

they have performed the wrong action, but because they have violated the *trust* of others. This trust is grounded in the expectation that agents have a fundamental disposition to accord deontic constraints a certain weight against desiderative constraints in their practical deliberations. It is this disposition that, at root, enables them to adopt rules to govern their conduct, which in turn allow them to avoid suboptimal interaction patterns. When the distinction between deontic, doxastic, and desiderative constraints is sharply drawn, the way that these rules are integrated into the agent's deliberations, along with the role of internal and external control in social integration, can be clearly specified.

VII Conclusion

An enormous amount of effort has been expended in recent years in an attempt to show that instrumentally rational agents can adopt commitments, special choice dispositions, or some other mechanism that will allow them to work their way out of suboptimal outcomes. The underlying motivation has been to explain how it is that we are able to avoid the extremes of uncooperative behavior that simple instrumental theories of rationality consistently predict. Over the years, the arguments advanced have become increasingly baroque.²⁷ The model presented here is comparatively simple. It shows how agents are able to cooperate with one another, even when more desirable outcomes can be achieved through defection. By integrating preferences for actions into the agent's overall utility function, without thereby obscuring the distinction between preferences for actions and preferences for outcomes, the model is able to provide an extremely straightforward representation of the way that norms might function as deontic constraints in social choice. The fact that the model is also able to supply deliberative micro-foundations for a 'focal point'-style solution to coordination problems is an unexpected bonus. It functions as indirect support for the proposed analysis, however, insofar as it fulfills the traditional expectation, among sociological theorists, that a solution to the 'problem of order' will require an account of both cooperation and coordination.

Received: December, 2000

Revised: June, 2001

²⁷ Compare David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press 1986) with David Gauthier, 'Assure and Threaten,' *Ethics* 104 (1994) 690-721.

*Unmasking Descartes's Case for the Bête Machine Doctrine*¹

LEX NEWMAN

The University of Utah

Salt Lake City, UT 84112-9156

USA

Among the more notorious of Cartesian doctrines is the *bête machine* doctrine — the view that brute animals lack not only reason, but any form of consciousness (having no mind or soul). Recent English commentaries have served to obscure, rather than to clarify, the historical Descartes's views. Standard interpretations have it that insofar as Descartes intends to establish the *bête machine* doctrine his arguments are palpably flawed. One camp of interpreters thus disputes that he even holds the doctrine. As I shall attempt to show, not only does Descartes affirm the doctrine, his supporting arguments are not palpably flawed — even if they ultimately come up short. It will indeed emerge that, in making his case, Descartes employs interesting argumentative strategies that have not been duly appreciated.

On my reconstruction, Descartes's account centers around the following probabilistic argument:

- (1) Experiments with mechanical physiology establish that purely mechanical causes are sufficient to explain all behavior of brutes, save that which is apparently reasoned.

¹ I am grateful to Robert Audi, Stephen M. Downes, Paul Hoffman, Nicholas Jolley, Alan Nelson, Ram Neta, Marleen Rozemond, and anonymous referees for the *Canadian Journal of Philosophy* for feedback on the ideas in this paper. I have also benefited from discussions with audiences in philosophy colloquia at Kansas State University and the University of California, Irvine.

- (2) Experiments with the behavior of humans and brutes establish that the apparently reasoned behavior of brutes is best explained as deriving from purely mechanical causes, not a faculty of reason.
- (3) Entities that are explanatorily superfluous (e.g. souls and other substantial forms) are to be eliminated.
- (4) The souls of brutes (and thus all consciousness in brutes) are to be eliminated. [(1),(2),(3)]

The present paper is organized around this argument.

In §I, I discuss the analogical character of the argument and its importance in light of Descartes's targeted audience. As conveyed in the formulation above, Descartes's case for the *bête machine* doctrine is broadly based in an analogy between humans and brutes. This understanding of the project challenges conventional wisdom according to which Descartes is inattentive to this analogy. This standard objection typically rests on the assumption that, had he noticed the strength of the analogy, he'd also have noticed that his Draconian, eliminativist arguments apply with equal force to brutes *and* humans. This line of objection traces back at least as early as the Fifth Replies where Gassendi complains that, though Descartes appeals to 'blind impulse' to explain the behavior of brutes, he appeals to 'some quite different' explanation of analogous forms of human behavior (CSM 2:188, AT 7:269-70).²

In §II, I address premise (1), a step that Descartes aims largely at scholastic readers. The tendency in the literature has been to ignore this step, focusing exclusive attention on the famous 'two tests' (discussed below) which purport to deny brutes a rational soul. At its worst, this tendency leaves Descartes with no argument against scholastics, an audience that already denies brutes a rational soul while maintaining that they have sensitive souls which provide sensory awareness.

In §III, I turn attention to premise (2) and Descartes's famous two tests of rationality. The locus classicus for his treatment of the *bête machine* doctrine is part 5 of his *Discourse on the Method*. He there proposes 'two very certain means of recognizing' that some creatures lack reason and are thus 'not real men' (CSM 1:139-40, AT 6:56). On standard interpreta-

tions the two tests are intended to reveal *whether or not* an organism has a faculty of reason.³ This understanding of the tests commits Descartes to the view that some forms of human behavior cannot be imitated by machines. So construed, Descartes is open to the criticism that he severely underestimates the imitative abilities of machines,⁴ a bad miscalculation even given his 17th-century understanding of mechanist principles. I shall challenge this standard account, arguing that the two tests of *Discourse 5* are intended only as negative tests — i.e. tests designed to reveal that a behavior is *not* reason driven (being instead an automatism), but not that it *is* reason driven. The Descartes emerging from my account has a far more sophisticated view of the imitative abilities of machines than on standard interpretations.

In §IV, I take up issues surrounding premise (3) and its importance to Descartes's broader case. In thus attributing to Descartes an appeal to the principle of parsimony, my interpretation again challenges conventional wisdom. On standard accounts Descartes does not invoke parsimony; it is indeed widely agreed that there's an intolerable logical gap between the premise set he intends to invoke, and the conclusion in (4). This widely held interpretive thesis has engendered two interpretive camps. One camp has Descartes affirming the *bête machine* doctrine, thus finding him guilty of the egregious non sequitur.⁵ Another camp exculpates him, by denying that he endorses the *bête machine* doctrine. On this latter interpretation, Descartes holds that the behavior of brutes is automatic, but not that brutes are mere automata (allowing them sensory awareness).⁶ As I'll argue, numerous unpublished texts (including early

3 Cf. J. Bennett, 'Thoughtful Brutes,' Presidential Address, American Philosophical Association (1988), 199; N. Chomsky, *Cartesian Linguistics* (New York: Harper and Row 1966), 78n.9; J. Cottingham, 'Cartesian Dualism: Theology, Metaphysics, and Science,' in J. Cottingham, ed., *Cambridge Companion to Descartes* (Cambridge: Cambridge University Press 1992), 250; D. Radner and M. Radner, *Animal Consciousness* (Buffalo: Prometheus Books 1989), 41; M. Rozemond, 'The Role of the Intellect in Descartes's Case for the Incorporeality of the Mind,' in Stephen Voss, ed., *Essays on the Philosophy and Science of René Descartes* (Oxford: Oxford University Press 1993), 99; and M.D. Wilson, 'Animal Ideas,' Presidential Address, American Philosophical Association (1995), 12.

4 Cf. Cottingham, 'Cartesian Dualism,' 250.

5 Cf. L.E. Loeb, *From Descartes to Hume* (Ithaca: Cornell University Press 1981), 112f.; D. Radner and M. Radner, *Animal Consciousness*, 79; and Z. Vendler, *Res Cogitans* (Ithaca: Cornell University Press 1972), 152ff.

6 Cf. J. Cottingham, 'A Brute to the Brutes? Descartes's Treatment of Animals,' in J. Cottingham, ed., *Descartes* (Oxford: Oxford University Press 1998), 230; and S. Gaukroger, *Descartes* (Oxford: Clarendon 1995), 289.

2 'AT' = C. Adam and P. Tannery, eds., *Oeuvres de Descartes* (Paris: J. Vrin 1904); 'CSM' = J. Cottingham, R. Stoothoff, and D. Murdoch, eds., *The Philosophical Writings of Descartes* (Cambridge: Cambridge University Press 1984); 'CSMK' = Volume III of CSM for which Anthony Kenny is a contributing translator. References to both AT and CSM are to the volume and page.

work on which premise (1) draws) establish that Descartes intends to eliminate sensitive souls in brutes (along with other scholastic substantial forms) by appeal to parsimony — thus avoiding the alleged non sequitur. That so many commentators have missed this is likely owed to the aforementioned neglect of Descartes's anti-scholastic program.

Though my central aim is to reconstruct (not to defend) Descartes's argument, a charitable reconstruction inevitably aims at plausibility. The strength of Descartes's position is best appreciated in the precise metaphysical and epistemological context in which the argument of *Discourse 5* is embedded. I shall thus assume his central metaphysical doctrines as sketched out in *Discourse 4* and more carefully in the *Meditations*.⁷ Accordingly, there are two fundamental, really distinct kinds of being: *mental* (i.e. thinking, or conscious) and *material* (i.e. mechanical). In that human persons consist of both kinds, our anatomical behavior may result from either of two, really distinct kinds of causes: willings, or local motion. Though Descartes purports to establish the metaphysical doctrines of *Discourse 4* with metaphysical certainty (cf. AT 6:37f),⁸ his inquiry assumes a less ambitious grade of certainty early in *Discourse 5* (cf. AT 6:45f) upon turning attention to the case for the *bête machine* doctrine.

I Correcting Misinformed Appeals To Analogy

Descartes's rejection of souls in brutes is the result not of inattention to the close analogy between their behavior and ours, but of what he takes as the strength of the analogy. In his view, analogical arguments purporting to establish that brutes have souls are typically grounded in mistaken assumptions about the causes of human behavior. Some such assumption informs the two kinds of views he looks to target — commonsensical views, and the doctrines of the Schools.

According to Descartes, a pervasive and commonsensical assumption has it that the fundamental mover shaping human behavior is the soul:

We have all found by experience that many bodily movements occur in obedience to the will, which is one of the faculties of the soul, and this has led us to believe that the soul is the principle responsible for all bodily movement. (CSM 1:314, AT 11:224; *Description of the Human Body*)⁹

On this basis, commonsensical judgments are formed — by analogy — as to the causes of behavior in brutes:

Most of the actions of animals resemble ours, and throughout our lives this has given us many occasions to judge that they act by an interior principle like the one within ourselves, that is to say [according to the common, rash opinion], by means of a soul which has feelings and passions like ours. (CSMK 99, AT 2:39; 1638 letter to Renier for Pollot; cf. AT 5:276)

In his widely read *Apology for Raymond Sebond*, Montaigne popularizes arguments of this sort. Putting heavy emphasis on analogy, he urges that 'we must infer from like results like faculties.'¹⁰ We see even 'in our cruder works, the faculties that we use, and that our soul applies itself with all its power; why do we not think the same thing of them [brutes]?' (ibid., 333) 'Why do we imagine in them this compulsion of nature, we who feel no similar effect?' (ibid., 337) Though Montaigne's defense of rationality in brutes is subtle and philosophical,¹¹ I shall nonetheless refer to his position as 'commonsensical.' The appellation is fitting, because he so aptly articulates that feature of commonsensical accounts that Descartes means to rebut. The principal problem lies not so much with Montaigne's *like effects, like causes* principle as with the underlying assumption as to the vast influence of the human soul.

For Descartes, scholastic accounts also attribute too much of human behavior to the regulation of the soul. The hylomorphic doctrines of Aristotelianism stem in part from a misguided anthropomorphization of the world: all manner of natural phenomena is explained by appeal to appetites, intentions, and various powers modeled after the workings of the human soul.

7 The general disfavor into which these doctrines have fallen render this an enormously generous assumption. Bear in mind that my aim is to clarify why Descartes would think — in the light of his philosophical system — that the *bête machine* doctrine provides the best account of observed behavior in brute animals.

8 In *Discourse 4*, Descartes contrasts *metaphysical* certainty with *moral* certainty. Morally certain matters are such 'that it seems we cannot doubt them without being extravagant' (CSM 1:130, AT 6:37-8; cf. *Principles* 4:205-6).

9 Descartes adds: 'Our ignorance of anatomy and mechanics has also played a major role here. For in restricting our consideration to the outside of the human body, we have never imagined that it has within it enough organs or mechanisms to move of its own accord in all the different ways which we observe' (ibid.).

10 M. Montaigne, *The Complete Essays of Montaigne*, D. Frame, trans. (Stanford, CA: Stanford University Press 1957), 336

11 As is that of other *nouveaux Pyrrhoniens* — cf. the defense of Pierre Charron (a disciple of Montaigne), in his *Of Wisdom*.

Descartes thus thinks his scholastic and commonsensical readers suffer a common malady. The false assumptions of each group motivate an excessive appeal to souls for the explanation of human behavior. In turn, these mistaken assumptions corrupt their efforts to understand the behavior of brutes by appeal to analogy.

A related (and mistaken) assumption further corrupts efforts to reason, by analogy, from effects to causes. The assumption has it that each kind of behavioral effect results from a unique kind of cause — an assumption at work in Montaigne's appeals to the *like effects, like causes* principle. Counterexamples arise in connection with the imitative abilities of machines. Descartes holds that in humans a single kind of behavioral effect (where *kinds* are individuated in terms of the productive result — the overt behavior — rather than the productive process) might be triggered by either kind of causal principle: local motion stemming from pure mechanics, which for Descartes involves none other than blind pushes/collisions; or local motion stemming from the will, a stem which provides a distinctive purposive element.¹² Though, as Descartes conceives them, purely mechanical systems are incapable of the purposive operations afforded by a faculty of reason, he maintains the thesis that such systems are able to *imitate* or *exhibit* the behavioral effects of such operations. Call this the Imitation Thesis. Assuming the Imitation Thesis, Montaigne is wrong that a given kind of effect is always matched by a single kind of cause. The behavior of the final piece in a chain of dominoes may be indistinguishable whether the initial motion of the chain is triggered fully mechanically by levers, or intentionally by an agent's will (e.g., via the mediation of her limbs).¹³ The Imitation Thesis underscores the difficulties involved in arguing from effects to causes. The situation is strongly parallel to that which Descartes confronts in his treatments of skepticism. Since (as Descartes holds) the same kind of effect — viz. an apparently normal and waking experience — may be

caused by dreams or by waking, there are grave problems in establishing the true cause solely on the basis of effects.

Difficulties remain. Behavioral similarities at play in the Imitation Thesis would appear to undermine not only the Montaigne-style appeal to analogy, but also Descartes's. Since we sometimes produce by reasoned volition the same kinds of behavioral exhibitions that on other occasions are produced in us by pure mechanics, it would seem hopeless to try to establish that analogous behavioral exhibitions in brutes are *always* the result of mechanics alone. As will emerge, Descartes is aware of the difficulties and thinks he overcomes them.

II Mechanical Causes Are Sufficient To Explain Non-reasoned Behavior

Recall premise (1):

- (1) Experiments with mechanical physiology establish that purely mechanical causes are sufficient to explain all behavior of brutes, save that which is apparently reasoned.

Descartes's case for (1) is aimed more at a scholastic audience than a commonsensical one: Scholastics already accept the central point of (2), namely that brutes lack reason. As I contend, commentators have tended to neglect Descartes's treatment of (1), focusing disproportionate attention on his famous 'two tests' (to be discussed in §III). I want now to focus on Descartes's case for (1), giving special attention to his anti-scholastic moves. This attention anticipates the discussion to come in §IV, where I'll address his invocation of the principle of parsimony.

The shift from the natural philosophy of scholasticism to the new mechanical philosophy involves an elimination of myriad entities of hylomorphic explanation. As Descartes observes (in *Principles* 4:201), explanations in terms of mechanics are 'much better than explaining matters by inventing all sorts of strange objects' including (as the French edition adds) "'prime matter,'" "substantial forms" and the whole range of qualities that people habitually introduce, all of which are harder to understand than the things they are supposed to explain' (CSM 1:287, AT 8a:324-5).¹⁴

12 To More, Descartes writes: 'There are two different principles causing our movements. The first is purely mechanical and corporeal, and depends solely on the force of the spirits and the structure of our organs.... The other, an incorporeal principle, is the mind or that soul which I have defined as a thinking substance' (CSMK 365, AT 5:276).

13 In the Fourth Replies, Descartes writes: 'both in our bodies and those of the brutes, no movements can occur without the presence of the all the organs or instruments which would enable the same movements to be produced in a machine. So even in our own case the mind does not directly move the external limbs, but simply controls the animal spirits which flow from the heart via the brain into the muscles, and set up certain motions in them; for the spirits are by their nature adapted with equal facility to a great variety of actions' (CSM 2:161, AT 7:229).

14 Cf. the *World*, AT 11:7 and 11:25-6, for more on how Descartes understands scholastic hylomorphic explanation.

Descartes distinguishes two sorts of automatistic human behavior, both of which he explains by pure mechanics. The one sort is explained entirely by the mechanical disposition of our physiological constitution, as in Descartes's explanation of 'heartbeat, digestion, nutrition, [and] respiration when we are asleep' (CSM 2:161, AT 7:229f; Fourth Replies).¹⁵ The other sort is explained by the mechanical disposition of our physiological constitution in tandem with external, corporeal stimuli, as when the hands instinctively protect the head from external injury; in such cases, the sense organs transmit a mechanical signal that 'reaches the brain and sends the animals spirits [back] into the nerves in the manner necessary to produce this movement even without any mental volition, just as it would be produced in a machine' (ibid.). For Descartes, an approved characterization of either sort of automatistic behavior would have it that something is happening to the body, though such happenings are not under the occurrent direction of an adjoined soul.

Scholastic theories generally agree with Descartes in characterizing both kinds of behavior as automatistic. Unlike Descartes, they hold that the automatic processes involved are under the direction of the soul, namely the powers of vegetative and sensitive souls. Heartbeat and digestion require the guidance of vegetative souls. And such behavior as involves the pursuit of, or flight from, the objects of sense — as when sheep flee wolves — requires the powers of sensitive souls.¹⁶ The theory entails that sensory awareness of the objects of pursuit or flight is as indispensable to the production of such behavior as are the physiological contributions, themselves under the direction of the soul though they involve no cognition (cf. *ST* 1a.78.1).

In making his case for (1), part of the difficulty facing Descartes is that an appeal to analogy would seem to favor the scholastic theory. Introspective reflection on, for instance, our own reaction to wolves, strongly suggests (even if misleadingly) that conscious awareness of the wolf is an integral part of the correct explanation of our behavioral response.

In *Discourse 5*, Descartes offers no systematic case for (1). Instead, he summarizes relevant highlights from an earlier, unpublished work — the *Treatise on Man* — drawing out its analogical implications (cf. AT 6:45-6). The earlier work incorporates a thought-experiment involving a hypothetical man-machine. Initially, the man is created with only a

mechanical body, but no soul: 'I suppose [its] body to be nothing but a statue or machine made of earth, which God forms with the explicit intention of making it as much as possible like us' (CSM 1:99, AT 11:120). Though the man-machine lacks all powers of a soul, God nonetheless 'places inside it all the parts required to make it walk, eat, breathe, and indeed to imitate all those of our functions which can be imagined to proceed from matter and to depend solely on the disposition of our organs' (CSM 1:99, AT 11:120). Descartes's aim is, in part, to convey the plausibility of a fully mechanical explanation of the workings and behavior of the human body:

I should like you to consider that these functions follow from the mere arrangement of the machine's organs every bit as naturally as the movements of a clock or other automaton follow from the arrangement of its counter-weights and wheels. In order to explain these functions, then, it is not necessary to conceive of this machine as having any vegetative or sensitive soul or other principle of movement and life, apart from [the motion of its physiological parts]. (CSM 1:108, AT 11:202)

Descartes thinks his account in the *Treatise on Man* provides analogical evidence in support of (1). Assuming the plausibility of a fully-mechanical explanation of the man-machine's behavior, a similar explanation of the behavior of brute animals is also plausible. In *Discourse 5*, Descartes writes (referring back to the man-machine of his earlier treatise):

And when I looked to see what functions would occur in such a body I found precisely those which may occur in us without our thinking of them, and hence without any contribution from our soul (that is, from that part of us, distinct from the body whose nature, as I have said previously, is simply to think). (CSM 1:134, AT 6:46)

As the text continues, Descartes draws out the intended analogical implication. The functions occurring in the man-machine

are all those in which it may be said that brutes resemble us without [the benefit of a faculty of] reason: without which [faculty of] reason, I could find none of those functions which, being dependent on thought, are the only ones which belong to us as [real] men; whereas I found all these later, having supposed that God created a rational soul, and that he joined it to this body in the particular manner which I described. (AT 6:46; my translation)

Two aspects of the implied analogical reasoning deserve comment. First, Descartes leaves unclarified the kinds of behavior of 'which it may be said that brutes resemble us without reason' and are thus subject to a fully mechanical explanation. This is not surprising given the diversity of his intended audience. For the scholastic reader, the relevant kinds include *all* behavior of brutes. For the commonsensical reader, a significant

15 Cf. the *Treatise on Man*, AT 11:202, and *Passions* 16 (AT 11:341f). Though such behavior is typically automatistic, it sometimes issues from the soul; Descartes cites willful changes in respiration (AT 11:197) and pupil adjustment (AT 11:361f).

16 Cf. Thomas Aquinas, *Summa Theologica* (abbreviated *ST*) (New York: Blackfriars and McGraw-Hill 1964-), 1a.80.1, 1a.81.2-3.

proportion of the behavior of brutes is thought to be under the direction of the organism's own reason. In any case, to whatever extent the reader supposes that a faculty of reason is at play, the later argument of *Discourse 5* (that expressed in [2]) is intended to show otherwise. Second, the analogical implication should not be overestimated. Descartes says, of the human bodily movements on which he builds the analogy, that they 'may' (*peuvent*) — but not that they *always do* — 'occur in us without our thinking about them.' Consequently, the intended conclusion (at this early stage of the argument, at any rate) is no stronger than that the analogous behavior of brutes *may* occur in them without their thinking about them. As (1) conveys, purely mechanical causes are *sufficient* to explain a wide range of the behavior of brutes; it does not follow (not without additional assumptions), however, that such causes provide the *correct* explanation. Nonetheless, if Descartes can establish (1), it follows that consciousness in brutes is needed to explain their behavior only if the behavior is apparently reasoned. To explain unreasoned behavior, appeals to *sensory* consciousness are unneeded. More generally, scholastic vegetative and sensitive souls are explanatorily superfluous.

Notwithstanding its importance to Descartes's larger aims, the paltry extent to which a case for (1) is made in *Discourse 5* (quite unlike the detailed treatment of the man-machine in the *Treatise on Man*)¹⁷ amounts to little more than a promissory note as to the explanatory completeness of mechanical physiology. The only physiological mechanical workings he writes about relate to the heart and circulation, leaving it to the reader to extrapolate the remaining details (cf. AT 6:46-7). This would seem to be a critical mistake. For a scholastic audience, there is no easy extrapolation from the mechanical explanation of circulation to a similar expla-

nation of behavior for which sensory awareness is regarded as essential. In further support of (1), Descartes does also note that the purely mechanical account

will not seem at all strange to those who know how many kinds of automaton, or moving machines, the skill of man can construct with the use of very few parts, in comparison with the great multitude of bones, muscles, nerves, arteries, veins and all the other parts that are in the body of any animal. For they will regard this body as a machine which, having been made by the hands of God, is incomparably better ordered than any machine that can be devised by man, and contains in itself movements more wonderful than those in any such machine. (CSM 1:139, AT 6:55-6)¹⁸

But the persuasiveness of this appeal would be lost on those not already versed in mechanical explanations — those for whom the comparison of muscles, nerves, and arteries, to levers, pulleys, and wheels, would be strange.

Today's reader is apt to miss the significance of the fact that, in his *Discourse* treatment, Descartes mostly asserts (1) rather than arguing for it. From our vantage point, the explanatory power of mechanical physiology (an updated version, of course) is more or less taken for granted. But for Descartes's commonsensical and scholastic readers, this is a grave omission and helps explain some of the objections he was to encounter — objections to the effect that the powers of a sensitive soul (or sensory awareness, at any rate) *are* needed to explain the behavior of brutes. Among these is the following criticism represented by Arnauld, in the Fourth Objections:

I fear that this view [the *bête machine* doctrine] will not succeed in finding acceptance in people's minds unless it is supported by very solid arguments. For at first sight it seems incredible that it can come about, without the assistance of any soul, that the light reflected from the body of a wolf onto the eyes of a sheep should move the minute fibres of the optic nerves, and that on reaching the brain the motion should spread the animal spirits throughout the nerves in the manner necessary to precipitate the sheep's flight. (CSM 2:144, AT 7:205)

The sixth objectors add:

So far are we from accepting that all [the behavior of brutes] can be satisfactorily explained by means of mechanics, without invoking any sensation, life or soul, that we are willing to wager anything you like that this is an impossible and ridiculous claim. (CSM 2:279, AT 7:414)

17 In his *Treatise on Man*, Descartes shows special interest in explaining seemingly purposive external behavior. For instance, he writes that 'the effect of [corporeal] memory here that seems to me the most worthy to be considered consists in this, that without there being any soul in this machine, it can naturally be disposed to imitate all the movements of real men, or other similar machines' (AT 11:185). The account then unfolds (AT 11:185-197) with Descartes explaining how mechanical physiology *alone* can account for a variety of behavior appropriate to various stimuli, via a wholly reflex mechanism, including: complex eye adjustments, and the corresponding corporeal signals sent to the brain; behavior in which the movements of the limbs, the head, and the eyes, are coordinated (e.g. when the hand is burned, the reflexive response includes tears, a wrinkling of the face, and a disposition of the voice to cry (AT 11:192f)); behavior that 'serves to pursue desirable things, or to avoid the harmful' (AT 11:193); behavior that in real men serves only 'to manifest the passions' (*à témoigner les passions*) (AT 11:194), including laughing and crying; leg movements requisite to walking (AT 11:196f); and much more.

18 Cf. the *Treatise on Man*, AT 11:131f, and the thought-experiment in the letter of April or May 1638 (to Renier for Pollot).

III Mechanical Causes Provide the Best Explanation of Apparently Reasoned Behavior

Assuming premise (1), all but the apparently reasoned behavior of brutes can be explained by pure mechanics. In the 10th and 11th paragraphs of *Discourse 5* (and other parallel writings), Descartes makes his case for (2):

- (2) Experiments with the behavior of humans and brutes establish that the apparently reasoned behavior of brutes is best explained as deriving from purely mechanical causes, not a faculty of reason.

Again, Descartes's project is best understood as analogical. The primary aim of the present Section is to understand why he holds that *all* apparently reasoned behavior of brutes is best understood as analogous to purely mechanical human behavior — why we should judge that *none* of it stems from reason.

The remainder of §3 is organized as follows. In §III.1, I lay the groundwork for my claim that the two tests are negative tests. In §III.2, I clarify what Descartes means by 'reasoned behavior'. In §§III.3 through III.5, I explain and defend my interpretation of the two tests as special cases of a negative test that I call an *unmasking* test. In §III.6, I offer brief remarks about Descartes's view of language as the possible basis of a positive test.

1. Descartes on what machines cannot do

In the tenth paragraph of *Discourse 5*, Descartes extends his discussion of the hypothetical man-machine. Generalizing to all machines, he explains two tests for exposing machines *as* machines, based on what machines cannot do:

If any such machines bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together other signs, as we do in order to declare our thoughts to others.... It is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do. Secondly, even though such machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they were acting not through understanding but only from the disposition of their organs. For whereas reason is a universal instrument which can be used in all kinds of situations, these organs need some particular disposition for each particular action; hence it is for all practical purposes impossible for a machine to have enough different organs to make it act in all the contingencies of life in the way in which our reason makes us act. (CSM 1:139-40, AT 6:56-7)

Since the two tests are supposed to derive from what machines *cannot* do, a clarification is in order as to what precisely Descartes takes as the relevant impossibility. On this count, the passage can be read in two ways. It can be read as referring to the kinds of *exhibitions* of behavior that are impossible for machines to produce. It can also be read as referring to the kinds of *processes* by which it is impossible for machines to produce an exhibition of behavior. (The two kinds of readings are not mutually exclusive.) As I want to suggest, the passage is best read in the second way, as referring (exclusively) to constraints on the kinds of productive processes available to machines. In the present section, I'll defend my reading based on its plausibility in the context of the above passage. In §§III.3 through III.5, the reading will emerge as superior based on the details of the examples Descartes develops in other passages.

On my reading, Descartes is not denying that machines can *exhibit* sophisticated language-like behavior — *as if* signifying genuine thought. What he denies is that machines can, in fact, use words *in order* to refer to objects of thought: machines, in that they lack thought, 'could never use words, or put together other signs, as we do *in order* [*pour*] to declare our thoughts to others' (CSM 1:140, AT 6:56; italics added).¹⁹ Explanations in terms of final causality provide the wrong model for understanding purely mechanical behavior. As he conceives of machines, it is strictly impossible for them to adapt means (*qua* means) to ends (*qua* ends), appearances to the contrary notwithstanding.

In addition to this claim about strict impossibility, I take Descartes to be making a separate claim about what is *impractical*. The range of human behavior that machines are able to mimic is owed to the complexity of their mechanical predispositions. He thus regards it as 'practically impossible' (*moralement impossible*) — though he does not say *metaphysically* impossible — for a machine with the limitations of the hypothetical man-machine to have enough predispositions to perfectly imitate everything that our reason enables us to do (AT 6:57). Though Descartes is less than clear, I take him to be alluding to the anatomical limitations that stem from being constructed so as to mimic the appearance and behavior of real humans — thus moving with two legs, working with two hands, producing sounds in the way we do, and so on. But it is not credible that he would intend a clarification of the kinds of

19 For Descartes, the mechanical workings in question — *qua* modes of body — would not even be 'signs' were they not so regarded by a mind. Cf. J.R. Searle, *The Rediscovery of the Mind* (Cambridge, MA: The MIT Press 1992), who makes a similar point about symbols and syntax (ch. 9).

machines that are practically impossible — say, for *us* to create — to shed light on the kinds of machines that *God* did in fact create — much less on the kinds of machines that *God could* create.²⁰ Indeed, while characterizing his hypothetical man-machine, Descartes writes:

We see clocks, artificial fountains, mills, and other such machines which, although only man-made, have the power to move of their own accord in many different ways. But I am supposing this machine to be made by the hands of God, and so I think you may reasonably think it capable of a greater variety of movements than I could possibly imagine in it, and of exhibiting more artistry than I could possibly ascribe to it. (CSM 1:99, AT 11:120)

As I read him, then, Descartes does indeed maintain that there are limitations in what machines can do. But producing impressive exhibitions of behavior is not among these limitations.

The position I attribute to Descartes is not the naive view that *every* machine can imitate every kind of human behavior. To the contrary, less complex machines might be unable to exhibit sophisticated behavior. But insofar as a machine does lack some imitative ability, this inability would be owed to its particular design rather than to essential constraints on any possible machine. It would be possible (surely metaphysically possible, for God) to build a more complex machine that would exhibit the behavior in question.²¹

So understood, Descartes cannot be said to have underestimated the imitative abilities of machines. His views turn out far more sophisticated than some commentators have supposed.²²

20 In *Principles* 4:205, Descartes characterizes *morally certain* matters as 'having sufficient certainty for application to ordinary life, even though they may be uncertain in relation to the absolute power of God' (CSM 1:289-90, AT 8a:327).

21 Given standard mechanist doctrines that Descartes helped establish — whereby all possible variety of secondary quality ideas can be produced (in stimulus-response fashion) by none other than variations of corporeal motion — he surely did not underestimate the potential indefiniteness in the variety of output responses that can be mapped to sequences of just a few kinds of input stimuli (cf. *Principles* 2:23). Note, too, that given the implications of his non-atomist version of mechanism, it is unlikely that Descartes underestimates the potential for miniaturization in machines (cf. *Principles* 2:20, 2:34-5).

22 In contrast with my reading, Cottingham's reading of *Discourse 5* commits Descartes to assumptions about the abilities of machines that, says Cottingham, we today know to be false ('Cartesian Dualism,' 250).

2. Descartes on 'reasoned' behavior — strictly speaking

A further refinement is in order, concerning Descartes's conception of *reasoned behavior*. There is some sense in which even purely mechanical behavior might be reasoned, or purposive. Let's then distinguish this weaker sense of the term, from the strict sense to which Descartes refers in his claims about what machines cannot do.

Descartes denies that machines have the ability to behave *in order* to pursue *their own* ends. He does not thereby deny that they might behave *in accordance* with an end, nor that their behavior might help *bring about* an end. Yet both of these cases involve a kind of purposiveness. Descartes sometimes alludes to such quasi-teleological elements of mechanical behavior by appeal to the purposes of the machine's designer (cf. AT 7:84f, 6:55f, 11:120), or those of the machine's user (cf. AT 7:84f) — themes also discussed by Aquinas.²³ While rebutting an objection to the effect that 'various animals, such as bees, spiders, and dogs' (all of whom he denies a rational soul) exhibit 'wonderful instances of sagacity' even to the extent of appearing 'as though reasoning by way of exclusion' (ST 1a2æ.13.2), Aquinas writes:

[A]ll the things moved by reason display the order of reason [*ordo rationis*], though they themselves are without reason, for instance, an archer flights an arrow and it goes straight to the target as though it were itself endowed with a directing reason [*rationem dirigentem*]. The same appears in the movement of clocks and other works of human art. Now artificial works are to human art as all natural things are to divine art. And so, like the things made by human ingenuity, the things moved by nature display an order, as observed in [Aristotle's] *Physics*. Hence the examples of sagacity in animal behavior are from a natural inclination to carry out the intricate processes planned by supreme art. That is why we call some animals clever or intelligent, not because they are endowed with reason or choice. (ST 1a2æ.13.2)

There are, then, at least two different senses in which a behavior may be *reasoned* or *purposive*. In the one sense, the behavior issues from an occurrent, conscious volition of the behavior. Call such behavior 'reasoned₁' and 'purposive₁' — such reasons and purposes being consciously entertained by the behavior. Since the archer's arrow lacks a conscious mind, its behavior could not be reasoned₁/purposive₁. Likewise, the behavior of the circulatory and digestive systems of the human body is not typically reasoned₁/purposive₁, but is instead (what Descartes calls) purely mechanical. I might form an occurrent volition to the effect that

23 Though of course via his hylomorphic conception of the beings Descartes regards as pure machines.

my blood will continue to circulate normally, but given that such continuance does not *issue from* my volition, nor is my circulatory system 'itself endowed with a directing reason' (to use Aquinas's way of putting it), this is not reasoned₁ behavior. There is, however, another sense in which the behavior of the archer's arrow is reasoned and purposive: namely, the sense in which it accords with, and indeed helps bring about, the purposes of the archer. Similarly, our circulatory and digestive systems are marvelously purposive, but their purposiveness is owed (according to Descartes) to their creator.²⁴ More generally, the reasons₁ and purposes₁ of designers are at play in the behavior of the well-functioning machines they design. Call such behavior 'reasoned₂' and 'purposive₂' — there being reasons₁/purposes₁ because of which the behavior occurs, though they are not the behavior's own.

As I understand Descartes's account, purely mechanical behavior is reasoned₂, but not reasoned₁. An organism's own consciousness is integral to producing reasoned₁ behavior, but such consciousness (if there be any) is idle in the production of its purely mechanical behavior.²⁵

The classification of behavior — as reasoned₁ versus purely mechanical — is often difficult. In some cases the difficulty stems from a wholly *genetic* role for reason₁. Consider that in the normal course of driving a car I do not consciously attend to the purposive way in which my limbs coordinate the accelerator, clutch, and gear shift. Such limb movements do not issue from any conscious, occurrent volition of mine, even though the disposition from which they do issue was acquired with the help of previous such volition — past, repeated instances of reasoned₁ behavior (i.e. conditioning, in the relevant sense).²⁶ Since occurrent volition does not factor into the production of acquired reflex behavior of this sort,

24 For Descartes, *all* purely mechanical behavior displays the order and purpose of the divine will. In his Sixth Meditation discussion of dropsy error, Descartes notes that even where machines (whether clocks or physiological systems) appear to be malfunctioning, they behave in accordance with their God-given natures (AT 7:83ff).

25 The distinction between reasoned₁ and reasoned₂ behavior is similar to, but not exactly the distinction between acting on a reason and acting in accordance with a reason. As I mean to construe it, reasoned₂ behavior is not merely in accordance with reason — not merely reasonable. Rather, there is a sense in which such behavior occurs *because of* a reason — it is reasoned — though the guiding reasons are not the behavior's own.

26 In other cases, the disposition is acquired without any volitional contribution from the behavior. In a 1630 letter, Descartes describes what he regards as such a case: 'I reckon that if you whipped a dog five or six times to the sound of a violin, it would begin to howl and run away as soon as it heard that music again' (CSMK 20, AT 1:34).

Descartes classifies it as 'purely mechanical' (cf. AT 7:229f) — a sort that he thinks includes some cases of walking, talking, singing, and more (cf. AT 4:573, 7:229f, 11:362, 11:368-70). In contrast, commonsensical classifications of such behavior are more apt to track a behavior's apparent subtlety. Where purely mechanical behavior appears very subtle, it's perhaps inevitable that we'd describe it in terms apropos of reasoned₁ behavior — e.g. in terms of the behavior's own occurrent beliefs, desires, and the like. From Descartes's standpoint such descriptions are, strictly speaking, no more than a *façon de parler*.

3. Unmasking an automatism as an automatism

Let us assume with Descartes that machines cannot engage in reasoned₁ behavior, though they can exhibit it. Descartes's two tests are designed to unmask some (but not all) such exhibitions of behavior as automatisms, as I want now to explain.²⁷

The following classification scheme will help to clarify the account:

- (A) exhibitions of behavior that are implausible to explain as other than reasoned₁;
- (B) exhibitions of behavior that are implausible to explain as other than purely mechanical;
- (C) exhibitions of behavior that, *prima facie*, are plausibly explained as reasoned₁ or as purely mechanical; two cases:
 - (i) the behavior is unmasked as type (B), upon further experiment;
 - (ii) the behavior is not unmasked upon further experiment.

So construed, the divisions between categories are *epistemic*: judgments concerning whether a given behavior belongs in, say, type (C) rather than type (A), depend on one's beliefs as to the explanatory power of mechanical physiology. Descartes has various arguments intended to persuade his

27 An analogical argument focusing on behavior would seem to offer the most promising strategy available to Descartes. Granting his substance dualism (among our assumptions, for purposes of understanding his project), there's no possibility of identifying structures of the central nervous system in humans in which consciousness resides, so as to infer to the existence or non-existence of consciousness in brutes based on the presence or lack of similar such structures.

readers to classify, as type (C), many forms of behavior that they'd have previously classified as type (A).²⁸ Given the interpretive aims of the present Section, I shall assume the success of such arguments.

Consider (A). In claiming that Descartes's two tests are intended only as *negative* tests of reason₁, I mean to rule out the conclusion that a behavior is of type (A) as a possible outcome of the tests.²⁹ Bear in mind that a behavior may *be* reasoned, while not falling into type (A). In §III.6, I shall return to a discussion of category (A) and whether it includes any human behavior. For now, suffice it to say that Descartes includes no behavior of brutes in type (A).

Consider (B). Descartes thinks that vast numbers of brutes exhibit none other than purely mechanical behavior as falls under (B). He takes it as implausible to deny that brutes 'such as oysters and sponges' (CSMK 304, AT 4:576), or 'worms, flies, caterpillars and other [such] animals move like machines' (CSMK 366, AT 5:277). Bear in mind, Descartes does not purport to establish the *bête machine* doctrine with metaphysical certainty. And insofar as we're to argue by analogy, he thinks the behavior of such organisms is akin to that of our circulatory or respiratory systems, or perhaps the statues at the Royal Gardens at Saint-Germain whose motion he attributes to a hydro-mechanism (cf. AT 11:130f).

Consider (C), the interesting category for our purposes. As I interpret Descartes, much of the behavior of brutes belongs in this category. But he thinks that careful experiment and reflection reveal that all of the type (C) behavior of brutes is also of type (Ci) — thus, resolving ultimately into type (B). In a 1646 letter, he writes:

28 *Discourse 5* is weak in this regard, though, in his post-*Discourse* writings, Descartes improves. For example, in the Fourth Replies he offers the following argument: 'When people take a fall, and stick out their hands so as to protect their head, it is not reason that instructs them to do this; it is simply that the sight of the impending fall reaches the brain and sends the animal spirits into the nerves in the manner necessary to produce this movement even without any mental volition, just as it would be produced in a machine' (CSM 2:161, AT 7:229-30). Elsewhere, Descartes strengthens the introspective force of the argument, adding that 'even if we expressly willed not to put our hands in front of our head when we fall, we could not prevent ourselves' (CSMK 302-03, AT 4:573; cf. also *Passions* 13).

29 The standard interpretation of Descartes has it that his two tests are intended to yield either a positive or a negative outcome — i.e., that they are intended to reveal *whether or not* a behavior is reasoned. Cf. Bennett, 'Thoughtful Brutes,' 199; Chomsky, *Cartesian Linguistics*, 78n.9; Cottingham, 'Cartesian Dualism,' 250; Radner and Radner, *Animal Consciousness*, 41; Rozemond, 'The Role of the Intellect,' 99; and Wilson, 'Animal Ideas,' 12, all of whom are committed to the claim that Descartes intends the tests as both positive and negative tests.

I know that animals do many things better than we do, but this does not surprise me. It can even be used to prove they act naturally and mechanically, like a clock which tells the time better than our judgment does. (CSMK 304, AT 4:575)³⁰

It is in connection with type (Ci) behavior that unmaskings occur — impressive exhibitions of behavior which, on closer inspection, are best explained as purely mechanical.

Unmaskings rest on what I shall refer to as the Universal Instrument Thesis. As Descartes writes, 'reason is a universal instrument which can be used in all kinds of situations'³¹ (CSM 1:140, AT 6:57). I take his unmasking strategy to arise from the following, partly analogical considerations. As revealed in part by experience, reason₁ allows us to adapt an indefinite variety of means in the pursuit of an indefinite variety of ends, in an indefinite variety of contexts. This universal adaptability of reason₁ affords us the ability to confront new, even unique challenges with new tactics specially devised for such purposes. Notably, we're able to use our anatomy *in order* to signify indefinitely many meanings. Though no machine can produce signing behavior *in order* to achieve an end, this inability might remain *masked* in machines that are adept at mimicking the relevant behavioral result. But in machines that are inept in this regard, their lack of reason₁ — their lack of universal adaptability — might be exposed in a manner that unmasks their behavior *as* mechanical.

On my account, Descartes's two tests involve two kinds of unmaskings. Before turning to his own treatment, consider an example of category (Ci) behavior stemming from recent experiments with wasps, an example that is arguably more perspicuous than any that Descartes explains:³²

When the time comes for egg laying, the wasp *Sphex* builds a burrow for the purpose and seeks out a cricket which she stings in such a way as to paralyze but not kill it. She drags the cricket into the burrow, lays her eggs alongside, closes the burrow, then flies away, never to return. In due course, the eggs hatch and the wasp grubs feed off the paralyzed cricket, which has not decayed, having been kept in the wasp

30 Descartes's early notebooks record related thoughts: 'The high degree of perfection displayed in some of their actions makes us suspect that animals do not have free will' (CSM 1:5, AT 10:219).

31 He adds, 'spirits are by their nature adapted with equal facility to a great variety of actions' (CSM 2:161, AT 7:229).

32 The example is cited by D. Dennett, 'Mechanism and Responsibility,' in T. Honderich, ed., *Essays on Freedom and Action* (London: Routledge & Kegan Hall 1973), who uses it to illustrate a rather different point.

equivalent of deep freeze. To the human mind, such an elaborately organized and seemingly purposeful routine conveys a convincing flavor of logic and thoughtfulness — until more details are examined. For example, the wasp's routine is to bring the paralyzed cricket to the burrow, leave it on the threshold, go inside to see that all is well, merge, and then drag the cricket in. If, while the wasp is inside making her preliminary inspection, the cricket is moved a few inches away, the wasp, on emerging from the burrow, will bring the cricket back to the threshold, but not inside, and will then repeat the preparatory procedure of entering the burrow to see that everything is all right. If again the cricket is removed a few inches while the wasp is inside, once again the wasp will move the cricket up to the threshold and reenter the burrow for a final check. The wasp never thinks of pulling the cricket straight in. On one occasion, this procedure was repeated forty times, always with the same result.³³

Where a behavior is reasoned₁, the Universal Instrument Thesis renders it incredible that the organism would fail to use reason₁ to devise new means of achieving its intended ends. Of course, various explanations might be devised to maintain the hypothesis that such behavior is reasoned₁, but Descartes thinks that the more plausible explanation (insofar as analogically grounded) is that such exhibitions of behavior stem from a purely mechanical process in which the needed adaptation is lacking.

Unmasking strategies can be characterized in the form of a *reductio ad absurdum*. An animal exhibits *apparently* reasoned₁ behavior. Assume the behavior is reasoned₁. The Universal Instrument Thesis warrants the prediction that the organism will exhibit a wide range of other reasoned₁ behavior (i.e. of that behavior that it has an anatomical capacity to perform).³⁴ Experiments intended to elicit the predicted behavior systematically fail to do so. To avoid absurdity, the assumption is denied. The original behavior is unmasked.³⁵

Given Descartes's own examples, he does not require that unmaskings take the precise form as occurs in the wasp example, nor that the results

be as palpable; nor does he hold that it is practically possible to unmask every machine. The automatism of extraordinarily complex machines might be such as to elude even the most ingenious attempts (by humans) at being unmasked, thus resolving into category (Cii) behavior. What all successful unmaskings have in common is the exhibition of behavior seeming to stem from the organism's own reasons and purposes — i.e., 'until more details are examined.'

Dramatic unmaskings might occur in connection with behavior that, *prima facie*, *surpasses* that of humans. Montaigne had a quite different view of apparently superior behavior in brutes. Objecting on analogical grounds, he asks: 'Why do we attribute to some sort of natural and servile inclination these works which surpass all that we can do by nature and by art?' (333) In a revised edition of the 'Apology' his view is conveyed in an embellishment of the *like effects, like causes* principle:³⁶

There is no apparent reason to judge that the beasts do by natural and obligatory instinct the same things that we do by our choice and cleverness. We must infer from like results like faculties, and from richer results, richer faculties, and consequently confess that this same reason, this same method that we have for working, the animals have it also, or some better one.³⁷ (336-7; italics added)

Descartes offers a very different analysis:

It is also very remarkable that, although many animals might show more ingenuity [*d'industrie*] than we do in some of their actions, they nonetheless show none at all in many others: so that what they do better does not prove that they have a mind [*l'esprit*] — for, in that case, they would have more [ingenuity] than any of us and would do better in all things — but rather [it proves] that they have none at all. (AT 6:58-59; my translation³⁸)

33 D.E. Wooldridge, *The Machinery of the Brain* (New York: McGraw-Hill 1963), 82-3

34 The parenthetical qualification is crucial, as it is unreasonable to expect behavior for which the organism is anatomically unfit. In some passages, Descartes shows a sensitivity to such concerns (cf. AT 6:57f, AT 4:575).

35 There is a *weak* and a *strong* reading as to the immediate conclusion we're to draw from an unmasking. On the weak reading, an unmasking shows that the *behavior* in question is not reasoned₁. On the strong reading, it shows that the *behavior* in question lacks a faculty of reason — because (as this reading would have it) any organism in possession of a faculty of reason would employ it in the scenarios at play in the unmasking. Though there are some textual grounds for the stronger reading, I believe that Descartes is best interpreted according to the weak reading.

36 The superiority of the behavior of brutes is among Montaigne's central themes: 'we recognize easily enough, in most of their works, how much superiority the animals have over us and how feeble is our skill to imitate them' (333).

37 The italicized remarks were added to the 1595 edition of Montaigne's *Essays*. The changes signal no more than a shift in emphasis, as they are entailed by the original version of the causal principle, namely, from *like effects, like causes*.

38 The passage poses translation difficulties because of the ambiguous reference in connection with the remark, '*car, à ce compte, ils en auraient plus qu'aucun de nous.*' (They would have more of what than any of us?) Most translations take the remark to refer back to '*esprit*,' thus resulting in the awkward claim that brutes would have more mind than any of us. (I can make no clear sense of the claim that 'x has more mind than y.') CSM avoids this, but only with the somewhat strained translation of *esprit* as 'intelligence.' I take the referent to be '*industrie*,' resulting in a very natural translation (albeit too literal) in connection with my unmasking interpretation — brutes would have more ingenuity than any of us.

This analysis suggests a revision to Montaigne's causal principle (though even this is misleading): *from richer effects, purely mechanical causes*. What Descartes regards as incredible is *not* that machines should outperform reasoning₁ humans — recall him saying, 'I know that animals do many things better than we do, but this does not surprise me.' What is incredible is that a being who regularly performed such feats using reason₁ would fail to adapt it in a variety of contexts. Descartes takes the available empirical data to show that brutes 'surpass us only in those of our actions which are not guided by our thought' (CSMK 302, AT 4:573).³⁹

An important consequence of the present interpretation warrants clarification. One kind of strategy (one Descartes thinks is wrong-headed) for ascertaining the underlying causes of a behavior involves a simple *pass-or-fail* test based on the degree of subtlety of exhibited behavior: where the behavioral subtlety is above a (presumably arbitrary) threshold, the behavior is judged as reasoned₁; where below, it is judged as purely mechanical. Descartes *must* reject this approach in view of the Imitation Thesis, a thesis that allows for dual kinds of causes of a given level of apparent subtlety. On my interpretation, there is no case in which Descartes judges the behavior of brutes as purely mechanical on the basis of any such *pass-or-fail* test. Where behavior is none other than type (B), it *is* correct that he judges it purely mechanical because of falling below a threshold of sorts;⁴⁰ but it is not a threshold above which the same behavior would fall into type (A). And where a behavior is unmasked, it would miss the point entirely to say that Descartes's conclusion stems from setting the passing grade threshold (for what counts as 'apparently reasoned₁') too high. Type (Ci) behavior (i.e. that which is unmasked) *does* appear reasoned₁; exhibiting impressive behavior is a *prerequisite* to an unmasking.

Descartes's two tests — his two kinds of unmaskings — both draw on the hypothetical man-machine thought-experiment introduced earlier in

connection with (1). The one is based on linguistic behavior; the other on non-linguistic behavior, as in the wasp example. I now consider each.

4. The linguistic unmasking test

Further reflection on the hypothetical man-machine suggests linguistic unmasking strategies based on the kinds of linguistic behavior it would be adept at imitating. What are these more easily imitated kinds of linguistic behavior? Where human language occurs in *reactive* contexts, namely those in which it is responsive to external stimuli (linguistic or otherwise), it is amenable to perfect imitation by a suitably designed stimulus-response machine. Descartes indeed portrays a hypothetical 'talking' automaton that exhibits a limited degree of such responsiveness:

We can certainly conceive of a machine so constructed that it utters words, and even utters words which correspond to bodily actions causing a change in its organs (e.g. if you touch it in one spot it asks what you want of it, if you touch it in another it cries out that you are hurting it, and so on). (CSM 1:140, AT 6:56)

Descartes often characterizes these *reactive* contexts in terms of the *passions*:⁴¹

And we must not confuse speech with the natural movements which express passions and which can be imitated by machines as well as by animals. (CSM 1:140-1, AT 6:58)

Even in us all the motions of our limbs which accompany our passions are caused not by the soul but simply by the machinery of the body. The wagging of a dog's tail is only a movement accompanying a passion, and so is to be sharply distinguished, in my view, from speech. (CSMK 374, AT 5:344-5)

In contrast, Descartes notes what a mechanical signer could *not* do. Again, he locates the inability of a machine in its means of behavior production, not in the degree of apparent subtlety of the behavior produced. Though the mechanical signer could *exhibit* language-like behavior, it is 'not conceivable' that it should produce such strings of words 'so as [pour] to give an appropriately meaningful answer to

39 It's worth noting that Descartes intended to run further experiments but evidently never found time. In a 1645 letter (probably) to the Marquess of Newcastle, Descartes writes: 'The treatise on animals, on which I began work more than fifteen years ago, cannot be finished until I have made many observations which are essential for its completion, and which I have not yet had the opportunity to make (nor do I know when I shall have it)' (CSMK 274, AT 4:326). Essentially the same point is made to More (AT 5:344), in the penultimate year of Descartes's life.

40 Though this need not be arbitrary where such behavior (cf. oysters) is well below any plausible threshold range.

41 In *Passions* 21 and 25, Descartes explains that he uses 'passion' with dual reference. In discussions related to brutes and consciousness, he uses the term in (what he calls) the more general sense, roughly the sense whereby our passions are those occurrences that happen to us purely mechanically, rather than being made to happen by our thought (cf. *Passions* 1).

whatever is said in its presence' (CSM 1:140, AT 6:56-7; my italics). Whereas reason₁ enables new meaningful word combinations to be generated via consideration of word meanings, a machine can generate only those combinations for which it is mechanically programmed. And since machines lack thought, they are incapable of stimulus-free expression of their thoughts.

Of course, we can imagine a machine that is cleverly designed to initiate language-like behavior in a seemingly stimulus-free manner. It periodically begins conversations that are not triggered by environmental stimuli. But this possibility does not undermine Descartes's case. His case depends on the thesis that some machines can be unmasked, but not on the thesis that every machine can. For those machines with design limitations of a particular sort, appropriate empirical investigation might unmask their exhibitions of language-like behavior by exposing behavioral patterns that are best explained as the effects of purely mechanical processes.

In devising a strategy Descartes builds on the thought that, where an organism exhibits language-like behavior only in stimulus-dependent contexts, this is symptomatic of a lack of reasoning₁ faculties. Recall that unmaskings take the form of a reductio. Given: brutes exhibit language-like behavior of type (C). Assume that such signing is indeed reasoned₁ behavior.⁴² Given the Universal Instrument Thesis, the brutes in question should also exhibit stimulus-free expressions of thought. As Descartes observes:

Since dogs and some other animals express their passions to us, they would express their [non-passionate] thoughts also if they had any. (CSMK 303, AT 4:575)

Descartes takes it that human language users (on whom the analogy rests) routinely express thoughts that are not palpably dependent on corporeal stimuli. Yet, according to Descartes, there is no credible evidence of such behavior in brutes. In a letter to More, he notes that

animals easily communicate to us, by voice or bodily movement, their natural impulses of anger, fear, hunger and so on.⁴³ Yet in spite of all these facts, it has never

42 In effect, assume with Montaigne that the signing of 'blackbirds, ravens, magpies, and parrots,' and other brutes, 'testifies that they have an inward power of reason' (cf. 339-40). Montaigne extends his position to all manner of bodily signs, not just verbal ones (332). Descartes would have us extend our assumption to the same (cf. AT 6:58, 4:574).

43 Prima facie, this remark implies that Descartes does not mean to deny brutes such consciousness states as anger, fear, hunger, and the like. I shall return to this in §IV.1.

been observed that any brute animal attained the perfection of using real speech, that is to say, of indicating by word or sign something relating to thought alone and not to natural impulse. (CSMK 366, AT 5:278)

In the Newcastle letter, he adds that 'there has never been known an animal [non-human] so perfect as to use a sign to make other animals understand something which bore no relation to its passions' (CSMK 303, AT 4:575). I take Descartes to mean, in these passages, that the language-like behavior we observe in brute animals *always* arises in obviously stimulus-dependent contexts. Assuming that Descartes is right,⁴⁴ the initial assumption is discredited.⁴⁵ The apparently linguistic behavior turns out better explained as purely mechanical.⁴⁶

On one kind of objection, the better explanation of the lack of (what Descartes would countenance as) 'real speech' is not that brutes lack reason₁, but that they lack the anatomical features needed for signing. Descartes anticipates the objection while pressing analogical considerations:

This [lack of non-passionate signing] does not happen because they lack the necessary organs, for we see that magpies and parrots can utter words as we do, and yet they cannot speak as we do: that is, they cannot show that they are thinking what they are saying. On the other hand, men born deaf and dumb, and thus deprived of speech-organs as much as the beasts or even more so, normally invent their own signs to make themselves understood by those who, being regularly in their company, have the time to learn their language. (CSM 1:140, AT 6:57-8)

On another kind of objection, Descartes is guilty of special pleading. He requires that brutes exhibit stimulus-free instances of language-like

44 Bennett (202-6) offers an interesting overview of some recent experimental work (by David Premack, et al.) that purports to establish that brutes exhibit stimulus-free signing behavior. Among other things, the discussion highlights the difficulties involved in establishing signs as stimulus-free.

45 Descartes does not deny that mechanical systems may incorporate — what is, from the point of view of their designer — *signaling*. He indeed characterizes the motions of animal spirits in our nerves as signals (cf. AT 7:86ff), though of course purely corporeal signals.

46 In significant respects, the procedure is similar to Alan Turing's Imitation Game. See A. Turing, 'Computing Machinery and Intelligence,' *Mind* 59 (1950) 433-60. But where Turing's test aims at unmasking the causes of an exhibition of behavior as involving thinking (i.e. insofar as the notion of thinking, on his view, is to be meaningful), Descartes's tests aim at unmasking the causes as purely mechanical. (I should add that there is no easy way to link Turing's concept of *thinking* with any concept in Descartes's scheme.)

behavior, even though — as he's surely aware — many competent human language users seem to speak only when spoken to, while others (e.g. infants) may exhibit no linguistic behavior at all. According to the objection, Descartes's appeal to the Universal Instrument Thesis requires that reasoning brutes *do* employ reason, in every situation they *can*, though this requirement is not met even in humans. The objection is a good one. Unfortunately, Descartes says very little that would serve to clarify his response. Nevertheless, one kind of reply is suggested in his claim that 'it would be incredible that a superior specimen of the monkey or parrot species should not be able to speak as well as the stupidest child — or at least as well as a child with a defective brain' (CSM 1:140, AT 6:58). Given the *superior specimen* remark, Descartes may hold that the reasoned₁ behavior of *even one* member of a species is sufficient to forestall the unmasking of *any* member of the species. This suggestion comports well with a remark to More:

Infants are in a different case from animals: I should not judge that infants were endowed with minds unless I saw that they were of the same nature as adults; but animals never develop to a point where any certain sign of thought can be detected in them. (CSMK 374, AT 5:345)

On this reading, exhibitions of linguistic behavior in brutes fall into type (Ci) (and thus resolve into type (B)) in part because of a stimulus-dependence that (according to Descartes) runs across the entire species; in contrast, exhibitions of linguistic behavior in human infants fall into type (Cii) (remaining masked) because of exhibitions by other humans.

5. The non-linguistic unmasking test

Yet further consideration of the hypothetical man-machine suggests a non-linguistic unmasking strategy. Since its behavior would stem from the predisposition of its mechanical parts (AT 6:57), it would inevitably come up short in the kinds of responses it *could* produce when compared with the *universal* adaptability of reason₁. Consequently, machines that exhibit impressive behavior are susceptible to unmaskings insofar as they lack mechanical predispositions needed to exhibit a wide variety of such behavior.

Descartes's *Discourse 5* discussion of non-linguistic unmaskings is largely unsatisfying.⁴⁷ He addresses none of the myriad examples of

specific behavior alleged (in the popular literature by Montaigne, et al.) as best explained by appeal to a rational faculty. Descartes's aim seems to be not so much to unmask specific such behavior, in piecemeal fashion, as to convey a general strategy by which any such type (C) behavior is to be unmasked. Pick your favorite examples of brutes that 'show more skill than we do' (CSM 1:141, AT 6:58) — e.g., the swallows and their seemingly stupendous compass and calendar abilities that enable them to arrive in San Juan Capistrano every spring. Assume that such behavior is in fact reasoned₁ — that, e.g., the swallows employ reason₁ in *judging* when and where to arrive. In keeping with the Universal Instrument Thesis, Descartes thinks we should then expect that the brutes in question 'would excel us in everything' (i.e. in every kind of behavior of which they are anatomically capable); or, at any rate, that they would exhibit signs of similarly subtle reasoning₁ across a variety of contexts. We find otherwise. Descartes elsewhere notes that

though their movements are often more regular and certain than those of the wisest men, yet in many things which they would have to do to imitate us, they fail more disastrously than the greatest fools. (CSMK 99, AT 2:40)

The best account is thus 'that they act naturally and mechanically, like a clock which tells the time better than our judgement does' (CSMK 304, AT 4:575); 'a clock, consisting only of wheels and springs, can count the hours and measure time more accurately than we can with all our wisdom' (CSM 1:141, AT 6:59).

Though, in his *Discourse 5* handling of non-linguistic unmaskings, Descartes does not treat any specific examples, in later correspondence he does — finally addressing a variety of allegations (as arise in the

suggesting is to be understood as presenting an independent unmasking — to be a continuation of the earlier argument addressing linguistic behavior. Descartes might thus intend to treat the universal lack of language in brutes (as he thinks he's now shown) as primary evidence relevant to the non-linguistic unmaskings. So understood, the reasoning (again, analogical) might run as follows. As he's already argued, 'it patently requires very little reason to be able to speak' — a claim he supports by analogy to various humans with marginal intellects (CSM 1:140, AT 6:58). But, in accordance with the Universal Instrument Thesis, it is implausible that any organism that does not use language would excel us in *any* manner via the assistance of reason. The better explanation of such 'superior' behavior is that it is produced purely mechanically. Assuming this is the correct reading, Descartes's linguistic unmasking test is perhaps intended to apply to every brute animal that exhibits any behavior of type (C) — a not implausible thesis, given how pervasive are passionate signings among brutes.

⁴⁷ One plausible explanation of the seemingly cursory treatment is that Descartes intends the discussion of non-linguistic behavior — a discussion I have been

Montaigne-related literature) of specific, reasoned₁ behavior. To the Marquess of Newcastle and to More, respectively, Descartes writes:

Doubtless when the swallows come in spring, they operate like clocks.⁴⁸ The actions of honeybees are of the same nature;⁴⁹ so also is the discipline of cranes in flight, and of apes in fighting, if it is true that they keep discipline. Their instinct to bury their dead is no stranger than that of dogs and cats which scratch the earth for the purpose of burying their excrement; they hardly ever actually bury it, which shows that they act only by instinct and without thinking. (CSMK 304, AT 4:575)

I am not disturbed by the astuteness and cunning of dogs and foxes,⁵⁰ or by all the things which animals do for the sake of food, sex, and fear; I claim that I can easily explain all of them as originating from the structure of their bodily parts. (CSMK 365, AT 5:276)

6. Language as a positive test

Though on my reading of *Discourse 5*, Descartes does not there characterize positive tests of reasoned₁ behavior, in his 1646 letter to the Marquess of Newcastle he proposes that language — and *only* language — might serve as the basis of a positive test:

In fact, none of our external actions *can* show [*puisse assurer*] anyone who examines them that our body is not just a self-moving machine but contains a soul with thoughts, *with the exception* of spoken words, or other signs that have *reference to [à propos des]* particular topics without expressing any passion. (CSMK 303, AT 4:574; italics added)

The feature of genuine language use that he thinks resists perfect imitation by mechanical signers is the occurrence of signs that are *context-appropriate while being stimulus-free*.⁵¹

48 Montaigne rhetorically asks, 'Do the swallows that we see on the return of spring ferreting in all the corners of our houses search without judgment, and choose without discrimination, out of a thousand places, the one which is most suitable for them to dwell in?' (333)

49 Montaigne asks, 'Is there a society regulated with more order ... than that of the honeybees? Can we imagine so orderly an arrangement of actions and occupations as this to be conducted without reason and foresight?' (332)

50 Montaigne recounts the dog whose behavior results from an 'act of pure logic' involving the 'use of propositions divided and conjoined' (339). Of foxes, he suggests that they use 'the same reasoning' as humans (337).

51 Cf. Chomsky, 4ff.

I add also that these words or signs must not express any passion, to rule out not only cries of joy or sadness and the like, but also whatever can be taught by training to animals.⁵² If you teach a magpie to say good-day to its mistress when it sees her approach, this can only be by making the utterance of this word the expression of one of its passions. For instance it will be an expression of the hope of eating, if it has always been given a tidbit when it says it. Similarly, all the things which dogs, horses and monkeys are taught to perform are only expressions of their fear, their hope or their joy,⁵³ and consequently they can be performed without any thought. (CSMK 303, AT 4:574-5)

Prima facie, the Newcastle letter reveals a commitment to a positive test of reason₁ — thus putting Descartes at odds with my earlier contention that he does not hold a naive view of the imitative abilities of machines. I'm inclined to resist this prima facie reading. The Newcastle letter contains the only passage in which Descartes discusses such a test, and its brevity and wording on the subject leave us with little more than speculation concerning important details. The questions left unanswered are numerous and significant. Does his commitment to a positive test mark a change of view, or does he take it to be continuous with his *Discourse 5* account? Does he mean that such behavior would elude any possible machine, or does he mean to refer only to those machines which (like his hypothetical man-machine) produce all their behavior by means of anatomical features like ours? Is Descartes's point that it would be 'practically impossible' for a machine to exhibit such behavior (cf. §III.1), or does he mean something stronger? Precisely what would Descartes count as a context in which linguistic behavior is *both* context-appropriate *and* stimulus-free? (What context-making feature does not *also* count as a candidate stimulus?) What proportion of normal humans does he think ever exhibit such type (A) behavior? Among those humans that do, what proportion of their normal speech does he think counts as type (A)? Does any *normal* speech count, or does Descartes mean to refer to *unique* linguistic contexts?⁵⁴ Does he think that any *actual* human speech counts as type (A), or is he instead claiming that some *merely possible* human signing would count as the 'external actions [that] *can* show

52 Elsewhere, it is clear that Descartes thinks brutes can be *taught* in the sense of being *conditioned* (by means of rewards and punishments) via none other than purely mechanical processes. (See note 26.)

53 Again (cf. note 43), such remarks might seem to imply that Descartes allows brutes such conscious states as fear, hope, and joy. I shall return to this in §IV.1.

54 Alan Turing's famous example of a linguistic exhibition that passes his own (Turing) test (446) would not qualify as type (A) behavior in view of the stimulus-dependent character of the witness's responses to the interrogator.

[*puisse assurer*] anyone who examines them that our body is not just a self-moving machine but contains a soul with thoughts'?

A mere clarification of these unanswered questions serves to counter-balance the *prima facie* reading. I'll not here speculate as to the answers to such questions, since a variety of plausible answers are consistent with my interpretation.

IV Descartes's Appeal To the Principle of Parsimony

A non sequitur in Descartes's case arises, if the premise set in support of the *bête machine* doctrine is exhausted by (1) and (2). In connection with these premises, Descartes argues that neither sensory consciousness nor a faculty a reason, is necessary for the explanation of any behavior of the brutes. (1) and (2) exhaust the premises explicitly offered in *Discourse 5*, but the *bête machine* doctrine does not follow from their conjunction. This premise set does not rule out that brutes have epiphenomenal (i.e. causally idle) consciousness.⁵⁵ The apparent non sequitur is all the more egregious insofar as one takes the case for (2) to exhaust Descartes's case for the *bête machine* doctrine.

This seeming logical gap between Descartes's actual case, and the conclusion of the *bête machine* doctrine, has fostered misunderstandings in the scholarly literature. One camp of interpreters acknowledges that Descartes does officially hold the *bête machine* doctrine, while accusing him of slipping between the wide and the narrow sense of 'thought':⁵⁶ he argues only that brutes have no thought in the narrow sense (= reason/intellect), but then concludes that they have no thought *at all* (including sensory awareness). A second camp of interpreters has Descartes rejecting the *bête machine* doctrine, maintaining that his official view is to deny brutes reason but not sensory awareness.⁵⁷ We should want to dispel the worries of the first camp, out of charity. We should

55 Given my *weak* reading of the unmasking tests, whereby Descartes intends to unmask behavior not behavers (cf. note 35), the alleged non sequitur arises not only in connection with epiphenomenal *sensation*, but also epiphenomenal *rationality*. On a more standard interpretation, however, since Descartes's two tests are supposed to show that brutes lack a faculty of reason, the non sequitur arises only in connection with epiphenomenal *sensation*. I shall thus focus my attention on this latter understanding of the alleged non sequitur.

56 Cf. Radner and Radner (79), Loeb (112f), and Vendler (152ff).

57 Cf. Cottingham, 'A Brute to the Brutes?' 230; and Gaukroger, 289 and 454n.165.

want to dispel those of the second, on textual and doctrinal grounds — though I shall have to argue this point.

The worries of neither camp are groundless, given the *express* argument of *Discourse 5*. But such worries are dispelled in the context of the larger corpus: there are clear and copious texts establishing the tacit step needed to justify the conclusion of the *bête machine* doctrine, along with Descartes's motive for being elliptical. In §IV.1, I argue that a strategy of preempting the non sequitur by denying that Descartes holds the *bête machine* doctrine is ill-grounded. In §IV.2, I argue that Descartes is appealing to the principle of parsimony — an appeal that avoids the non sequitur. In §IV.3, I will touch on a remaining line of objection.

1. Brutes and sensory awareness

In various places Descartes refers to such passions, in brutes, as 'their hope or their joy' (CSMK 303, AT 4:574-5), and their 'impulses of anger, fear, hunger and so on' (CSMK 366, AT 5:278). As one prominent commentator observes, Descartes's reference to 'impulses of anger, fear, hunger' and so on, is 'quite extraordinary' and unexpected coming from 'a man who is supposed to believe animals are "without feeling or awareness of any kind"' (Cottingham, 'A Brute to the Brutes?' 231), adding that nowhere 'does Descartes commit himself to the monstrous thesis that they [brutes] have no feelings or sensations' (ibid., 230). As already noted, if Descartes did not hold the *bête machine* doctrine the egregious non sequitur is avoided. It has thus seemed attractive to some commentators to deny that Descartes holds the doctrine.

Though in attributing passions to the brutes Descartes might seem to refer to *conscious* passion, there is compelling textual evidence to the contrary. In all cases of sensation, appetites, and passions, which involve (in humans) both a corporeal and an incorporeal component (cf. *Principles* 1:48), the relevant terminology — e.g., 'vision,' 'hunger,' 'pain' — is understood, by Descartes, to have dual reference. When speaking strictly, Descartes refers such terminology to a mental component alone: in the Second Meditation, he remarks that 'what is called "having a sensory perception" is strictly just this [the conscious seeming], and in this restricted sense of the term it is simply thinking' (CSM 2:19, AT 7:29). But in other, less strict contexts, Descartes is clear in referring such terminology to the physiological processes that give rise to the corresponding passions in humans.⁵⁸ To Mersenne, Descartes notes that he

58 In response to a critic of the *Discourse* who supposes that Descartes allows that brute

does 'not explain the feeling of pain without reference to the soul,' though in brute 'animals it is [mechanical] movements alone which occur, and *not pain in the strict sense*' (CSMK 148, AT 3:85; my italics). Ironically, in the very letter (cited by Cottingham) referring to 'impulses of anger, fear, hunger and so on,' Descartes adds that he does not deny that brutes have 'sensation, *in so far as it depends on a bodily organ*' (CSMK 366, AT 5:278; my italics). While developing his thought-experiment of the hypothetical man-machine — a being which, by hypothesis, has no rational, sensitive, or vegetative soul — Descartes characterizes the machine as experiencing 'hunger' (*faim*), 'pain' (*douleur*), 'joy' (*joye*), 'anger' (*colere*), 'sadness' (*tristesse*), and 'other passions' (*autres passions*) (AT 11:199); Descartes elsewhere refers to its physiological states as 'internal passions' (*passions intérieures*) (AT 6:55). Note too that in *Principles* 4:190 (cf. *Passions* 47), Descartes refers to natural appetites which depend only on physiology — e.g., 'hunger and thirst which depend on the nerves of the stomach, throat and so forth' (CSM 1:281, AT 8a:317-18). As explained in the Sixth Replies, Descartes distinguishes three grades of sensory response. What he calls the 'first grade' includes only physiological — i.e. purely mechanical — aspects of sensation:

[It] is limited to the immediate stimulation of the bodily organs by external objects; this can consist in nothing but the motion of the particles of the organs, and any change of shape and position resulting from this motion. (CSM 2:294, AT 7:436-7)

Only this first grade of sensory response is explicitly said by Descartes to be 'common to us and the brutes' (CSM 2:295, AT 7:437).

On closer inspection, then, Descartes's reference to 'impulses of anger, fear, hunger,' and the like, do not signal a commitment to sensory awareness in brutes. Neither, of course, do such texts entail a commitment to the *bête machine* doctrine. But such references turn out to be neither extraordinary nor unexpected coming from someone who is committed to the doctrine.

Descartes does sometimes use 'soul'-talk in connection with brutes, but his career-long position is that the souls (so to speak) of brutes fully reduce to corporeality. In a 1637 letter, Descartes writes (of his *Discourse* 5 account):

'animals see just as we do, i.e. being aware,' Descartes clarifies that, in his view, brutes 'do not see as we do when we are aware that we see, but only as we do when our mind is elsewhere. In such a case the images of external objects are depicted on our retinas, and perhaps the impressions they make in the optic nerves cause our limbs to make various movements, although we are quite unaware of them. In such a case we too move just like automatons' (CSMK 61-2, AT 1:413-14).

I thought I had clearly explained, that the souls of animals are nothing but [*nihil aliud*] their blood, the blood which is turned into spirits by the warmth of the heart and travels through the arteries to the brain and from it to the nerves and muscles. (CSMK 62, AT 1:414; cf. AT 1:414f and 4:64f)

In the Sixth Replies he adds: 'I accept that the brutes have what is commonly called "life", and a corporeal soul [*anima corporea*] and organic sensation [*sensus organicus*]' (CSM 2:288, AT 7:426). To More, he clarifies that what 'can be called the corporeal soul [*anima corporea*]' is 'purely mechanical and corporeal, and depends solely on the force of the [animal] spirits and the structure of our organs' (CSMK 365, AT 5:276).

Why then does he speak of *souls* if he means to refer to corporeal processes? In part, he appears to be motivated by theological considerations.⁵⁹ Moreover, in the context of attempting to persuade a scholastic audience (who suppose that *souls* provide the explanatory principles of all the behavior of brutes), Descartes may mean to convey that the only explanatory principle that is needed — the only *soul*, if you will — is purely mechanical.

Doctrinal considerations prevent Descartes from allowing sensory awareness in brutes without also extending substance dualism to them: 'were I to concede that they [brutes] have thought,' writes Descartes, it would then follow that 'in other animals, too, the mind is distinct from the body' (CSM 2:287, AT 7:425-6); 'if they thought as we do, they would have an immortal soul like us' (CSMK 304, AT 4:576). It might, then, seem warranted to conclude that brutes *do* have souls. As emerges from §IV.2, however, Descartes's invocation of parsimony suggests that he thinks there are prohibitive ontological costs in allowing for explanatorily superfluous substances.

2. The strategy of subliminal parsimony

Clear texts establish that Descartes means to eliminate souls (= incorporeal entities) in brutes by means of the principle of parsimony, and, moreover, that he intends the elimination to be subsumed under a more general elimination (by parsimony) of scholastic *forms and qualities*.⁶⁰ Neither point has received due attention in the secondary literature.⁶¹ According to scholastic doctrine, sensitive souls (the locus of sensory

59 Cf. the 1637 letter (of October 3) to Plempius for Fromondus (AT 1:414f).

60 The one exception is the human soul, which Descartes regards as the substantial form of a human (cf. P. Hoffman, 'The Unity of Descartes' Man,' *Philosophical Review* 95 [1986] 339-70) and which he does not attempt to eliminate.

apprehension among brutes) are the substantial forms of brutes. Such forms are thus eliminated by parsimony only if the souls of brutes are. The problem of consciousness in brutes is — as Descartes construes it — a *special case* of the more general problem of forms and qualities. In an early letter (1630), Descartes writes:

As for animals' souls and other [autres] forms and qualities, do not worry about what happens to them. I am about to explain all this in my treatise. (CSMK 26, AT 1:154)

If, as I contend, Descartes intends to eliminate consciousness in brutes by means of the principle of parsimony, why does he not straightforwardly say as much in *Discourse 5*?

As already alluded to, Descartes's procedure is tailor made with the intention to win over a scholastic audience. I call the strategy he pursues *subliminal parsimony*, a strategy that involves the following essentials: (a) ignore forms and qualities in one's philosophical explanations, to avoid overt rejection of scholastic doctrine; (b) emphasize the explanatory power and simplicity of the mechanical hypothesis; if asked about forms, (c) assert that the existence of such entities (as forms and qualities) is by no means in question, but only their explanatory utility. Descartes looks to hold that the effect of this strategy is that such entities will eventually come to be regarded as explanatorily vestigial and thus eliminated in due course.

A variety of textual evidence suggests this. In a letter to Mersenne (28 Jan 1641), Descartes comments briefly on his motivation for not explicitly addressing forms and qualities in the *Meditations*:

I hope that readers will gradually get used to my principles, and recognize their truth, before they notice that they destroy the principles of Aristotle. (CSMK 173, AT 3:298)

While coaching his student Regius, who was in trouble with scholastic authorities for teaching Cartesian doctrine, Descartes elaborates on the strategy of subliminal parsimony in a letter (January 1642):

61 Garber broaches the subject of parsimony, observing that the case against animal souls is *analogous* to the case against substantial forms. See D. Garber, *Descartes' Metaphysical Physics* (Chicago: University of Chicago Press 1992). My claim is much stronger. Garber indeed worries that 'Descartes may not have explicitly made the connection' between parsimony and the *bête machine* doctrine (115f). M. Rozemond, *Descartes' Dualism* (Cambridge, MA: Harvard University Press 1998), ch. 4, offers a superb discussion of Descartes on parsimony, but she does not make the link to the *bête machine* doctrine.

Why did you need to reject openly substantial forms and real qualities? Do you not remember that on page 164 of my *Meteorology*, I said quite expressly that I did not at all reject them or deny them, but simply found them unnecessary in setting out my explanations? If you had taken this course, everybody in your audience would have rejected them as soon as they saw they were useless, and in the mean time you would not have become so unpopular with your colleagues. (CSMK 205, AT 3:492)

The relevant passage of the *Meteorology* reads:

But to keep the peace with the [scholastic] philosophers, I have no wish to deny any further items which they may imagine in bodies over and above what I have described, such as their "substantial forms", their "real qualities", and so on. It simply seems to me that my arguments will be all the more acceptable in so far as I can make them depend on fewer things. (CSM 1:187, AT 6:239; cf. AT 6:85, 7:248-9, and 11:25-6)

As the letter to Regius continues, Descartes urges Regius to write an 'open letter' in reply to Voetius.⁶² At this juncture, the coaching proceeds in a very concrete fashion: Descartes goes on, as he says, to 'sketch out the reply in the form I would think it ought to take' (CSMK 206, AT 3:494) — a sketch that includes the following:

I fully agree with the view of the learned Rector that those "harmless entities" called substantial forms and real qualities should not be rashly expelled from their ancient territory. Indeed, up to now we have certainly not rejected them absolutely; we merely claim that we do not need them in order to explain the causes of natural things. (CSMK 207, AT 3:500)

As I read *Discourse 5*, Descartes is employing this strategy of subliminal parsimony in his treatment of sensory awareness in brutes. In his unpublished *Treatise on Man*, in the precise context of arguing that vegetative and sensitive souls are unnecessary to explain the behavior of the hypothetical, man-machine, Descartes *does explicitly* invoke a version of the principle of parsimony: 'nature always acts by the simplest and easiest means' (CSM 1:108, AT 11:201). Yet, in the *Discourse* (a later work that is published), Descartes does not expressly invoke the principle of parsimony. Instead, his express conclusion is that forms and qualities are explanatorily *unnecessary*. He does of course explicitly conclude that brutes have no reason, but scholastics already accept that doctrine. In a 1637 letter, in response to criticism of his *Discourse 5* argument, Descartes writes that he withheld many lines of support for

62 Voetius (the Rector of the University of Utrecht) had tried to remove Regius from his Chair at the University.

his position 'partly for fear of seeming to want to ridicule received scholastic opinions' (CSMK 63, AT 1:415-16).

On my reading, we may reconstruct the final (implicit) steps of Descartes's argument as follows (where *souls* are understood not as reducible to corporeality, but as incorporeal entities):

- (3) Entities that are explanatorily superfluous (e.g. souls and other substantial forms) are to be eliminated.
- (4) The souls of brutes (and thus all consciousness in brutes) are to be eliminated. [(1),(2),(3)]

From (1) and (2), it follows that any form of consciousness in brutes — whether sensory or reasoned₁ — is explanatorily superfluous. Premise (3) warrants the elimination of explanatorily superfluous such states. Even this reconstruction does not yet fully represent the philosophical considerations moving Descartes. I have formulated line (3) in terms of explanatory superfluity, since line (1) entails no more than the relatively weak claim that sensory consciousness is explanatorily superfluous. Descartes's *own* view (never mind the official, published view, one that is politically more palatable) is that sensory consciousness is not only explanatorily superfluous, but that it is *causally idle* — just as is reason₁ — in the production of the behavior of brutes. As such, the version of the principle of parsimony required by his *own* position is the weaker — where formulated in terms of causal idleness.

3. An objection to Descartes's employment of parsimony

An interesting line of objection remains. Assume (as Descartes contends, on my reading) that the principle of parsimony warrants the elimination of causally idle sensory consciousness in brutes. Since Descartes's procedure rests on analogy, the principle of parsimony should warrant an analogous elimination of causally idle sensory consciousness in humans. But Descartes rejects the latter elimination. Evidently, the principle of parsimony does not warrant the elimination of causally idle consciousness in humans. To avoid absurdity, then, it seems we must reject the initial assumption as to what the principle warrants in the case of brutes.

The objection deserves careful treatment, though space does not permit such treatment here. Though I believe that a correct understanding of Descartes's theory of agency provides him with a plausible response to the objection, a reconstruction of the theory will have to wait for another day.

V Conclusion

Measured by contemporary philosophical standards (or fashions, at any rate), I have been very generous in the assumptions here granted to Descartes. Notably, I have granted his rendition of the distinction between reasoned (= reasoned₁) and purely mechanical causes, a rendition that is parasitic on his substantial distinction between mind and body. I have focused largely on *Discourse 5*, while making liberal use of his larger corpus of texts and doctrines. I have made no effort, however, to characterize the myriad (and often conflicting) accounts of later Cartesians,⁶³ though I have given special attention to the views of his immediate, intended audience — especially scholastics, and Montaigne-influenced readers. It is in the context of these assumptions and methods that I have reconstructed Descartes's case for the *bête machine* doctrine. I have argued that his version of the doctrine stems from an analogy based on, first, the explanatory power of mechanical physiology; second, the unmaskability of all apparently rational behavior among brutes; and third, the principle of parsimony. On the first two bases rests the thesis that neither sensory consciousness, nor reason₁, is necessary to explain the behavior of brutes (indeed, that reason₁ is idle in the production of their behavior). On the third basis rests the elimination of all explanatorily unnecessary entities, with the result that brutes are none other than machines. As I have argued, Descartes's case is not without defect, but it is much more formidable than standard interpretations would have it.

Received: August, 1999

Revised: July, 2000

Revised: April, 2001

63 L.C. Rosenfield does, in her very useful *From Beast-Man to Man-Machine* (New York: Oxford 1940).