefer to read Descartes as if he understood the conceptual foundations of his physics, and that we should reject Garber's interpretation rather than Descartes's.

bodies move towards the centre of the earth. If we were to think of the motion of heavy bodies as being caused by some kind of soul in each heavy body, that would amount to applying the wrong primitive notion. It would mean applying mental concepts to something that is extended and that should be explained, however it is done, by appeal to the concept of extension and to other concepts that are compatible with that basic concept.[24]

Elizabeth was not convinced by Descartes's initial response. She rephrased her original objection (20 June 1643) by confessing that 'it would be easier for me to attribute matter and extension to the soul than to attribute the ability to move a body, and to be moved by a body, to an immaterial being' (iii. 685; *M*. 151). Evidently, the mechanism by which two entities can interact, when they have as little in common as the mind and the body, remained as unintelligible to her as it had been earlier. Descartes tries once more to address the underlying issue in his reply. He concedes that one might wish to conceive of the soul as material, 'which is, strictly speaking, to conceive of its union with the body', but even in that case the problem remains once we acknowledge that it is separable from the body. The new factor in this letter (28 June 1643) is that we not only have three distinct primitive notions, but that they are acquired in different ways. Their distinct origins help to explain the fact that there is no overlap in the phenomena to which they may legitimately be applied.

> the soul is conceivable only by pure understanding; the body, that is extension, shapes and movements, may also be conceived by pure understanding on its own, but it can be conceived much better by the understanding assisted by the imagination; and finally, things that pertain to the union of the soul and body are known only obscurely by the understanding on its own or even by the understanding assisted by the imagination, but they are known very clearly by the senses. (iii. 691–2; *M*. 152)

Descartes draws the conclusion that people who never philosophize and who use only their senses 'have no doubt that the soul moves the body and the body acts on the soul'. However, they also think that body and soul are one and the same thing; 'in other words, they conceive of their union. For to conceive of the union of two things is to conceive of them as one thing' (iii. 692; *M*. 152). This might seem like a misguided concession by Descartes simply to appease an importunate royal correspondent who could not easily be ignored or dismissed. But Descartes goes on to explain that metaphysical thoughts are likely to mislead us when we conceive of the unity of the human being, and they may be responsible for obfuscating rather than clarifying 'the notion we have of the union of mind and body'. The reason is that we find it difficult to conceive 'very

[24] The conceptual confusion involved in the analogy between gravity and the soul's power to move the body is found in vii. 442, and Descartes to Elizabeth (21 May 1643) (iii. 667).

distinctly, and simultaneously, both the distinction and union of body and soul'. In other words, we find it difficult to conceive of mind and body as a single reality and at the same time to conceive of them as two distinct realities, 'which is self-contradictory' (iii. 693; *M*. 153).

It remains to be seen, in subsequent chapters, whether this correspondence with Elizabeth can be combined with Descartes's model of explanation to provide a coherent theory of the human mind. The most familiar interpretation of these letters is that they represent an honest recognition of the consequences of ontological dualism. However, they may also be read as a discussion of what happens when Cartesian science reaches the limits of its explanatory success and acknowledges that it cannot provide, in the 1640s, a unified theory of mental powers in terms of properties that are described as material (in the Cartesian sense of that term).

CONCLUSION

The interim conclusion from this chapter is that Descartes's project in natural philosophy is motivated primarily by his concept of explanation and, in particular, by his systematic critique of the widely established style of explanation that was prevalent in the schools. The principal claims made in this context were: (1) we can explain some phenomenon only by reference to something else that is already adequately understood; (2) it is not an explanation at all to postulate a specific 'form' or 'faculty' for every phenomenon that needs to be explained; (3) when constructing an explanation, we must not postulate more theoretical entities than are strictly necessary (Occam's razor); (4) there is no objection to postulating the existence of unobservable theoretical entities, on condition that they are described by analogy with well-known properties of macroscopic bodies; (5) the explanation of any phenomenon must be developed by relying on a set of concepts that are relevant to that type of phenomenon (though the implications of this vague rule remain to be made explicit); and (6) the explanation of all natural phenomena must ultimately come to a stop somewhere and there is no alternative, at that stage, but to appeal to fundamental dispositions of matter that are described in terms of the relevant *explananda*.

Neither Descartes nor his contemporaries can establish the first condition in an atemporal manner, as if what we can understand were completely independent of the current state of our best theories. However, this condition constitutes a standing challenge to anyone who wishes to introduce new theoretical entities: to explain what they mean by them and, especially, to explain how the adoption of such entities would contribute to an improved understanding of

some phenomenon. Thus, without being committed uncritically to the fruitfulness of corpuscular theories, Cartesian objections to real qualities and substantial forms remain valid. Such occult entities failed the most basic requirements of a good explanation because they were not adequately understood, were redundant, and were fundamentally non-explanatory. There is a real sense in which these objections represent a watershed in the history of ideas; there could be no going back to substantial forms, faculties, natures, or their equivalent in a genuine explanation of any phenomenon, including the human mind.

Descartes also accepted that any explanation must come to at least a temporary end at some point, at which one simply specifies those fundamental features of some reality that must be postulated in order to explain its manifest appearances. If this point is to avoid being premature or arbitrary, one needs some account of why one's explanatory progress happens to stop there rather than elsewhere. Descartes's correspondence with Princess Elizabeth presupposes, rather than explains, one of the points at which explanations of natural phenomena encounter a conceptual impasse. It remains to be seen whether he can provide a reason for this without relying on a metaphysics that is either discredited or unsupported by independent arguments.

One option available to Cartesian natural philosophy is to accept the limits of its chosen model of explanation by linking 'mechanical explanation' and 'material bodies' together, and then to face with an open mind the implications for what were generally classified as immaterial realities. Much would depend, at this juncture, on what Descartes wishes to claim about what is described as 'immaterial'. He could adopt Boyle's suggestion, in the *Christian Virtuoso* (Part 2), that 'immaterial' is a name not for another kind of reality that we know, but for something that (by definition or otherwise) we do *not* know or that we know very inadequately.

> For though superficial considerers take up with the vulgar definition, that *a spirit is an immaterial substance*, yet that leaves us exceedingly to seek, if we aim at satisfaction in particular inquiries. For it declares rather what the thing *is not*, than what *it is*; and is as little instructive a definition, as it would be to say, that *a curve line is not a strait one*, which sure will never teach us what is an ellipsis, a parabola, an hyperbola, a circle, or a spiral line, etc. (Boyle 1999–2000: xii. 474)[25]

This suggestion, if carried further (as it is by Hobbes), would result in an explicit denial that we know what we are talking about when we speak of immaterial

[25] Boyle illustrates his point by reversing the definition in favour of immaterial substances. If one knew only that 'a body is an unspiritual substance', one would know little about 'the distinct and particular natures of the sun, or a cloud, or of the stars, elements, minerals, plants, animals' (Boyle 1999–2000: xii. 475).

substances.[26] But one could also stop short of this conclusion and claim that, while we have some knowledge of material things and have a promising model for explaining them, our knowledge of what is immaterial is either too indirect or limited to provide an appropriate foothold for applying the same model of explanation. Kenelm Digby, who, like Descartes, set himself the task of establishing the immortality of the human soul, adopted this solution.

Digby published his *Two Treatises* in the same year as Descartes's *Principia* (that is, in 1644). In the *Treatise of Mans Soule*, Digby argues: 'Here we have passed the Rubicon of experimentall knowledge: we are now out of the boundes that experience hath any iurisdiction over: and from henceforth, we must in all our searches and conclusions rely only upon the single evidence of Reason' (1644: 415–16). Digby argues that everything material, including the motions and sensations of animals, falls in principle within the scope of mechanical explanation.[27] When one reaches the limits of explanation, one could invoke forms as place-holders for various features of a material body that one hopes to specify subsequently. Scholastic philosophers, however, misunderstood the function of forms and thought of them as entities that are distinct from the realities to be explained and comparable to immaterial souls.

> But later Philosophers, being very disputative, and desiring to seeme ignorant of nothing (or rather, to seeme to know more than any that are gone before them and to refine their conceptions) have taken the notions, which by our first Masters were sett for common and confused explications of the natures, (to serve for conveniency and succinctenesse of discourse) to be truly and really particular Entities, of thinges of themselves . . . for in truth, they turne all bodies into spirits, making (for example) heate, or cold, to be of it selfe indivisible, a thing by it selfe, whose nature is not conceivable; not the disposition or proportion of the partes of that body which is said to be hoat or cold, but a reall thing, that hath a proper Being and nature peculiar to it selfe; whereof they can render you no account. (Digby 1644: 351–2)

The damaging consequence of this strategy was that they collapsed the distinction between bodies and spirits by conflating the distinction between our ways of knowing them. For Digby, the introduction of faculties and powers is appropriate once we go beyond material things and attempt to know something about what is non-material.

> For wee having very slender knowledge of spirituall substances, can reach no further into their nature, then to know that they have certaine powers, or qualities; but can

[26] Hobbes (1991: 34): 'And therefore if a man should talk to me of . . . *Immateriall Substances* . . . I should not say he were in an Errour; but that his words were without meaning; that is to say. Absurd.'

[27] Digby (1644: 366): 'all the actions of sensible bodies may be reduced to locall motion, and to materiall action of one body unto another, in a like manner (though in a different degree) as the motion which we see in liveliest bodies.'

seldome penetrate so deepe, as to descend to the particulars of such Qualities, or Powers. Now our moderne Philosophers have introduced such a course of learning into the schooles, that unto all questions concerning the proper natures of bodies, and their operations, it is held sufficient to answere, they have a quality, or a power to doe such a thing. And afterward they dispute whether this Quality or Power, be an Entity distinct from its subiect, or no; and how it is seperable, or unseperable from it, and the like. (Digby 1644: preface)

Digby's concern is that if some feature of material bodies is explained by reference to a 'qualitie occult, specificall, or incomprehensible', and if these occult qualities turn out to be nothing more than features of material bodies that were misunderstood or poorly named, the mental properties from which he concludes the immateriality of the soul may be likewise explained by 'a corporeal occult quality'. In a word, the viability of any argument for the immateriality of the soul presupposes agreement about a significant distinction between (1) our knowledge of material things and their properties, and (2) our ignorance about what is classified as immaterial. It follows that the strategy that is inexcusable in the case of material bodies, including animals—namely, the introduction of occult powers—may be appropriate in the case of the immaterial, because of the indirectness of our knowledge of the latter and the lack of empirical information to support our reasoning.

Of course this move merely defers the problem of explaining the 'immaterial' unless it is supported by supplementary reasons. It suggests that we construct genuine explanations of what is material and that we offer only provisional explanations of what is immaterial.[28] There are two ways in which Descartes could justifiably adopt this option. One is to accept that all our explanatory efforts must eventually reach some limit. At that point, we cannot further explain the realities postulated to account for the explanandum with which we began, and we may appeal to explanatory entities that are described purely dispositionally in terms of their effects. Given the limited development of all the relevant sciences in the 1640s, it would have been reasonable for a natural philosopher to concede that, at that time, the human mind was beyond the 'Rubicon' of which Digby wrote. But even then we would not be reduced to complete silence; we could argue that human beings have faculties or powers that are identified merely as the kind of faculties that give rise to our mental experiences. A second option is to think of mental phenomena as not being completely natural events because they are not explained adequately by causal

[28] Cf. MacIntosh (1983: 333), on the tension caused for those who are both corpuscularian natural philosophers and Christians: 'This makes it difficult for them in dealing with humans, since they want to explain all the animal functions mechanically, while offering non-mechanical (non)-explanations for the rest.'

interactions between material bodies. Such events could be described as non-material and therefore not subject to structural explanations because, like human language, they involve *conventional* signs, the meaning of which cannot be analysed merely in terms of causal relations. I discuss the somewhat vaguely perceived significance of this option for Descartes's theory of mind in Chapter 6. Both options would allow Descartes to think of the scope of the material as being co-extensive with those natural phenomena for which we can, at least in principle, provide a scientific explanation, and to classify whatever lies outside the scope of the latter as 'immaterial'. It may be that mental phenomena are immaterial in both senses. This interim conclusion also leaves open the question whether the mind is explained, in any sense of the term, by reference to substance dualism.

INDEX