A Selection from an Unfinished Work of Metaphysics

Abraham Wine

First published 2 April 2026

Published on www.abrahamwine.com

On Knowledge

Introduction

This chapter establishes fundamental terms and doctrines that are used throughout this work. Some of the content of this chapter establishes the use of language with definitions and sometimes compares it to alternate uses. Some content restates existing doctrines to refresh or instruct the reader and to illustrate the conventions in the use of language in this work. Some doctrines take sides in existing metaphysical debates, though this is not the primary focus of the discussion. Doctrines serving these sorts of foundational purposes also appear throughout the remainder of this work with less frequency than in this chapter.

This chapter gives a metaphysical reconstruction of the origin of knowledge. It thus establishes this work as a positive endeavour to contribute to knowledge in some way, rather than a work of pure skepticism.

All of the doctrines of this chapter apply to human beings, and only some of them apply to other automata. The doctrines of this chapter are thus relatively a posteriori and epistemological, because they are modelled on how human beings have discovered human metaphysics. For other automata, gen-

erally, but not always, the more sophisticated the automaton, the more doctrines of this chapter there are that apply to it. A few of the doctrines apply to all automata in general, though what these are depends on what one takes to be the definition of an automaton. This definition is given and defended in the chapter On Representation.

On the Practical Foundation of Knowledge

The practical foundation of knowledge is experience. In the course of development, an individual learns how to live before they learn how to philosophize. Thus, experience is practically prior to any abstract logical foundation of knowledge.

The only possible exception to this would be a being that was configured to begin learning through abstract reasoning from first principles. Such a being might be God or a godlike being that was purely rational and that had no need for practical experience, or, on the other hand, a contrived computer simulation or other contrived sort of being that learned practically through reasoning from first principles. However, neither of these cases is relevant to the present discussion, which is mostly intended to apply to human beings, and more generally to beings like humans that are neither godlike nor experimental peculiarities, but that must live according to the practical exigencies of real life before they can philosophize.

The order of practical learning is chronological. Chronological order is logical to some extent, insofar as time is bound by the logical structure of causation. However, this logic applies to the causal structure of the development of the being, rather than necessarily to the logical structure of any doctrines formulated by the being. Thus it is that in the practical course of learning, most individuals learn many doctrines before they learn the doctrines on which they are logically dependent, generally because the dependent doctrines are more useful or easier to learn.

However, once an individual learns how to live through practical experience, they may retrospectively give a philosophical description of the foundations of knowledge that abstracts from the particular experiences that taught them how to think. The intended order of the retrospective philosophical account of the foundations of knowledge is logical. And when one tries to de-

termine the logical foundation of all knowledge, one is ultimately led to consciousness. The remainder of this chapter gives a brief description of the retrospective philosophical account of the foundations of knowledge starting with consciousness. The discussion is phrased dogmatically and definitionally, though its origin is epistemological.

On Consciousness

Consciousness is subjective existence.

The doctrine that consciousness exists is not an assumption. The celebrated argument of René Descartes *I think, therefore I am* demonstrates its existence. For if one doubts whether one exists, then one exists. Note that this argument is a subjective demonstration from the perspective of the reasoner that they exist, not an objective demonstration that a reasoner must exist in the first place.

Consciousness may also be called *experience* or *phenomenal consciousness*. In certain technical discussions in this work, consciousness may also be called *mental innate intuition* to enable certain linguistic parallels that facilitate the formulation of some of the doctrines of this work. In discussions outside these technical contexts, the shorter and more common terms *consciousness* and *experience* are more appropriate.

A mental phenomenon is an object of consciousness.

A mind is a maximal collection of mental phenomena that are objects of the same consciousness.

A mind may also be called *a consciousness*. According to the use of language of this work, a mind consists only of the mental innate intuition of mental phenomena, and any non-phenomenal correlates or substrates of phenomenal consciousness are not part of what is called the mind. Other metaphysical authors may use the term *mind* to encompass things other than merely phenomenal consciousness. The terms *mind* and *consciousness* are used in this work approximately interchangeably, but a few differences in their shades of meaning prevent one from only using only one term without using the other, even in the context of a work of metaphysics. In modern philosophical discourse, it is more common to use the term *consciousness* than the term *mind* to refer to phenomenal consciousness specifically. However, the term *mind* paral-

lels the adjective *mental*, for which the adjective *conscious* is not an exact substitute. In particular, it parallels the term *mental innate intuition*, which is useful for various purposes. For example, the term *intuition* allows the transitive verb *intuit*. An analogous transitive verb cannot be conveniently formed from the term *consciousness*. The term *mind* is also more common when speaking of the duality between the mind and the body, as in the terms *mind–body problem* and *mind–body dualism*, and this duality is a main focus of this work, and the varying interpretations of the particular elements invoked in dualism form some of the core doctrines of this work. The term *mind* is also shorter, and less associated with tasteless popular science or academic discussions, and more associated with classic philosophers. However, *consciousness* is the more unambiguous term. Both terms are used in this work.

On the Types of Mental Phenomena

A sensory perception is a mental phenomenon that is intuited passively.

A thought is a mental phenomenon that is intuited actively.

Volition is thought that appears to direct outer behaviour.

Imagination is the active recombination of mental phenomena.

A dream is a passive recombination of mental phenomena during sleep.

What it means for a mental phenomenon to be intuited passively rather than actively is that it appears as though it is an input from without rather than a product from within. This is not always a well-defined division, even in human beings. In human beings, mental phenomena that are intuited actively are less strong or lively than those that are intuited passively. However, it is conceivable that there could exist a being for which the mental phenomena that were intuited actively were more lively than those that were intuited passively. Thus, the more general division, which applies to minds that are not necessarily human, is that mental phenomena may be divided into mental input and mental output. Mental input is intuited passively, and mental output is intuited actively. But it is even conceivable that there could exist a being, the mental phenomena of which could not be categorized even thus.

On Naïve Realism

Naïve realism is the doctrine that mental phenomena are the external objects that they appear to represent.

Naïve realism is false. It may be refuted by appealing to various scenarios in which the senses are deceptive or imperfect. In such scenarios, mental phenomena as they appear differ from what reason obliges one to believe is the objective state of external reality. Thus, mental phenomena are not the objective state of external reality, but at best imperfect representations of it.

On Skepticism

Skepticism is doubt concerning the possibility of knowledge.

The doctrine that the mind exists is outside the scope of rational skepticism, because the argument *I think, therefore I am* cannot be rationally doubted. For if one doubts it, then one exists.

And beyond the fact that one is, one may also be certain that one is as one is. For example, if some mental phenomenon is present to the mind, the mind has immediate knowledge that the mental phenomenon exists thus. Perhaps one could doubt whether that mental phenomenon accurately represents some other thing in the way one believes it does. But since the essence of mental phenomena themselves is appearance, one cannot doubt that mental phenomena are as they appear. Thus, the fact that mental phenomena are as they appear is also outside the scope of rational skepticism.

However, once the mind endeavours to move outside the scope of immediate experience, any complete certainty disappears, and skepticism has more power. As for whether sensory perceptions accurately represent matter, whether memories accurately represent the past, whether beliefs are accurate or in error, this cannot be known for certain. Everyone has experienced situations in which their mental phenomena have not represented things as they thought they did. Thus, practical skepticism is a justified wariness about the soundness of representation. However, once one learns to avoid error, one may attain a reasonable level of confidence about certain beliefs, especially those that are mundane rather than subtle. Thus, reasonably justified beliefs are outside the scope of practical skepticism.

In addition to practical skepticism, there is philosophical skepticism, which is more difficult to surmount in principle, but less practically significant. Philosophical skepticism doubts even beliefs that practical skepticism admits to be reasonably justified. As an extreme case, philosophical skepticism may doubt the existence of anything at all other than the mind at the instant that it now exists. For beyond this, there is no certainty. For it is possible that nothing exists or has ever existed beyond the mind at the exact instant that it now exists thus. Such a hypothetical absurd but possible state of affairs is an example of a skeptical scenario. There are many possible skeptical scenarios, and it is impossible to prove conclusively that reality is not any one of them. Thus, philosophical skepticism cannot be conclusively refuted. However, it is not a serious impediment to philosophizing. To continue in the pursuit of knowledge, all the philosopher needs to do is acknowledge that it is possible that reality is a skeptical scenario, and proceed with philosophizing assuming that it is not.

It would not be within the scope of this chapter to discuss philosophical skepticism in more detail. It is discussed later in this work. For the present, let the above remarks suffice to establish that this work does not advocate philosophical skepticism. Thus, suppose that it is possible for mental phenomena to represent things.

On Representation

Mental intuition through representation is the process by which the mind mentally innately intuits a mental phenomenon in such circumstances that the mental phenomenon represents something.

A mental phenomenon is a mental representational state.

What a mental phenomenon represents, and whether it represents anything, depends not only on the mental phenomenon itself but also on the circumstances in which it is innately intuited. In general, the relevant circumstances may be all of reality, though it is often not necessary to take into account all of reality to determine what a mental phenomenon represents.

The distinctions between representation, belief, and knowledge are not clear-cut. Philosophers have debated these distinctions, and they will likely

continue to debate them. It is likely that there is no fundamental binary distinction between any of them.

The essence of a thing is what it is for it to be.

The structure of a thing is all that may be intuited through representation about it.

The mind innately intuits the essence of its mental phenomena. The essence of the mind is to intuit innately. The essence of mental phenomena is to be innately intuited. The essence of innate intuition is appearance.

The quality of a mental phenomenon is what it is like to mentally innately intuit it.

The essence of a mental phenomenon is qualitative.

This may also be described using the term *qualia*. In this work, the term *quality* is generally preferred over the term *qualia*, because when writing English prose, it is generally in good taste to use English words rather than loanwords. However, the term *qualia* is not a misuse of language.

Some parts of essence cannot be intuited through representation. Thus, structure is a proper subset of essence.

Semiotics is the philosophical study of representation, especially the circumstances under which it occurs.

This work passes over detailed semiotic questions, as they are at too fine-grained a level of investigation to be relevant to most of the questions discussed here. This work assumes that some semiotic system has been fixed, or could in principle be fixed, according to which it is determined how the circumstances determine what a mental phenomenon or a material representational substrate represents, if it represents anything. Thus, the semiotic system of this work abstracts from any theory of what a particular mental phenomenon with such-and-such characteristics would represent in any particular set of circumstances with such-and-such characteristics, except in a few cases where these characteristics are relevant to the abstract structure of the discussion.

On Being and Knowing

Being is that by virtue of which something is.

Knowing is a particular kind of representation typically characterized by being justified and accurate.

Ontology is the branch of metaphysics concerned with being.

Epistemology is the branch of metaphysics concerned with knowing.

Proper ontic differences are differences in essence that are causally indistinguishable.

Proper ontology is the philosophical study of proper ontic differences.

The terms *ontic* and *ontological* have distinct meanings. The term *ontic* refers to being itself, whereas the term *ontological* refers to the study of being. Thus, an ontological distinction represents, or at least purports to represent, an ontic difference. Similarly, the term *epistemic* refers to knowing itself, whereas the term *epistemological* refers to the study of knowing. Thus, an epistemological distinction represents, or at least purports to represent, an epistemic difference.

All knowing is a kind of being. For all knowing must be instantiated in the activity of some existing thing. In other words, all knowing is ontic.

Not all being is knowing. For there are kinds of being, such as that of inanimate objects, that do not involve representation, let alone knowing. In other words, not all being is epistemic. Or it is only epistemic in the sense of not knowing being a degenerate epistemic state.

All ontology is knowing or at least an attempt at knowing. In other words, all ontology is epistemic. Ontology is thus also ontic. Not all knowing is a kind of ontology, because there is much knowing that is not concerned with the philosophical study of being. Similarly, not all being is a kind of ontology.

All epistemology is being and knowing. Thus, all epistemology is ontic and epistemic. Not all knowing is a kind of epistemology, because there is much knowing that is not concerned with the philosophical study of knowing. Similarly, not all being is a kind of epistemology. It would be reasonable to say that in a strict sense all epistemology is a kind of ontology, insofar as all epistemology is philosophical and all knowing is a kind of being. Thus, all epistemology, the philosophical study of knowing, is also the philosophical study of being. However, when one discusses epistemology, one usually discusses it in the capacity of being epistemology rather than merely being ontology. Not all ontology is a kind of epistemology, because not all being is knowing.

The difference in meaning between the terms *epistemic* and *epistemological* is less clear-cut than that between the terms *ontic* and *ontological*, be-

cause being contains knowing, and knowing contains both ontology and epistemology. The inclusion of epistemology in knowing is closer than the inclusion of ontology in being. This is especially so in the proper, rather than the general, senses of the terms *ontic* and *ontological*. Proper ontic and ontological differences have no intersection. The former refer to differences in essence that are causally indistinguishable, whereas the latter refer to differences in philosophical doctrines. And the differences in philosophical doctrines may be entirely characterized causally. This means that, especially in the proper sense of the former two terms, it is generally unambiguous whether to use the term *ontic* or *ontological*, but it is sometimes ambiguous whether to use the term *epistemic* or *epistemological*. The former question is one of accuracy, whereas the latter question is sometimes merely one of emphasis, in particular, whether to emphasize how a particular instance of knowing philosophically about knowledge is knowing in the capacity of knowing in general, or to emphasize how it is knowing in the capacity of knowing philosophically about knowledge.

In this work, when speaking of ontic differences and ontology, the discussion is usually in the proper, rather than the general sense of these terms.

In this work, when claims are made about things according to their essence, instead of using the terms *in essence* or *essentially*, which have imprecise popular meanings that could cause confusion, the term *ontically* is used. The term *ontic* is used instead of the term *essential* for a similar reason. In this work, the terms *ontic* and *ontically* can refer to mental phenomena as substance, as well as referring to matter.

On Matter

Matter is that which sensory perception typically represents.

The doctrine that matter exists, or that there is any organization to external reality at all, may be doubted somewhat with skepticism, but this is not a serious impediment to metaphysical investigations beyond forcing one to acknowledge the possibility of a skeptical scenario.

The doctrine that matter exists may be more plausibly denied with idealism. Idealism is considered in more detail later in this work. Discussing ideal-

ism in more detail here would only obscure the intended train of thought, so for the present, assume that matter exists.

The material universe is the collection of all matter.

The physical state of matter is its structure.

The body is the continuous collection of matter correlated with the mind.

A substance is matter, mental phenomena, or any other kind of existing thing.

The mind cannot innately intuit matter. The mind can intuit the structure, but not the essence, of matter.

The terms *material* and *physical* have related meanings, but different shades of meaning. The term *physical* denotes that something relates to that which underlies the regularity of sensory perception, irrespective of whether or not this is taken to be matter. For example, the predictions of the physical sciences are unchanged even if idealism is true.

The body is an automaton. Automata are defined and analyzed in the chapter On Representation. Other bodies are correlated with other minds similarly to how the body is correlated with the mind. These bodies are automata. Other minds have mental phenomena similarly to how the mind has mental phenomena. The mind cannot innately intuit the mental phenomena of other minds. The mind can intuit through representation the structure, but not the essence, of the mental phenomena of other minds.

On Space and Time

A memory is a thought that appears to be an innate copy of other mental phenomena.

Subjective space is the outer structure of mental phenomena.

Subjective time is the inner structure of memory.

The subjective present is that in which mental phenomena appear to the mind, considered with respect to subjective time.

The past is that which memory typically represents.

The future is that in which memory could represent the present.

Objective space is that which is imperfectly represented by subjective space.

Objective time is that which is imperfectly represented by subjective time.

Objective space-time is the global manifold of which objective space and objective time are local projections.

An objective event is a point in objective space-time.

The future follows from the present. Depending on the circumstances, memory can represent either the past mental phenomena of which the memory is a copy, or the matter or mental phenomena that the copied mental phenomenon represented, if it represented anything, or the memory could represent something else entirely. There are regularities in how the future follows from the present. The way that the future follows from the present may be inferred by studying the past.

Subjective space and subjective time locally and imperfectly represent objective space-time. Subjective space has three dimensions. Subjective time has one dimension. Objective space-time is a four-dimensional manifold.

Matter exists in objective space-time.

A thing is a set of events.

A thing is not necessarily a continuous substance that endures over time. A thing is a subset of all that exists.

On Dualism

Introduction

Pure philosophical skepticism having been rejected in the last chapter, this chapter further specifies the dogmatic stance of this work as dualism, the doctrine that minds and matter exist, rather than materialism, idealism, neutral monism, or some other alternative. For though the previous chapter discusses the mind and matter, and is phrased in a way that suggests they are distinct, that chapter does not positively claim that they are distinct, and one could coherently interpret its discussion assuming that the mind is material. In philosophy, one should not assume that the mind is non-material without careful justification, for at first appearance it is conceivable that one is really talking about the same thing in different ways when one speaks about the mind and matter. Indeed, many philosophers are of this opinion. However, it is more widely believed that the mind and matter are distinct.

This chapter endeavours to show that the mind and matter are distinct. It thus establishes philosophical dualism as the main dogmatic stance of this work. It does this by arguing that it is possible that the body could exist exactly as it does in reality without consciousness existing. That is to say, it is

conceivable that there could exist aphenomenal copies, which are automata that behave identically to the automata in reality but which have no mental phenomena correlated with them.

Once dualism is established, the remaining doctrines of this work are advanced from its perspective.

On Dualism

Dualism is the doctrine that the mind exists and that matter exists, and that the two are distinct.

Substance dualism is dualism claiming that the mind and matter are different substances.

Property dualism is dualism claiming that the mind and matter are different aspects of one substance.

Many of the discussions of this work involving dualism apply as well to substance dualism as they do to property dualism. Indeed, it is argued later in this work that the distinction between substance dualism and property dualism may be chimeric, that is, a distinction of words that does not reflect any underlying difference in reality.

This work advocates dualism.

On Materialism

Materialism is the doctrine that everything is material.

Materialism may also be called *physicalism*. In some contexts, the term *physicalism* may have a different meaning from the term *materialism*. Indeed, in some contexts, physicalism may be consistent with dualism rather than materialism as defined here. For example, physicalism may be interpreted as an affirmation of epiphenomenalistic dualism rather than interactionistic dualism, or of determinism rather than indeterminism. The epiphenomenalistic dualism advocated in this work is in some ways closer to materialism than to interactionistic dualism. Epiphenomenalism is discussed in a later chapter. In this work, the term *physicalism* is avoided in favour of more precise terms such as *materialism*, *epiphenomenalism*, and *determinism*.

Absolute materialism claims that consciousness does not exist.

Reductive materialism grants that consciousness exists, but claims that it is material.

In non-metaphysical contexts, the term *materialism* generally has a moral-philosophical connotation relating to the preoccupation with wealth or other possessions. This meaning is not intended in this work.

This work does not advocate materialism. However, the focus of this work is not on arguing against materialism, but on dialectically assimilating it with dualism using interpretation.

On Idealism

Idealism is the doctrine that the mind exists and that matter does not exist, and that appearances are not illusory.

Idealism differs from skepticism in that it believes that there is an underlying organization to reality. It differs from dualism in that it does not believe this organization is due to matter. The typical formulation of idealism claims that God directly organizes the regularity of sensory perception without creating any matter to do so. This form of idealism is due to George Berkeley, and is generally the one invoked when this work discusses idealism. However, one could be an idealist without believing in God if one believed something else was responsible for the regularity in sensory perception, such as a non-anthropomorphic and non-material law of causation taking the place of God, or a computer simulation in another level of reality.

It would distract from the intended train of thought of this work to discuss idealism in detail here. Idealism is discussed later in this work. For the present, suffice it to say that idealism is very similar to dualism in most respects, and that most of the doctrines of this work, though they superficially appear to reference matter, could be rephrased without much difficulty to fit into an idealistic framework that avoided invoking the existence of matter.

This work does not advocate idealism. However, it acknowledges it as a viable alternative to dualism.

On Monism

Monism is the doctrine that everything is one substance.

Property dualism, materialism, and idealism are all kinds of monism. Of them, materialism is the most monistic and property dualism is the least. Materialism is more monistic than idealism because idealism posits the existence of God, and God is a very different kind of mind from the others in the universe, and under this view of reality there is thus a duality between God and the other minds of the universe. Idealism is more monistic than property dualism because God is more like a mind than matter is, even if the mind and matter are interpreted as different aspects of the same substance, because such a resemblance is part of what makes God more than merely matter.

On Neutral Monism

Neutral monism is the doctrine that mental phenomena and matter are one substance, and that phenomenal properties are the internal properties of this substance, and that physical properties are its extrinsic properties.

The panpsychistic version of this posits more specifically that fundamental microphysical entities have phenomenal properties, and that phenomenal properties are the substance that underlies the causal structure of reality, and that this underlying means these phenomenal properties are non-epiphenomenal.

Neutral monism may be interpreted as a form of property dualism, but its proponents tend to frame it with a more monistic slant than property dualism.

Neutral monism may also be called *Russellian monism* or *the identity theory*. The term *Russellian monism* is an homage to Bertrand Russell, who advocated a form of neutral monism. However, Russell himself called this doctrine *neutral monism* in his work *The Analysis of Matter*, and this term is more descriptive.

Reductive materialism becomes neutral monism if it is taken to its logical conclusion. For the materialistic doctrine that consciousness is material, if taken to its logical conclusion, can only be rehabilitated as the doctrine that matter, either some matter or all matter, is phenomenal, once one understands that one cannot deny the existence of the phenomenal. Thus, neutral monism is a kind of materialism, and the only kind that is consistent with the facts of experience.

At first appearance, neutral monism may appear to have some ontological or aesthetic superiority over dualism, especially epiphenomenalistic dualism, because neutral monism is less arbitrary. The biggest problem with neutral monism is the combination problem. This is the problem of how the mental phenomena of the microphysical constituents of reality, if these even have mental phenomena, combine into the complex minds of automata. Later in this work, it is argued that the combination problem is fatal for the main ontological benefits that neutral monism might seem to have over dualism, and that dualism should therefore be preferred. However, to enter into this argument in detail in this place would only obscure the intended train of thought. Thus, for the present, suffice it to say that the mental phenomena of automata are likely distinct from the mental phenomena of any of their constituent parts, and that there is therefore little sense in saying that the former or the latter mental phenomena constitute the internal properties of matter.

This work does not advocate any form of monism. However, it acknowledges some forms of monism as viable alternatives to dualism.

On the Monistic Rehabilitation of Materialism

All arguments for dualism have a fatal problem. Ultimately, one cannot conclusively refute materialism, because one form of materialism, neutral monism, is viable. It is possible that consciousness is material if the material universe is constituted by phenomenal properties. Indeed, if consciousness is material, then it is necessary that the material universe is constituted by phenomenal properties, because it is certain that phenomenal properties exist. Thus, neutral monism is the only viable rehabilitation of materialism. Thus, all the above arguments that consciousness can be separated from the body, or that phenomenal properties are distinct from physical properties, are doomed to be invalid in this special ontological case. However, ignoring this case for now, the argument for dualism may be put forward as follows.

On the Distinction between the Mind and Matter

The most important metaphysical debate between the foundational systems of metaphysics described above is not over whether the mind exists. For all ra-

tional thinkers, even skeptics and materialists, will grant this. Nor is the debate over whether matter exists. For most philosophers do not seriously believe in skepticism or idealism beyond acknowledging them as possibilities. The main debate is between dualism and reductive materialism. Both grant that the mind exists, and both grant that matter exists, but dualists claim that the mind is non-material, and materialists claim that the mind is material.

The argument *I think, therefore I am* shows that consciousness exists. Thus, anyone who wishes to argue for materialism cannot outright claim that consciousness does not exist. And indeed, most authors who call themselves materialists do not do so. Instead, they try to show that the argument *I think, therefore I am* does not apply to the thing the dualist claims it applies to.

It is in this distinction between dualism and materialism that the term *phenomenal consciousness* is useful. The dualist asserts that consciousness is phenomenal, and the materialist asserts that consciousness is material and that what appears to be phenomenal is really material. Or a materialist who wished to slant the use of words another way could assert that phenomenal consciousness exists, but that phenomenal consciousness is material.

On the Common Belief in Dualism

Dualism is more widely believed than materialism, both among philosophers and non-philosophers.

If a non-philosopher is asked whether they believe that they are more than matter, they will usually say that they are something more. If asked why, they may give various justifications for this belief.

The dualist may say that their experience cannot be matter, because this is intuitively obvious, because subjective existence is something other than the physical behaviour of matter. However, this is begging the question.

If they are religious, they may cite faith in the existence of God as a refutation of materialism. However, faith in the existence of God, though personally meaningful, is not sufficient for rigorous philosophical reasoning, because it is not certain whether or not God exists. And even belief in the existence of God based on a well-founded argument, such as the presence of morally significant phenomenal consciousness in the universe, is not sufficient for rigorous philosophical argumentation in this context, because one would thus

take the existence of God as evidence for the existence of phenomenal consciousness, and and the existence of phenomenal consciousness as evidence for the existence of God, which is a circular argument. And even if one has some argument for the existence of God that is independent of the existence of consciousness, this is still not sufficient to establish dualism. For this argument in itself would only establish that God exists, and it would not establish whether God created a universe in which materialism holds for all automata. For it is conceivable that it could have pleased God to do so. Thus, such an argument for the existence of God may be a refutation of atheistic materialism, so to speak, but not of theistic materialism, or of materialism that is agnostic about the existence of God.

The dualist may appeal more vaguely to spirituality or some other intuitive feeling that they are more than matter. But this is not adequate for rigorous philosophical argumentation.

None of these common arguments is sufficient to establish dualism. Rigorous philosophical argumentation is required.

On Dualism and the Existence of Consciousness

If a dualist has some acquaintance with philosophy, they may claim that the argument *I think, therefore I am* establishes dualism in favour of materialism, because it demonstrates that the mind exists.

Now, this argument is an important step in establishing dualism. However, in itself it is not sufficient for this purpose, because though it establishes that consciousness exists, it does not in itself show that consciousness is not material. Indeed, the argument leaves the nature of consciousness undetermined beyond the fact that it exists. This is evidenced empirically by the fact that many intelligent philosophers acquainted with that argument nevertheless profess materialism. They grant that consciousness exists, and that this argument establishes that what they call consciousness exists, but they claim that consciousness is not what the dualist thinks it is.

An absolute materialist may claim that there is nothing that it is like to be anything. They could claim this in various ways. They could claim that there is nothing that it is like to be conscious. They could claim that there is no more to consciousness than certain functional or physical properties. They

could claim that consciousness is not phenomenal. They could claim that phenomenal consciousness does not exist. They could claim that phenomenal consciousness exists, but that the essence of the phenomenal is material function. They could even claim that there is something it is like to be conscious, but that what it is like is merely the behaviour of matter, nothing more.

If such a claim is merely verbalistic and does not have a definite meaning, it may be disregarded as a sophism. If such a claim is made meaningfully, it may be refuted by application to the facts of experience. For consciousness is thus as one experiences subjective existence, and if the materialist claims that it is not thus, then they are false. Thus, absolute materialism, which endeavours to show that consciousness is not thus, is false.

More difficult to refute is reductive materialism. Indeed, it is ultimately impossible to disprove reductive materialism, because, in its most rational form without any sophistries, it becomes neutral monism, which is a viable doctrine.

On Early Arguments for Dualism

In ancient and medieval times, even great philosophers did not have a clear conception of the duality between the mind and matter. René Descartes was the first person to give a rigorous definition of dualism and to give rigorous arguments in favour of it. His arguments are edifying and historically significant, but they are of less relevance to the modern debate between dualism and materialism, because the forms of materialism adopted by modern philosophers are sufficiently subtle that they are not refuted by the arguments of Descartes. In his *Meditations on First Philosophy*, Descartes shows that the mind is distinct from the various peripheral parts of the body, such as its limbs, and that mental perceptions of the body are not the body itself, but representations of it. But modern materialists grant this as well. These arguments are only refutations of naïve realism, not of modern materialism. In order to establish dualism and refute materialism, one must show not only that the mind is different from external objects, or the limbs, or the spinal cord, but that the mind is different from the part of the brain with which it is most closely associated. The association between the mind and the brain is described in more detail in the chapter On Correlation, but here suffice it to say for the purpose arguing whether or

not the mind is material, and as most would grant without detailed philosophical investigation, that the mind is more closely associated with some part of the brain than with any other part of the body or of the brain. Few of the arguments of Descartes endeavour to demonstrate this more precise doctrine. Indeed, most would be better described as taking it for granted. However, an argument that applies even to the distinction between the mind and the part of the brain with which it is most closely associated is contained in Meditation 6 of the *Meditations on First Philosophy*. In it, Descartes claims that the mind is indivisible, but that matter is indivisible, and that therefore the mind cannot be matter. This argument is dubious. It seems that Descartes is positing the existence of a simple metaphysical subject distinct from the mental phenomena that it innately intuits. And indeed, if such a subject exists, this argument shows that it is distinct from matter. Whether such a subject exists is discussed in more detail later in this work, but here, suffice it to say that it is dubious whether such a subject exists. In any case, the more important question in the present train of thought is whether mental phenomena themselves, which most would grant are compound, are distinct from matter.

And other great philosophers of the early modern age, though they also discussed dualism and the refutation of naïve realism, did not give arguments against the form of materialism now under consideration, because materialism itself had not yet emerged as a doctrine in its own right. Materialism denying the existence of consciousness or claiming that it was nothing but matter first arose around the end of the nineteenth century, and the debate between dualism and materialism developed in the twentieth century, and continues now in the twenty-first, though now, as previously, dualism tends to be more widely believed.

On Contemporary Arguments for Dualism

One can refute the more irrational aspects of materialism with arguments that are extensions of the earlier arguments to refute naïve realism. Thus, one can show that consciousness is not any part of the body apart from perhaps some particular part or aspect of the brain. One can show that the existence of phenomenal consciousness cannot be denied. However, difficulty arises when the dualist endeavours to show that phenomenal consciousness is not identical

with some part or aspect of the brain, or more generally, of the body of the automaton in question.

A dualist may claim that material objects have certain physical properties, such as mass and charge, which consciousness, or what it is like to be subjectively, does not have. Therefore, consciousness is not material.

A response to this is as follows. The materialist does not necessarily claim that consciousness is one particular material object in the sense of a particle or a collection of particles. Perhaps it is a more abstract aspect of physical structure that is not in one definite position or that does not contain one definite piece of matter. The properties of mass and charge would not apply to such an aspect, but it would still be merely physical.

A dualist may claim that what it is like to be has nothing to do with how something behaves extrinsically. It is conceivable that the body could behave just as it does, for this is merely characterized extrinsically, but where there is nothing going on inside or in addition to the movement of atoms and other microphysical constituents of reality.

A response to this is as follows. Such a dualistic argument is merely begging the question using rhetorical language.

A more rigorous argument for dualism must go beyond rhetoric and first appearances.

On Criteria for Establishing Dualism

Dualism differs from materialism insofar as it claims that the mind is distinct from matter. But it is difficult to argue for or against this claim without some criterion for determining when two things are distinct. And indeed, the definition of distinction forms part of the debate. Philosophers have come up with various ways to precisely discuss these issues. A few of the criteria and definitions they use are as follows.

One can claim two things are distinct if one can conceive of one existing exactly as it exists without the other existing. One can claim that this is only a sufficient condition, or that it is also a necessary condition. One can claim two things are distinct if it is possible that one could exist exactly as it exists without the other existing. One can claim that this is only a sufficient condition, or that it is also a necessary condition. One can claim that the conceivab-

ility criterion is sufficient for the possibility criterion to hold. One can claim that the possibility criterion is sufficient for the conceivability criterion to hold. One can claim that they are equivalent. One can claim, prior to any application in characterizing the distinction between the mind and matter, or even prior to characterizing distinction at all, in general that conceivability implies possibility, or that possibility implies conceivability, or that they are equivalent. One can distinguish between conceivability at first appearance, also called *prima facie* conceivability, and ideal, or clear and distinct, conceivability. One can distinguish between positive conceivability, in which one conceives of something as possible, and negative conceivability, in which one cannot conceive of it as impossible or as involving a contradiction. One can distinguish between primary conceivability, involving the primary intension, or intrinsic signification, of the representational states used in the conception, and secondary conceivability, involving the secondary intension, or contextual signification, of the representational states used in the conception. One can distinguish between primary possibility, involving the primary intension of the representational states used to define the possibility, and secondary possibility, involving the secondary intension of the representational states used to define the possibility. And one could make further distinctions in this discussion.

Thus, there are many fine distinctions involved when one endeavours to argue for or against the claim that the mind is distinct from matter. Contemporary academic philosophical arguments between dualism and materialism generally involve many such fine distinctions and subtle arguments involving the above concepts and others. The separability criterion is of great importance. Practically, most arguments about whether the mind is distinct from matter turn into arguments about whether the mind could be separated from or subtracted from matter in principle. One can use the following doctrines.

Consciousness is explanatorily irrelevant for determining physical behaviour. It is conceivable that the mind could have a different quality while the body remained identical. For any being that has a mind and a body, it is conceivable that there could exist a being that was physically identical to that being but which did not have any consciousness.

All these doctrines are very closely related to dualism, and according to some interpretations are logically equivalent to it. It is hard to demonstrate these doctrines without presupposing or demonstrating dualism, and it is hard to demonstrate dualism without presupposing or demonstrating at least one of

these doctrines, or a doctrine like one of them. All these doctrines appear obvious to most philosophers who consider them, and most do not see any need to demonstrate them. But then again, dualism appears obvious to most philosophers who consider it, and most do not see any need to demonstrate it. But in principle, if one wishes to philosophize rigorously, one should have some rational justification for believing these doctrines other than a practical intuition that they are obvious.

Thus, practically, most arguments for or against dualism ultimately become arguments over whether, for some person or other being with consciousness, here could exist a purely material being that was physically identical to it in all respects, but which, being purely material, did not have phenomenal consciousness. Such a purely material being may be called an aphenomenal copy.

On the Possibility of Purely Material Automata

It is likely that purely material versions of dual automata are naturally impossible. That is, it is likely that in reality, purely material versions of what would ordinarily be dual automata do not exist, nor could they have existed in the past, nor could they exist in the future, nor could they exist with different initial material conditions but the same psychophysical laws. Furthermore, as is argued above, and as is generally agreed by many philosophers, it is nevertheless logically possible, that is, conceivable, that purely material automata could exist, for example, in an alternate universe with different psychophysical laws. For according to the separability criterion for being distinct, if dualism is true, then consciousness can in principle be separated from the body.

Someone might also argue that, depending on how automata are defined, an automaton could be purely material without being an aphenomenal copy, because it would be simple enough that, according to the ordinary laws of nature, it is not dual. The question of whether aphenomenal copies are possible may be expressed as the question of whether it is possible that an automaton that would be dual according to the ordinary laws of nature could be purely material. This formulation avoids the question of whether some simple automata in reality are naturally purely material but not properly speaking aphenomenal copies. However, as this work endeavours to show, since all

automata are dual in reality, a posteriori this formulation amounts to the same thing as the formulation of asking whether purely material automata are possible. Since the expression *the purely material version of what would ordinarily be a dual automaton* is unwieldy, it may be replaced by the term *aphenomenal copy*.

Some philosophers claim that aphenomenal copies are logically impossible, that is, inconceivable. For example, a materialist might claim that consciousness is merely material innate intuition, and that all automata are thus purely material. According to such a claim, since all automata are purely material, none of them are aphenomenal copies, because they are no longer the purely material version of what would ordinarily be dual, because ordinarily they are purely material. This position is mistaken, because it ignores evidence of first-person experience. However, the materialist may nevertheless generate much subtle argumentation to claim that the evidence of first-person experience is illusory.

On Consciousness and Logical Necessity

An objector may claim that aphenomenal copies are only conceivable at first appearance, but they are not conceivable by an ideally rational agent. Just as how someone who is not reasoning properly may conceive at first appearance of a true mathematical theorem being false, but it is not logically possible for the theorem to be false, similarly, there must be some as-yet hidden logical structure to reality that means that aphenomenal copies are logically impossible, though they appear to be logically possible.

A response to this is as follows. This argument is at best an inductive aesthetic suggestion based on an intuitive analogy that aphenomenal copies are not ideally logically possible, because it does not provide any rigorous argument beyond this intuitive analogy to support its claim.

But even supposing as a concession that there were a more substantial argument to support the claim, it will nevertheless necessarily be flawed for the following reason. There is a fundamental difference between the possibility of mathematical theorems and the possibility of matters of fact. Mathematical theorems are not concerned with how things are a posteriori, but with the a priori structure of how things would be if certain a posteriori conditions were

met. Thus, the truth of mathematical theorems is analytic a priori, because at their foundation, they only involve the logical analysis from definitions characterizing how certain matters of fact might be, not any assertion of any matters of fact themselves to which the definitions apply. Whereas for atomic matters of fact, it is conceivable and logically possible that any of them could not be the case, as Ludwig Wittgenstein described. Perhaps if a fact is a compound proposition with a logical structure, and if some other fact related to the first fact is the case, it might be impossible for the first fact not to be the case according to the logical structure of their relation. But the atomic matters of fact themselves are all logically independent. An atomic matter of fact may be an event in space-time, or the value of a wave function at some point in phase space, or something else, dependent on what the fundamental theory of physics turns out to be. Or an atomic matter of fact may relate to mental phenomena. Thus, there is no way that hidden a priori structure could make it impossible for aphenomenal copies to exist in the same way it is impossible for a true mathematical theorem to be false, because the existence of aphenomenal copies may be reduced to a conjunction of atomic matters of fact.

An objector may claim that under the apparent independence of what are typically called atomic matters of fact, there may be some hidden logical structure that means they are really logically connected, and thus not atomic. Thus, perhaps there is only one atomic fact for the whole universe, and all other facts, which appear to be several atomic facts to a limited and imperfect human reasoner, are all theorems that follow from the logical structure of the universe a priori. Or perhaps there is no atomic fact at all, for the existence of even one atomic fact at all, at least as the author has characterized atomic facts, would imply that that fact could not be the case, or perhaps that it could be otherwise in some way other than non-existence. Perhaps all that is the case is necessary a priori, and all appearance of the possibility of an alternate scenario is an illusion.

A response to this is as follows. This is an exotic ontological stance that likely cannot be tested empirically or judged conclusively through a priori reasoning. Also, note that the objector is at best appealing to the conceivability at first appearance of such an exotic ontological scenario, which is self-defeating, or at least hypocritical, given that they are arguing that all possibility at first appearance beyond necessary demonstration is illusory. For they have given no necessary demonstration that such an ontological scenario is the case,

and according to their very position, such a necessary demonstration is the only valid way to establish that such a scenario is possible. For according to their view, the only way to establish that a scenario is possible is to establish that it is necessary a priori.

Furthermore, the truth of such an exotic ontological stance has little to do with dualism beyond objecting to certain ways of using words in typical dualistic arguments. Dualism claims that consciousness is distinct from matter, and in order to do this, it typically uses the criterion that two things are distinct if one could exist without the other. However, if one adopts the exotic ontological stance that nothing could exist other than how it exists in reality, then one may still affirm dualism, but one merely needs to use a different criterion to establish it other than the typical one of the possibility of one existing without the other. Indeed, being somewhat facetious here, one could use the criterion that two things are distinct if one could exist without the other if the exotic ontological stance happened to be false, perhaps due to some unforeseen error the objector had not thought of. More seriously, one could use the criterion that two things are distinct if neither in itself logically entails the other. For though two theorems may be both necessarily true, perhaps because they both follow from true axioms, yet they may still in themselves be logically independent. And if the supposedly necessary theorem of the existence of consciousness is logically independent of the supposedly necessary theorem of the existence of the body, then this could be interpreted to mean that consciousness and the body are distinct. One could then say that the aphenomenal scenario is logically possible, not in the typical sense, but in a weaker sense that means the two theorems are logically independent.

One might show that the two supposed theorems are logically independent as follows. Insofar as it is even possible to reason at all about such an exotic ontological scenario, the way to show the two supposed theorems are logically independent would be very similar to showing that one thing could exist without the other under the more standard ontological framework where alternate scenarios are possible. Thus, this ontological variant becomes merely an aesthetic quibble about how to use words, and in principle any dualistic argument for the logical possibility of an alternate aphenomenal scenario could be translated into an equivalent argument for the logical independence of two necessary theorems, one about the existence of consciousness, one about the existence of the body.

An objector may claim that no two such theorems can be independent, and that from any theorem about reality, all the rest follow. At least, and more relevant to the purpose of this argument, any necessary theorem about the body logically entails any necessary theorem about consciousness.

A response to this is as follows. At this point, the objector is just dogmatically asserting what they are supposedly endeavouring to prove rationally using a circumlocution to obfuscate the unjustified assertion. They are saying the following. Suppose there were a hypothetical universe where the claim the objector is endeavouring to prove is true. In such a universe, the claim the objector is endeavouring to prove is true. But reality is that universe. Therefore, the claim the objector is endeavouring to prove is true. This argument is merely an assertion that they are right, but it is dressed up with a tautological syllogism to make it look less like a plain dogmatic assertion.

An objector may claim that there is nothing that is logically possible but not naturally possible, because things are as they are, and it is not meaningful to think about anything else.

A response to this is as follows. It is true that things are as they are. This is a tautology. However, as some of the thought experiments in this work and other works indicate, it can sometimes be useful and meaningful to consider hypothetical alternatives to what exists in reality. And the objector has not even argued why they believe it is not meaningful to consider alternate possibilities. They have merely dogmatically asserted this claim.

An objector may claim that since physics has succeeded very well so far in explaining nature, if it has not yet succeeded in doing so for consciousness, that should only be interpreted as a temporary hurdle.

A response to this is as follows. Such inductive reasoning is plausible at first appearance, but it is not a necessary argument, and if stronger evidence exists in favour of the opposite stance of dualism, as it does here, the opposite stance should be adopted despite the inductive reasoning giving an intuitive aesthetic suggestion otherwise.

On Consciousness and A Posteriori Necessity

A materialist wishing to take on a logically weaker claim, and thus a philosophically stronger position, may instead acknowledge that aphenomenal cop-

ies are logically possible, that is, conceivable, but claim that they are metaphysically impossible, according to some definition of what is metaphysically possible that is different from the definition of what is logically possible. When contemporary philosophers do this, they generally invoke the notion of primary and secondary intensions. Or they may argue that aphenomenal copies are not possible in some other subtly defined way.

On Skepticism concerning Kinds of Possibility

As the academic philosophers witness in great detail in their publications, for all subtle materialistic argumentation that aphenomenal copies are not possible in some way, much subtle dualistic argumentation may be generated in turn to reaffirm that aphenomenal copies are in fact possible in that way, or that the type of possibility is not properly defined, or some other argument. Such dualistic argumentation may very well be valid. This work at least agrees with its conclusion that aphenomenal copies are possible, because thought experiments involving purely material automata underlie some of its core interpretive doctrines, and if one is conducting thought experiments, one would hope that the scenarios in question are possible in some way. However, it is doubtful whether subtle metaphysical argumentation about the nature of possibility practically supports the claim that aphenomenal copies are possible. For one who did not doubt their possibility in the first place may merely find it tiresome, and one who did doubt their possibility is unlikely to be convinced by a particular syllogism or other argument in a complex edifice of distinctions regarding types of possibility set up by their opponent, when the whole edifice is already in doubt for the much more basic reason that it was set up by their opponent. Furthermore, in metaphysics more than in any other branch of human knowledge, one must always be on guard to avoid falling into chimeric argumentation. The more subtle, fine, and technical distinctions that one makes, the more likely it becomes that one is not talking about anything at all, but merely following certain patterns of using words that happen to be in fashion in the circle of discussion to which one belongs. And it is possible that some of these distinctions of possibility that contemporary philosophers have put forward in their voluminous arguments for and against dualism are chimeric.

On Referring to Purely Material Automata

Dualism, and thus the possibility of purely material automata, having been established, this section elaborates on purely material automata, clarifying how they are defined and referenced in this work.

The thought experiment of purely material automata and the purely material universe is not new. Several authors have already written about these ideas, and they are relatively common subjects of discussion in contemporary metaphysics. Purely material automata are often referred to as *philosophical zombies* or just *zombies*, and the purely material universe is often referred to as the *zombie universe*.

In this work, the term *zombie* is avoided because it is not primarily descriptive, but rather primarily facetious or metaphorical. Such levity is unfitting for a precise metaphysical term. To a layperson, the term *zombie* in this context is even misleading, because it typically refers to an undead being, and may thus deceptively appear to refer to an undead being when such a meaning is not intended. Undead beings would likely have phenomenal consciousness of some kind, and thus not be philosophical zombies, and philosophical zombies would not necessarily be undead beings in the typical sense, though there could exist a philosophical zombie version of an undead being, which did not have phenomenal consciousness. In the opinion of the author, it is an embarrassment to philosophy that such a facetious and misleading term has been accepted as standard terminology for such a fundamental notion as that of an automaton identical to a dual automaton in all respects but lacking phenomenal consciousness directly correlated with it.

Now, the term *purely material automaton* is more serious and descriptive, being simply based on the attributes of the thing described. In contexts where the differences in the shades of meaning of the terms *zombie* and *purely material automaton* are not important, the latter term is a suitable replacement for the former. However, the term *zombie* expresses a different shade of meaning than the term *purely material automaton*. For the term *purely material* means that the automaton in question consists only of matter, whereas the term *zombie* means that the thing in question is an alternate version without phenomenal consciousness, but otherwise identical to a thing that has phenomenal consciousness in reality. A posteriori these two terms may have the same sec-

ondary intensions in most or all contexts, but a priori their primary intensions differ.

Thus, the term *purely material automaton* is not a complete replacement for the term *zombie* in all contexts. But the term *zombie* is flawed. Therefore, it would be beneficial to find some other term that is a suitable replacement for the term *zombie*.

Etymologically, the term *aphenomenal* denotes that a thing lacks phenomena. Metaphysically, the term *aphenomenal copy* may thus have the same meaning as the term *philosophical zombie*. A zombie could also be called *an aphenomenal automaton*, when that of which it is an aphenomenal copy is not at the focus of the discussion. The zombie universe may be called *the aphenomenal universe*. The term *aphenomenal* may thus be used to replace the term *philosophical zombie* in most contexts other than those specifically referencing the latter term as a term rather than using it to denote an aphenomenal automaton.

The term *aphenomenal* has a number of advantages that make it more suitable than the term *zombie*. It is more descriptive, because a philosophical zombie, simply put, is an aphenomenal copy. This term is more grave, not only because the language is more formal, but more deeply because precision in description, being fitted to the task, is more grave than imprecision, which is whimsical. And gravity is more suitable for philosophy than levity.

Thus, this is a term that has the gravity and precision of the term *purely material automaton*, and which has the shade of meaning of the term *philosophical zombie*, in all contexts other than those specifically referencing the latter term as a term.

On Different Shades of Meaning in Referencing Purely Material Automata

Some of the contexts where the terms *purely material automaton* and *aphenomenal copy* have different shades of meaning are as follows.

One context where the terms have different shades of meaning is when the speaker claims there is an automaton that does not have mental phenomena correlated with it, even according to the laws of nature in reality. For example, a philosopher might claim that a liquid thermometer is a sufficiently simple

automaton that it does not have any mental phenomena directly correlated with it. The philosopher would then claim that this automaton is purely material. However, the philosopher is likely not claiming that the thermometer is an aphenomenal copy of a typically dual thermometer, because they are not claiming that the thermometer is a hypothetical alternate version of a supposedly typical thermometer with phenomenal consciousness. They are simply claiming that thermometers do not have phenomenal consciousness. This is a meaningful distinction in the primary intensions of the terms. However, a posteriori the terms mean the same thing for this purpose, because, as this work endeavours to show, all automata are dual in reality.

Another context where the terms have different meanings is when the speaker is considering the possibility of dual and aphenomenal automata with purely psychical states. The phenomenal consciousness of such an automaton is correlated with, and most naturally interpreted as being the substance underlying the causal structure of, its purely psychical states. In such a case, if the term *material* is interpreted to refer only to the physical universe, and to exclude the purely psychical states, then such an automaton lacking phenomenal consciousness would be an aphenomenal copy but not a purely material automaton. If purely psychical states exist, the thought experiment of the purely material universe thus would be better described as the thought experiment of the aphenomenal universe. Then, the paradox of dualism is the question of why the automata in that universe talk about their psychical states having phenomenal consciousness when these states do not have phenomenal consciousness correlated with them in that universe.

Another context where the terms have different meanings is when one is considering the possibility of aphenomenal copies from the perspective of a different ontological framework than dualism. According to dualism, an aphenomenal copy is a purely material automaton, and a purely material automaton is an aphenomenal copy, at least, if one is a panpsychist. But according to materialism, all automata are purely material, and none of them, properly speaking, are aphenomenal copies. According to idealism, no automata are aphenomenal copies, but aphenomenal copies are possible, but such aphenomenal copies would not be purely material, because the idealist claims matter does not exist. Rather, such aphenomenal copies would be purely theotaxic, that is, purely the organization of God without any material substance. This is revisited in more detail later in this work. According to neutral mon-

ism, no automata are aphenomenal copies in reality, but aphenomenal copies are possible, but such aphenomenal copies would not be purely material, because the neutral monist claims the substance of reality is phenomenal, not material. Rather, such aphenomenal copies would be pure causal flux without substance. At least, if the monist granted such a scenario to be possible, this is what they would claim an aphenomenal copy is.

On Objections to the Proposed Term

An objector may claim that the term *aphenomenal copy* is imprecise, because it does not in itself specify that the aphenomenal copy is not exactly the same thing as that of which it is a copy. For example, a materialist may claim that an automaton is purely material in reality, and they would thus claim that this automaton is aphenomenal, and thus also that a perfect duplicate of this automaton would be aphenomenal, and thus an aphenomenal copy. However, this aphenomenal copy would not be a zombie, because it does not differ from that of which it is a duplicate. Thus, the term *aphenomenal copy* is not a suitable replacement for the term *zombie* in this context.

A response to this is as follows. This is a pedantic criticism of little significance. It is true that a duplicate of an aphenomenal automaton would be an aphenomenal copy in a strict sense of the literal denotation of words. However, if one is specifying that something is an aphenomenal copy, then this would typically connote that the aphenomenal copy is not a perfect duplicate, and that the way it is not a perfect duplicate is precisely by virtue of its being aphenomenal. This connotation is what is intended by this term. Thus, a duplicate of an aphenomenal automaton is a copy, and it is also aphenomenal, but it is not quite, according to the typical connotation of words, an aphenomenal copy, because it is not the case that it differs from that of which it is a duplicate by virtue of being aphenomenal.

An objector may claim that the term *aphenomenal copy* is imprecise, because the so-called aphenomenal copy is not really a copy, because it differs from that of which it is a so-called copy.

A response to this is as follows. A copy need not be a perfect duplicate. And specifying that it is an aphenomenal copy naturally suggests that it differs from being a perfect duplicate precisely by virtue of its being aphenomenal.

An objector may claim that the term *non-phenomenal* should be used instead of the term *aphenomenal*, because the prefix *non-* is more common and well-understood than the prefix *a-* to denote negation.

A response to this is as follows. The word *phenomenal* comes from ancient Greek, and thus it is more fitting to use the prefix *a-*, which is the Greek prefix for negation, than the prefix *non-*, which is the Latin prefix for negation, when compounding from this term. Furthermore, in English use the two prefixes have slightly different shades of meaning. The prefix *a-* suggests absence, whereas the prefix *non-* can instead suggest replacement by an alternative. And it is the former meaning that is intended here. Thus, the prefix *a-* is here preferable even for semantic precision, not merely etymological purity.

An objector may claim that the term *aphenomenal* is undesirable because the prefix *a-* may be mistaken for the indefinite article *a*, especially in spoken conversation.

A response to this is as follows. This criticism applies in general to the prefix *a-*, which is a strong burden of criticism, given that this is a widely used and widely understood prefix. In practice, the term *aphenomenal* will usually be preceded by some word, such as the definite or indefinite article, that makes it clear, even in spoken conversation, that the prefix *a-* is not the indefinite article. Also note that the prefix *a-* in the word *aphenomenal* is pronounced like the vowel sound in *say*, whereas the indefinite article usually does not have this vowel sound, especially in its form *an*, which will in practice be the form preceding the word *aphenomenal*. This further reduces the risk of confusion.

An objector may claim that the term *zombie* is preferable because it is shorter than the term *aphenomenal copy*.

A response to this is as follows. The more just comparison would involve the term *philosophical zombie*, because this extended form is the one resorted to when it is necessary to indicate that the term *zombie* does not refer to an undead being in the typical sense. And the term *aphenomenal copy* is only a little longer than this extended term, which is a small price to pay for it being much more precise.

An objector may claim that the term *zombie* is preferable because it is more fun and lighthearted. This is good for its own sake, because even philosophers are people who can laugh and enjoy jokes, and it is also good for encouraging people to learn about philosophy, by making it not seem so cold and abstruse.

A response to this is as follows. Philosophers, in the capacity of being people, may in other contexts act with more levity than is appropriate in philosophical discourse. However, philosophical discourse should be the place where the philosopher strives to escape from the levity of everyday life and to make their thoughts as pure and rigorous as possible. Humour and other whimsical forms of self-expression are best reserved for other contexts where they are more appropriate. As regards the claim that levity in philosophical terminology can induce more people to learn philosophy the following may be said. Perhaps if one is trying to advertise philosophy, then a term with more levity may be more appealing to a certain kind of target audience. However, in serious philosophical writing, the communication of truth is more important than shallow considerations of popular appeal.

An objector may claim that it is vain to try to change the use of words, and that one should simply follow custom and use the term *zombie* instead of the term *aphenomenal copy*.

The response to this is as follows. If the objector dislikes the term *aphenomenal copy*, they may simply substitute for it the term *zombie*. The author does not claim that this term will gain any traction against the term *zombie* among other writers. He is content to use language as he sees fit, and to let others use language as they see fit. So long as everyone can understand everyone else reasonably well, there is little need to worry about such things very much.

However, the term *purely material automaton* is preferred in this work to the term *aphenomenal copy* when the difference in the shades of meaning of these terms is not relevant to the discussion.

On the Assumptions of the Ensuing Discussion

This chapter has endeavoured to establish dualism as the foundational dogmatic stance of this work. It establishes dualism over skepticism and materialism with confidence. However, the debates between dualism, idealism, and neutral monism are subtle, and these latter two doctrines should be admitted as at least possible, though the author is more sympathetic to dualism. However, as with idealism, it would obscure the intended train of thought to discuss neutral monism in detail here. And, as with idealism, neutral monism preserves the

important characteristics of dualism that are significantly invoked in the ensuing discussion. Thus, though the ensuing discussion is superficially dualistic, it may be interpreted more deeply as compatible with neutral monism or idealism. However, it would be prolix and confusing to phrase the ensuing train of thought in a way that remained agnostic with regard to whether dualism, idealism, or neutral monism is true, and it would contribute little to the accuracy of the doctrines described. Instead, the method adopted in this work is to postpone more detailed discussions of idealism and neutral monism until later in this work, and for the present to follow the train of thought of this work from the dualistic standpoint. Later, the discussions of idealism and neutral monism should indicate how all of the dualistic doctrines can be interpreted from the perspective of those other doctrines. Thus, taking dualism as true in the ensuing discussion is not completely based on rational argumentation justifying this doctrine. It is partly a hypothesis adopted for the sake of convenience of expression. But the hypothesis is only adopted for the sake of convenience of expression, and not dogmatically. Thus, until idealism and neutral monism are discussed later in this work, the discussion may be interpreted by the dualist as true, and by an idealist or identity monist at least as speculation on what would be true if dualism were true. However, since idealism and neutral monism bear close similarities to dualism, many of the doctrines should be interpreted by the idealist or identity monist as more than just speculation presupposing a hypothetical assumption with which they disagree. These doctrines can still retain much of their meaning if they are interpretively transposed into the framework of the other system. But this is discussed in detail later in this work.

On the Coherence of Purely Material Automata

According to dualism, purely material automata are possible. It is another question whether their behaviour would be coherent, and how exactly it would be coherent. In order to establish this stronger claim, it is necessary to investigate whether the mind has any causal efficacy on the body. For if it does, it would appear that the behaviour of purely material automata would be incoherent. This investigation is contained in the next chapter, which argues that the mind does not have causal efficacy on the body, and that the behaviour of

purely material automata is thus coherent, and should be interpretable without taking into account the existence of mental phenomena.

On Epiphenomenalism

Introduction

Dualism having been established and elaborated on in the last chapter, this chapter further specifies the dogmatic stance of this work, arguing for epiphenomenalism, the doctrine that mental phenomena do not have causal efficacy on matter, rather than interactionistic dualism or other alternatives. This dogmatic stance having been described and justified, the remainder of this work is put forward assuming, and in parts further justifying, epiphenomenalistic dualism.

The main doctrines of this work rely on a model of reality in which the material universe is causally closed, and in which the mind thus does not have causal efficacy on the material universe. This doctrine is epiphenomenalism. Epiphenomenalism is not a new doctrine. No scientific evidence refutes it, and, on the contrary, scientific evidence leads in the direction of it being true. Many scientists and philosophers believe in epiphenomenalism, and this has been so since at least the end of the nineteenth century. This chapter describes and defends epiphenomenalism.

On Causation

A model is accurate if it correctly predicts the state of reality.

A model is natural if it models complex details with simple laws that are fitting for the purpose.

A model assumes that something exists if it takes the state of that thing into account when deriving predictions from inputs.

A thing has causal efficacy on a second thing if, to naturally and accurately model the behaviour of the second thing, the model must take into account the state of the first thing.

A thing is a causally closed system if the rest of the universe does not have causal efficacy on it.

If a thing has causal efficacy on a second thing, it may be said that it influences the behaviour of the second thing, or causes the behaviour of the second thing to be as it is.

The natural and accurate model in the definition of causal efficacy need only exist in principle, and whether or not it is practically possible for a human being to determine this model or apply it does not affect the definition of causal efficacy.

A precise formalization of what it means for a model to be natural is beyond the scope of this work, nor does the author possess such a formalization. A model is unnatural if it is unnecessarily complex. For example, an unnatural but accurate model would have no general laws at all. It would include the complete state of the universe implicitly in its physical constants and laws without giving any organized structure to these constants and laws, thus having no explanatory power besides being a recordation of fact. This model would be unnatural because it would have much unnecessary complexity.

This definition of causal efficacy is not given here for its own sake, but for the sake of discussing the causal efficacy or lack thereof of the mind on the automaton with which that mind is correlated. Thus, the definition of causal efficacy given above need not be meticulously precise, so long as it suffices for the ensuing discussion on mental causal efficacy. Much more could be said about causal efficacy in general, and each of the doctrines of this section could be the subject of metaphysical debate. However, such debate is beside the point of this work, which is not overly concerned with the exact details of the foundations of metaphysics.

An objector may claim that this definition of causal efficacy does not characterize what it is for causation to be causation. It only specifies a criterion for detecting causal efficacy, if only in principle. But in addition to the behaviour that is observed that informs whether causation is occurring, there is also the necessary power of causation itself, which is the essence of true natural laws, but which this definition ignores, thus missing an important part of reality.

A response to this is as follows. As David Hume showed, there is no hidden power that underlies causation. For even if one could identify some internal essence or power to things that was supposedly what underlied causation, one could still conceive of that exact essence or power existing without the causation then occurring. The best possible account of causation is to describe the occurrences that instantiate it.

On Purely Mental Phenomena

A purely mental phenomenon is a mental phenomenon that has no material correlate.

That is to say, a purely mental phenomenon is a mental object with no corresponding material object. It is not known whether purely mental phenomena exist. A purely mental phenomenon is usually considered to be, but is not necessarily, a mental phenomenon that has causal efficacy on the body. However, this need not necessarily be thus. For example, it is conceivable that there could exist a purely mental phenomenon that simply existed and was not correlated with the material universe in any way, neither arising because of it, nor interacting with it.

Many of the arguments in this work rely to some extent on the assumption that purely mental phenomena do not exist.

On Mental Causal Efficacy

It is reasonable to hypothesize that if a mental phenomenon has causal efficacy on matter, then it has causal efficacy directly on the material correlate or on what is in immediate proximity to the material correlate, and on other matter indirectly through the material correlate.

If a mental phenomenon has causal efficacy on matter, then it is purely mental. This may be demonstrated by demonstrating its contrapositive. Let the mental phenomenon be the particular mental phenomenon, universally quantified, considered for the purposes of the discussion. Let the mental phenomenon not be purely mental. Thus, the state of the mental phenomenon is a function of the state of the material universe. Then, in any model of the behaviour of the material universe, whenever the state of the mental phenomenon appears, it may be replaced by the function of the state of the material universe. Therefore, the behaviour of the material universe may be completely modelled assuming that the mental phenomenon does not exist. Therefore, the mental phenomenon does not have causal efficacy on matter. By contraposition, if the mental phenomenon has causal efficacy on matter, then it is purely mental. This needed to be demonstrated.

A consequence of this implication is the weaker doctrine that if minds in general have causal efficacy on their material correlates, then purely mental phenomena exist.

The converse of this implication is not necessarily true under some interpretations of the definition of a purely mental phenomenon. For it is possible that there could exist a purely mental phenomenon that is arbitrary and exists by itself as a closed system, and has no relation whatsoever to the material universe, either causal or correlative.

The existence of causal efficacy of the mind on matter is also independent of the existence of free will. The argument is similar, though not strictly identical, to the argument for the independence of the existence of purely mental phenomena and the existence of free will. The possibility of the non-existence of causal efficacy and the existence of free will follows from the possibility described in the section On Free Will, and the implication that a mental phenomenon that is not purely mental has no causal efficacy, as does the possibility of the non-existence of causal efficacy and the non-existence of free will. Next, suppose the mind has causal efficacy on the body through its purely mental phenomena, and the body cannot be modelled as a purely material automaton, but the behaviour of all mental phenomena, even of purely mental phenomena, is predetermined by God or by some non-anthropomorphic law of necessity. Then, causal efficacy of purely mental phenomena exists, but free will does not. Next, suppose that the mind has causal efficacy on the body through its purely mental phenomena, and it is through this causal efficacy that

free will exists. Then both exist. Both could exist without the causal efficacy of purely mental phenomena being what enables free will. Therefore, the existence of free will and the existence of causal efficacy of purely mental phenomena are independent.

In the remainder of this work, the conjunction of the existence of purely mental phenomena and the non-existence of causal efficacy of the mind on matter is not discussed further.

If mental causal efficacy exists, it exerts a negligible influence on behaviour over a short period of time. The justification for this doctrine is empirical evidence. All empirical evidence so far has failed to detect mental causal efficacy. Therefore, if mental causal efficacy exists, it must have a negligible influence over a short time so as not to be detected even by careful observation. If it exists, it must be unapparent either in scope or in intensity of difference in structure from the material correlate, or both.

On Epiphenomenalism

Epiphenomenalism is the doctrine that the mind does not have causal efficacy on the body.

This chapter clarifies the exact meaning of this doctrine as it is understood in this work, and argues that it is valid. Later parts of this work provide further evidence for epiphenomenalism by building a theory of purely material behaviour that is consistent with it, and which solves one of the greatest problems facing epiphenomenalism, that of explaining why automata tend to be dualistic if the mind is an epiphenomenon. However, most of the doctrines of this work are logically compatible with interactionism if some care is taken in the analysis.

In ancient times, it was generally believed that the mind had causal efficacy on the body. This was partly because it was not properly understood that the mind is not a part of the material universe. It is now known that the behaviour of the body is mostly or completely automatic, and opposition to the claim that it is completely automatic only claims that the amount of causal efficacy the mind has is relatively minor. The claim that the function of the body is completely automatic was first formulated in the nineteenth century. At that time, epiphenomenalism was more often called the *automaton theory* or the

conscious automaton theory. The term *epiphenomenalism* is more common in current usage.

One particularly well-known early formulation of epiphenomenalism is that of Thomas Henry Huxley. In an essay, he compared the relation between the mind and the body to that between the steam whistle of a locomotive engine and its machinery. He also compares the relation of the mind to the body to the relation of the sound of a bell to the bell itself. These comparisons are well-known, and his essay expressing them is often cited as a standard early formulation of epiphenomenalism. Given that it is so often cited, it is a shame that the comparisons Huxley uses are not completely accurate. For the steam whistle that accompanies the work of a locomotive engine has some small influence on its machinery, though in the short term this influence is negligible for most purposes. For example, the mass of the water molecules in the steam exerts some gravitational force on the machinery of the locomotive, and the pressure of the sound waves of the whistle puts some pressure on the body of the locomotive, and so on, all of which affect the machinery in some slight way. Thus, it is correct to state that the consciousness of animals is as without power of modifying the working of their body as the steam whistle which accompanies the work of a locomotive engine is without influence on its machinery, for the same reason that it is correct to state that zero is less than a particular very small positive real number. However, just as the latter comparison is an insufficient characterization of the number zero, the former comparison is an insufficient characterization of epiphenomenalism. Strictly speaking, the comparison leaves room for consciousness to have some slight effect on the body equal to or less than that of the steam whistle on the machinery of the locomotive. But such a scenario would be better described as interactionistic dualism. Similarly, the sound that a clock gives out when it is struck has some slight influence on the clock itself. For the vibrations of the air interact with the clock and influence its behaviour to some small extent. At least, this is so if the sound is interpreted as a material process. If the sound is interpreted as a conscious experience of sound, then this imprecision is avoided, but the metaphor is no longer a metaphorical illustration, because the so-called metaphor is using the very thing it is supposed to illustrate to illustrate that thing. However, this remark verges into pedantry. Huxley was writing in the imprecise, rhetorical style of an essayist, and the reader is presumably expected to gather that he means that consciousness has no influence on the body whatsoever.

Thus, this remark should be interpreted, not as a real criticism of what Huxley wrote in the capacity of being an essay, but as a clarification of the precise meaning of epiphenomenalism.

On Purely Psychical States

A psychical state is a state with causal efficacy that is directly correlated with a mental phenomenon.

A purely psychical state is a psychical state that is not a physical state.

Thus, a psychical state is the structure of a mental phenomenon.

The psychical universe is the causal structure of the mental universe.

If the mental phenomenon is in direct epiphenomenal metaphysical correlation with a material representational substrate, then the psychical state of this mental phenomenon is just the metaphysical structure of its material representational substrate. If the mental phenomenon is not in direct epiphenomenal metaphysical correlation with a material representational substrate, then its psychical state is a purely psychical state.

A purely psychical state is in causal interaction with the material universe.

A demonstration of this is as follows. If the material universe had no causal efficacy on the purely psychical state, then the mental phenomenon would be disembodied. There is no empirical evidence or theoretical justification for such a mental phenomenon to exist. Thus, this case may be taken as absurd and naturally impossible according to reasonable supposition. Thus, the material universe has causal influence on the purely psychical state. If the purely psychical state had no causal efficacy on the material universe, then the mental phenomenon would be in direct epiphenomenal correlation with some aspect of the material universe. But it is reasonable to assume that if a mental phenomenon is in metaphysical correlation with some aspect of the material universe, then this aspect is a material representational substrate. For mental phenomena are fundamentally representational. Therefore, this case may also be taken as absurd and naturally impossible according to reasonable supposition. Thus, a purely psychical state has causal efficacy on the material universe. Thus, the purely psychical state is in causal interaction with the material universe. This needed to be demonstrated.

The properties of purely psychical states and a criterion according to which it may be justly claimed that a state is psychical rather than physical are as follows. A psychical state could be considered purely psychical and not physical if it lacked characteristics typical of the physical universe, such as position and momentum. In such a situation, the structural laws of nature would be naturally divided into laws of the behaviour of physical states, laws of the behaviour of mental states, and laws of their interaction. The physical states would have typical physical characteristics, such as position and momentum, and would obey physical laws, such as conservation of momentum and energy, at least if the causal efficacy of psychical states did not break these conservation laws. By contrast, the psychical states would have a different set of structural characteristics and follow different laws. These psychical characteristics and laws would be sufficiently different from physical characteristics and laws that the psychical structure would be most naturally considered as forming a psychical universe separate from, but causally connected with, the physical universe.

It is hard to hypothesize about what these supposed psychical characteristics and laws would be, because all scientific evidence currently suggests that no such characteristics or laws exist. But for example, one psychical characteristic could be a variable representing the liveliness of mental phenomena, in a way that was not merely a function of the physical state alone. Then, a psychical law could be a law governing how the liveliness variable interacts dynamically with the physical state of the automaton, or with other psychical variables.

On Psychical State Automata

A psychical state automaton is an automaton with a purely psychical state.

The proximal material state of an automaton with a purely psychical state is the physical part of the automaton with which the purely psychical state is in direct interaction.

The distal material state is the complement of the proximal material state in the physical part of the automaton.

The proximal material state of an automaton with a purely psychical state is analogous to the material representational substrate of an automaton

with causally ineffective psychical structure. The distal material state of an automaton with a purely psychical state is analogous to the material integration of an automaton with causally ineffective psychical structure.

For the purpose of discussing psychical state automata, it is assumed that purely psychical states exist, at least for the automaton to which the discussion applies. This should not be interpreted to mean that this work advocates that purely psychical states exist. On the contrary, it appears likely that they do not. Therefore, the following discussion should be interpreted as hypothetical speculation.

On a Principle of Psychophysical Interaction

The principle of representational psychical interaction is the doctrine that a purely psychical state can only interact with the proximal material state according to the metaphysical structure of representation, not with any of the fundamental physical properties of the proximal material state directly in a way that is straightforward to characterize physically.

That is to say, a purely psychical state may only interact with the proximal material state modulo the abstraction of metaphysical structure. A purely psychical state cannot interact with the basic physical properties of the proximal material state, nor with any of its physical properties except insofar as they have representational structure.

This may be taken as a definition of the metaphysical structure of the interaction between matter and purely psychical states. After all, if purely psychical and physical states interact in a certain way, then that is the way that they interact, and the principle of organizational invariance is merely recording that fact. Above, when speculating on what a purely psychical state might be like, it is said that it might have properties that could not be straightforwardly characterized as physical. This is an almost aesthetic doctrine, determining how to define purely psychical states in the first place.

However, given natural assumptions on the nature of representation and automata, this principle may be interpreted to yield propositions about the properties of purely psychical states. That is to say, this definitional doctrine may be extended to propositions about reality using hypotheses about representation. For example, if one assumes that purely psychical states can only re-

flect the integration of information according to the time rate of the organizational realization of the physical part of the automaton, then this means that the subjective time rate for memories characterized purely psychically does not intrinsically specify physical time rates. Different physical realizations of the same physical integrative structure may have different time rates of change, such as different rates of a simulation, and though their purely psychical states could differ even at points where the physical states were organizationally isomorphic, the psychical states could only differ modulo the difference in time rate of change. And a similar speculation could be made about other physical attributes, such as position, mass, and charge. However, this is only speculation, and likely chimeric speculation at that, because it is likely that purely psychical states do not exist.

In this respect, the discussion for purely psychical states is similar to that for causally ineffective psychical states, and the discussion on the necessarily non-physical attributes of mental phenomena in the section On the Principle of Organizational Invariance in the chapter On Correlation is relevant to the above observation, though not logically equivalent to the doctrine.

On the Causal Closure of the Physical and Psychical

The union of the physical universe and the psychical universe is causally closed from phenomenal consciousness.

A demonstration of this doctrine is as follows. If the physical material universe is causally closed, then the psychical structure of the mind is isomorphic to the metaphysical structure of the material representational substrate, and phenomenal consciousness is an epiphenomenon, and the proposition that the union of the physical universe and the psychical universe is causally closed holds. Therefore, suppose that the physical universe is not causally closed. If the non-physical state that causally influences the material universe is not psychical, but of some other exotic substance not directly correlated with the mind, then still the psychical structure of the mind is isomorphic to the metaphysical structure of the material representational substrate, and the mind does not have causal efficacy. Therefore, suppose that at least some part

of the non-physical state that causally influences the physical universe is psychical, that is, directly correlated with the mind.

Consider some particular automaton for which this is the case. The representational substrate, that is, the structural correlate of the phenomenal consciousness of this automaton, is partly physical and partly, or entirely, psychical. The proximal material state is not isomorphic to the structure of the mind of this automaton. It is unlikely that such a purely psychical structure exists, but it is conceivable. And it may seem that mental causal efficacy would exist in such a state of affairs. And if such a state of affairs is conceivable, then it may seem that mental causal efficacy is possible.

The reason why this is not the case is that phenomenal consciousness is separate from any psychical state that it is the correlate of, just as phenomenal consciousness is separate from any physical state that it is the correlate of. In reality, the mind knows that there is something that it is like to mentally innately intuit its mental phenomena, irrespective of whether the structure of those mental phenomena has a purely psychical component, or whether it is isomorphic to the material representational substrate. But it is conceivable that there could exist an automaton with the same physical and psychical structure, but for which there was nothing it was like to intuit innately. There would be no phenomenal consciousness correlated with its psychical structure, even though its psychical states had the same structure as the version that was not an aphenomenal copy, which had phenomenal consciousness. This could be the case, for example, in a scenario where God cognized the structure of the physical and psychical states but chose not to create phenomenal consciousness correlated with this structure. Thus, physical and psychical structure could exist without phenomenal consciousness existing. Therefore, physical and psychical structure are causally closed from phenomenal consciousness. This needed to be demonstrated.

Thus, epiphenomenalism is valid a priori, provided that epiphenomenalism is properly defined. Even an interactionist dualism positing purely psychical states is compatible with this a priori epiphenomenalism.

The above a priori demonstration of epiphenomenalism is influenced by a discussion by David Chalmers in the section Is This Epiphenomenalism? in the chapter Naturalistic Dualism in his work *The Conscious Mind*. The discussion given there has been here reformulated and elaborated to fit the context of this work. Until he read that discussion, the author of this work mistakenly

thought that interactionistic dualism was logically possible, and he only rejected it based on the independent evidence for epiphenomenalism that some of the later doctrines of this work provide. But after reading that passage, he saw his error, and realized that epiphenomenalism is valid a priori, provided one defines epiphenomenalism properly.

On the Rehabilitation of Interactionistic Dualism

Interactionism is traditionally defined in a way that it is incompatible with epiphenomenalism. However, since epiphenomenalism is true a priori, and since there is an element of interactionism that is not necessarily false, and which could thus use a name, this suggests the following rehabilitation of the term *interactionism*.

Explanatory epiphenomenalism is the doctrine that the existence of phenomenal consciousness is explanatorily irrelevant for modelling physical behaviour.

Non-epiphenomenalism is the negation of epiphenomenalism.

Interactionism is the doctrine that purely psychical states exist.

Non-interactionism is the negation of interactionism.

Epiphenomenalism is true a priori. Non-epiphenomenalism is false a priori. Interactionism is unlikely based on a priori and aesthetic considerations. Non-interactionism is likely for the same reason.

Epiphenomenalism is independent of interactionism. These two different dimensions of doctrine may be varied independently to yield four doctrines. Epiphenomenalistic interactionistic dualism is possible but unlikely. Epiphenomenalistic non-interactionistic dualism is possible and likely. Non-epiphenomenalistic interactionistic dualism is impossible a priori. Non-epiphenomenalistic non-interactionistic dualism is impossible a priori.

Thus, mental phenomena do not have causal efficacy on physical states in the capacity of being phenomenal.

Explanatory epiphenomenalism may be called for brevity *epiphenomenalism*, because this is the version of epiphenomenalism that matters most and is valid a priori. As for whether there is some ontological or aesthetic sense in which the existence of mental phenomena is still necessary for causa-

tion to occur in some way, such a notion is likely chimeric, and even if it is true, it is still compatible with much of the remainder of this work.

On the Defence of Epiphenomenalism

The above a priori demonstration of epiphenomenalism should in principle suffice in itself. However, since epiphenomenalism has many influential critics, in addition to this positive demonstration of this doctrine, it may be worthwhile to give a more detailed critique of various arguments against it. This is done in some of the following sections. The general method is to show that any supposed alternative to epiphenomenalism is either false or compatible with epiphenomenalism. In the former case, the alternative cannot stand. In the latter case, the debate becomes merely that of how one should use the term *epiphenomenalism*, and there is no need to argue about this. If someone wants to use it with a different shade of meaning, they may do so, and they will hopefully be able to understand the meaning of this work when this work claims that epiphenomenalism is valid a priori without feeling too much aesthetic dissatisfaction at this use of language.

On Consciousness and Measurement

A detailed discussion of quantum mechanics is not within the scope of this chapter, the purpose of which is to establish epiphenomenalism. A detailed scientific discussion of quantum mechanics would not even be within the scope of this work. There is a philosophical discussion of quantum mechanics later in this work. However, within the scope of this chapter, it suffices to show that quantum mechanics is compatible with epiphenomenalism. Most of quantum mechanics has little to do with epiphenomenalism. The point where they come into tension is the quantum-mechanical measurement process. Thus, within the scope of this chapter, it suffices to show that this process is compatible with epiphenomenalism. Most viable interpretations of the measurement process are compatible with epiphenomenalism. Therefore, within the scope of this chapter, it is not necessary to express any opinion in favour of any particular interpretation of this process. The only interpretation that comes into tension with it is the interpretation claiming that consciousness causes

collapse. Therefore, within the scope of this chapter, it suffices to argue against this interpretation.

Now, authority is not a good reason to believe anything in philosophy. However, it is not a terrible reason, either. And authority is not in favour of the theory that consciousness causes collapse. For this theory is not in fashion among academic physicists and philosophers, who are the most trusted and respected authorities on the interpretation of quantum mechanics among rational thinkers. Rather, the theory tends to be associated with superstitious mysticism, sensationalized popular science, and delusional sophistry. Thus, authority alone suggests that this theory is flawed. However, this is not enough for the purposes of this discussion. For reputable authority has erred in the past, and at first appearance it is conceivable that it could now err with regard to this theory. And the inclination of authority is not unanimous. Some reputable thinkers acknowledge this theory as a possibility, and some of its advocates are important rational thinkers. Therefore, the subject must be examined in itself, not merely relative to authority.

In this connection, note that no scientific evidence requires this theory to be true. Many alternate interpretations of quantum-mechanical measurement are viable explanations of all quantum-mechanical phenomena observed to date. For example, there are interpretations that involve no collapse of the wave function, such as the hidden variable interpretation associated with Bohm or the many-worlds interpretation associated with Everett. And there are interpretations that involve the collapse of the wave function, but where this collapse occurs when some physical criterion is satisfied. Such a physical criterion need not involve consciousness, and its satisfaction may be only imperfectly correlated with the presence of consciousness. For example, such a physical criterion may be a gravitational effect from the mass of the system involved. Roger Penrose suggested such an interpretation. And a predominantly gravitational effect cannot be perfectly correlated with the presence of consciousness, because the physical processes that give rise to consciousness are not purely gravitational, and likely not even predominantly gravitational. Another physical criterion could be the number of particles in the system, any one of which has some small probability of collapsing. One such interpretation is that due to Ghirardi, Rimini, and Weber. And a quantitative reckoning of this kind cannot be perfectly correlated with the presence of consciousness, because the physical processes that give rise to consciousness are not merely

based on the number of particles interacting in a system. Now, neither is that objective collapse theory, but its proposed criterion for collapse has a less intricate structure than the process responsible for consciousness.

Thus, there is no need to posit that consciousness causes collapse, and no empirical evidence now exists that requires this interpretation as opposed to any other existing viable interpretation.

On Collapse before Conscious Measurement

An objector may claim that consciousness must cause collapse because measurement causes collapse, and measurement is a conscious process. For all measurement is done by a conscious being, and therefore consciousness has causal efficacy on the system being measured.

A response to this is as follows. Wave function collapse occurs before the direct involvement of the human mind, or even of the human brain. It occurs once the state of the measured system becomes correlated with a classically describable state in the measurement apparatus. This process occurs before the intuition of the classically describable state of the measurement apparatus by the mind or brain. If there are no people or animals present, but the measurement apparatus measures the quantum system, then wave function collapse occurs. If, afterwards, a person observes the classically describable state of the measurement apparatus that represents the previously existing state of the measured system before measurement, this does not significantly affect the wave function of the measured system, and in any case, wave function collapse has already occurred. Therefore, to explain why wave function collapse occurs, there is no need to posit that the mind of the observer has causal efficacy on the measured system. What has causal efficacy on the wave function of the measured system is the measurement apparatus that collapses it, which is a material system.

An objector may claim that until a mind observes the outcome of the measurement, the outcome is uncertain, not merely practically, but fundamentally, and the possible classically describable states of the measurement apparatus exist in a quantum superposition.

A response to this is as follows. As described earlier, there are a number of viable and more widely accepted interpretations of quantum mechanics that

do not involve this hypothesis. Among these are those not involving any collapse of the wave function, and those involving a physical criterion for collapse.

If an objective collapse interpretation is valid, then there is no such superposition of classically describable states in the measurement apparatus, as it would violate the basic physical law that causes the wave function of sufficiently large systems to collapse spontaneously, irrespective of whether a mind or a measurement apparatus measures the system. Thus, one definite classically describable state exists in the measurement apparatus, even if it is practically uncertain to the experimenter until they observe it. Essential non-localization of the wave function must not be confused with practical uncertainty. There is practical uncertainty, but not essential non-localization in this case.

It must be clarified here that the classically describable state is not necessarily completely definite. There is a sense in which, according to some objective collapse interpretations, the classically describable state, and the whole universe, is technically in a quantum superposition to some minuscule degree, in that the value of the wave function at various classical states is not necessarily zero. But its probability weight at these different states is negligible except at the actual state, where it resembles the Dirac delta function. However, even such a small but non-zero weight in the tails is aesthetically undesirable in a theory. This is discussed in more detail later in this work. In any case, significant quantum superposition cannot take place in classically describable states.

In hidden variable interpretations, reality is collapsed before measurement, so there is no meaning to saying that the mind leads to collapse, because collapse does not occur.

However, in other interpretations, one could claim that the possible classically describable states of the measurement apparatus are in a quantum superposition, though their standard formulations do not involve this assumption.

In the Copenhagen interpretation, the wave function is fully collapsed at the classical level, and there is not even a negligible amount of quantum superposition of classical states. Collapse occurs once the measurement apparatus measures the system. But in either case, there is no global non-localization of the wave function.

In the many-worlds interpretation, which posits a globally uncollapsed wave function that represents the state of an existing uncollapsed substance, there is never a quantum superposition of classically describable states on one

branch from the perspective of one mind, as this would violate the measurement postulate, now a branch postulate, for the same reason that it would violate it according to any other viable interpretation of quantum mechanics. The branching of the wave function from the perspective of one mind, the equivalent under this interpretation of wave function collapse, occurs when the measurement apparatus measures the system, prior to the mind intuiting the classically describable state of the measurement apparatus.

Furthermore, note that the claim that the state of the measurement apparatus is uncollapsed because consciousness has not yet observed it is incompatible with panpsychism, or at least a doctrine that the measurement apparatus has some rudimentary phenomenal properties. Such panpsychism is not now a widely accepted doctrine, so this incompatibility is not at first appearance a serious weakness of the claim. However, this work later advocates such a form of panpsychism. These later doctrines of this work show that the objector, in claiming that the measurement apparatus has no consciousness correlated with it, is false.

On Collapse with Conscious Measurement

But as a concession, suppose that the wave function really were uncollapsed after the measurement apparatus measures the system, but before the human mind intuits the measurement apparatus. If this were true, which is doubtful, then the correlation of observation by the human mind with wave function collapse would still be no evidence that the mind causes collapse, because the collapse would be caused by the brain process correlated with the mind rather than by the mind itself. Then there is still no need to posit that the mind has causal efficacy on matter.

Measurement, even conscious measurement by a human or animal, is not done by the mind alone, but by the mind and matter in tandem. Therefore, if measurement alters the wave function, all this implies is that either the mind, or matter, or both alter the wave function when measurement occurs. But this does not imply that the mind alters the wave function. If observation by the mind is correlated with wave function collapse, this need only be because the mind is metaphysically correlated with material innate intuition, and wave function collapse is at least statistically correlated with material innate intu-

ition. The material process is what causes collapse. There is no need to posit that the mind is the cause of wave function collapse.

All innate intuition is a tandem act of the mind and the material integration. The effect of measurement can thus be modelled as resulting from the material innate intuition of the experimenter. Thus, the quantum mechanical measurement process would be physically identical in the purely material universe, where the experimenter is a purely material automaton. That is, a purely material automaton would cause collapse just as well as a dual automaton.

There is no evidence that the mind causes wave function collapse. The full structure of the measurement process can be explained under the hypothesis of the material universe being causally closed. Therefore, quantum measurement is not evidence that the mind has causal efficacy on the material universe.

On the Cause of Collapse

An objector may claim that it is necessary to posit that consciousness causes collapse even if collapse occurs prior to observation by the human consciousness of the experimenter, because the faculty of intuition of the measurement apparatus itself has a sort of rudimentary consciousness correlated with it, and that it is this rudimentary consciousness of the measurement apparatus that causes wave function collapse, not the material process of the measurement apparatus in itself.

A response to this is as follows. Such a hypothesis that all material innate intuition has consciousness correlated with it is not now widely believed, so an objection based on assuming it to be the case is not of much force independently. But in fact, this work does argue that this hypothesis is true. However, even though it is true, this doctrine does not imply that it is the rudimentary consciousness of the measurement apparatus that causes the collapse of the wave function. For it is the material process of intuition that causes collapse, and the fact that this material process happens to be correlated with consciousness is not important for the physical process itself. Measurement would be the same if the measuring apparatus were a purely material automaton. And furthermore, the material innate intuition of the measurement apparatus is not necessarily perfectly correlated with wave function collapse.

An objector may claim that the most viable interactionistic dualistic theory is the theory that consciousness causes collapse because it may be a law of nature that mental phenomena cannot be in two states at once, and thus, when a superposition of states of the material correlate would lead to a superposition of states of mental phenomena, this is not permitted, and the mental phenomena force a collapse. This provides a natural explanation of why collapse happens that puts consciousness in a causal role.

A response to this is as follows. The argument by subtraction of phenomenal properties to yield identical causal laws applies. One could still subtract the phenomenal consciousness in the above model and obtain a purely physical account of collapse, characterizing it as occurring whenever certain physical conditions arise. As for whether these physical conditions also happen to be those that are correlated with mental phenomena of a certain kind existing, that is interesting, but it is merely an aesthetic, or at best ontological observation, and is an additional fact to the causal structure of the physical, and not entailed by it in itself. Thus, consciousness is a priori explanatorily irrelevant for modelling wave function collapse because it can conceivably be subtracted from any picture of reality to yield identical physical behaviour.

Another thing to note about the theory that consciousness cannot be in a quantum superposition is that it is incompatible with forms of panpsychism where fundamental particles have mental phenomena correlated with them. For by the same mechanism that supposedly explains why macroscopic systems are not superimposed, the claim that consciousness cannot be in a superposition would imply that microscopic systems would collapse because of the consciousness correlated with them. But in reality, microscopic systems can be in a quantum superposition. Therefore, these systems either do not have mental phenomena correlated with them or mental phenomena can be correlated with superimposed systems. But panpsychism is a reasonable hypothesis, and later in this work, a positive argument is put forward in favour of it. Thus, the incompatibility of this theory of collapse with panpsychism casts doubt on the theory of collapse.

An objector may claim that the way consciousness cannot be in a superposition does not need to be all or nothing. Perhaps there is a gradation of how degrees of consciousness lead to different probabilities of collapse, or only a certain amount of consciousness leads to collapse, or perhaps only protophenomenal properties are correlated with very small systems, and it is only

when these protophenomenal properties combine into true phenomenal properties that collapse occurs. All these possibilities adequately explain how microscopic systems can remain uncollapsed if consciousness resists being in a superposition.

A response to this is as follows. This alternative theory is just a standard collapse theory with an ontological frill appended. Standard collapse theories model how some physical condition can gradually lead to an increased probability of collapse or a cutoff past which collapse occurs. This supposedly new theory models how some mental condition can gradually lead to an increased probability of collapse or a cutoff past which collapse occurs. But these two kinds of theories are seen to be explanatorily equivalent when one replaces any condition involving a mental phenomenon with an equivalent condition expressed in terms of the state of the physical system. Thus, one could model collapse if consciousness were subtracted from the system entirely by referring to the physical condition. Thus, phenomenal consciousness is explanatorily irrelevant, even if the ontological frill of causation refers to anything in reality.

An objector may claim that there may be mental conditions leading to collapse that cannot be entirely characterized as a function of the physical state of the system. Thus, phenomenal consciousness is not explanatorily irrelevant, because the condition involving a mental phenomenon cannot be replaced with an equivalent condition expressed in terms of the state of the physical system.

A response to this is as follows. Perhaps purely psychical states exist, and perhaps these must be factored into a fundamental model determining when collapse occurs. However, the argument by subtraction of phenomenal properties still applies. The fact that there is something that it is like for these purely psychical states to exist is an additional fact beyond the fact that they interact with matter and cause collapse in some way, and it is not entailed by these other facts, and it is logically possible that it could not be the case while these other facts are the case. Therefore, even in such a state of affairs, which is quite likely not the state of affairs of reality, consciousness does not cause collapse.

On Ontological Interactionism

Interactionistic dualism based on overdetermination and other ontological considerations posits the same structure to the physical that epiphenomenalism posits, but with the ontological frill appended to the ordinary doctrines of epiphenomenalism that for some reason the existence of phenomenal consciousness is necessary for this causation to occur. The characteristic case is that of physics determining how things behave, but consciousness still being necessary in some ontological sense for that evolution to occur. Ontological interactionism is simply epiphenomenalism with an ontological frill, because phenomenal consciousness is still explanatorily irrelevant, because an aphenomenal scenario is still possible through the hypothetical subtraction of phenomenal consciousness.

An objector may claim that the hypothetical subtraction of phenomenal consciousness that the author imagines above misses the point that phenomenal consciousness is what underlies the causation occurring. If consciousness were not there, the causation would not happen that way. Thus, the causal role of consciousness cannot be played by a placeholder insubstantial variable. Thus, consciousness cannot be epiphenomenal. At least, one cannot show a priori that it is epiphenomenal, as the author endeavours to do.

A response to this is as follows. As David Hume showed, there is no sense in saying that some particular initial state of reality is logically necessary for some subsequent thing to occur. It is logically possible that any supposed causal factor that leads to some later event could be absent while that later event occurs, and it is conceivable that the supposed causal factor could be present while the later event does not occur. Therefore, there is no meaning to saying that consciousness is logically necessary for the collapse to occur. Thus, even if in nature consciousness were in constant conjunction with collapse, this would not change the fact that a hypothetical subtraction of consciousness is logically possible.

Even supposing that there is some ontological meaning to saying that consciousness is logically necessary, this does not change the fact that it can be conceivably subtracted without altering the causal structure. Perhaps, according to some natural ontological law of how reality should behave, perhaps even one that God has determined, if consciousness were not there, then the collapse would not occur. But it is still conceivable that the consciousness

could be absent while the collapse happens the same way, and this would be a consistent picture of reality, because the causal hypothesis is ontological, not structural. As for whether this conceivability is a true possibility, and if so, exactly what kind of possibility it is, this falls within the territory of metaphysical hair-splitting, into which this discussion has perhaps fallen too much already. Suffice it to say that reality with phenomenal consciousness subtracted would be logically possible, and even coherent and natural. Phenomenal consciousness is still explanatorily irrelevant.

An objector may claim that the phenomenal causation that underlies collapse could be non-deterministic, thus distinguishing it from physical causation. It would not be possible to subtract the phenomenal and obtain identical physical processes, because the non-deterministic aspect would not be there in the physical alone. It is impossible to construct a non-deterministic model that replaces the substance itself that causes the non-determinism, as the above response claims to do. Therefore, consciousness is not epiphenomenal. David Chalmers suggested such a theory in the chapter Consciousness and Its Place in Nature in his work *The Character of Consciousness*, though he did not positively advocate it.

A response to this is as follows. Injecting non-determinism into the discussion does not change the force of the relevant arguments already put forward. Any particular kind or instance of physical causation could conceivably be deterministic or indeterministic, and independently, any particular kind or instance of phenomenal causation could conceivably be deterministic or indeterministic. If physical causation could be indeterministic, then a coherent indeterministic causal structure could remain when phenomenal consciousness is subtracted. Furthermore, it is likely that non-deterministic causation is a chimera that philosophers tend to posit because of the misguided attempt to identify some logically necessary power underlying causation. If causation is non-deterministic, it is either not causation at all, but merely things happening without any pattern for determining how they happen, or it is ontically probabilistic behaviour, which also could conceivably happen otherwise.

An objector may claim that physical conditions may determine according to laws what will occur, but that mental phenomena are still necessary for causation to occur, and without the mental phenomena the causation would not occur.

A response to this is as follows. The above refutation of interactionistic dualism applies to this doctrine as well. This doctrine is really epiphenomenalism but with an ontological frill appended, to the effect that, for some reason, the philosopher can say with words that it is necessary for phenomenal consciousness to exist for physical causation to occur, though the causal structure of physical causation is the same as under epiphenomenalism. Both kinds of interactionistic dualism are merely ontological or aesthetic. They posit the same physical structure of reality that epiphenomenalism posits.

An objector may claim that the above arguments rely excessively on a particular abstract metaphysical way of thinking put forward by David Hume centuries ago in an age with far less scientific knowledge than modern society has. Philosophy has advanced since then, and philosophers should consider other views. Perhaps there could be a power or law that necessarily causes things to happen in some way, and perhaps phenomenal consciousness could have such a necessary causal power.

A response to this is as follows. Perhaps if, like a dogmatist appealing to authority, the responses had superficially appealed to the argument of Hume to dismiss anything that seemed counter to that argument, one could say that they relied excessively on it. But instead, this work has considered opposing views and refuted them according to their logical structure. If the argument used to refute them reliably converges on an argument given centuries ago by a celebrated philosopher, the best explanation for this is that that argument is valid. Furthermore, the above objection does not even give a particular argument for why some kind of causation could be logically necessary.

Ontological causal necessity is different from, and less real than, the typical causal necessity of physical objects. If two particles interact in a certain way, then for each particle, it is necessary to posit that it exists because it fills a certain causal role. At least the causal role exists, implicit in the interaction itself. Whereas to claim that the internal essence of the particle is necessary for it to play the causal role may be ontologically well-defined, but it is explanatorily irrelevant. And since, as David Hume showed, there is no power that makes causation logically necessary, it is chimeric to posit that mental phenomena, which may be subtracted from any account of physical causation, are necessary for that causation to occur, which is what neutral monism and this ontological last stand of purportedly interactionistic dualism claim. In any

case, physical causation is isomorphic according to all of the ontological theories.

For one to separate the two kinds of interactionistic dualism as if they were significantly different suggests one does not fully acknowledge how the quantum-mechanical interactionistic dualism he put forward is merely an ontological frill one can add on to epiphenomenalistic theory, rather than a competitor with it.

All this is not to say that the author completely eschews ontology. There is on the contrary a great deal of ontology in this work. But one engaging in ontology should not take it too seriously, or become confident that one is really talking about anything.

On Simulations and Ontological Hypotheses

This section describes a thought experiment that provides evidence against ontological theories of mental causation. The theory that consciousness causes collapse leads to absurdity in skeptical scenarios where reality is a simulation. The train of thought leading to this absurdity from this premise is as follows.

Suppose hypothetically that all of what the mind thinks is physical reality is being simulated by a computer. Thus, quantum superpositions do not exist ontically as superpositions, but are merely simulated as variables on that computer. A computer cannot store complex numbers to arbitrary precision, so it would likely use discretized approximation methods that are sufficiently precise that they cannot be detected by the mind in the simulation. Perhaps this could be avoided somewhat with quantum computing, but if it were avoided through and through, it would merely amount to reproducing the state of physical reality exactly with a physical state that was exactly like it, and this would not be a simulation. So there would need to be at least some element of what appears to be a quantum superposition really being simulated by a classically describable variable for this to qualify as a simulation. This assumes that the laws of physics in the simulating universe are like the laws of physics in the simulated universe, supposed here to be the reality of experience, or at least reasonably similar. If they are not, then there is little rational structure to the thought experiment, and the argument becomes a pedantic one about the definition of a simulation, and there is no need to enter into this.

Thus, suppose that the laws of physics of the simulating universe are similar to those of reality, and that the simulation is thus imperfect, and cannot represent all quantum states exactly using physical quantum states, but must use some approximation methods with classically describable states. In this case, a simulated person claiming that consciousness causes collapse amounts to claiming that in the simulating universe, the existence of consciousness causes certain classically describable variables in a computer simulation to change their state, sometimes quite drastically, in a way that is unnatural relative to the laws of physics of the universe in which the computer simulation is taking place, and only natural relative to the logic of the simulation, in particular, to its simulation of the laws of quantum mechanics. Indeed, in the universe in which the simulation takes place, consciousness, at least the kind of consciousness simulated by such simulations, does not cause real physical wave function collapse, if the laws of physics are even quantum mechanical in that universe. Instead, it is an absurd psychophysical law of that universe that consciousness causes complicated variable changes in certain kinds of computer simulations. This is logically possible, but aesthetically absurd. It is sensible to postulate that reality should not be thus. Therefore, assuming reality is not thus, the claim that consciousness causes collapse is incompatible with the possibility that what appears to be reality is being simulated on a computer. But this latter possibility is not so absurd, and it should be considered a virtue of any theory if it is at least compatible with this possibility. Therefore, it is a weakness of the theory that consciousness causes collapse that it is not compatible with this possibility.

Note that the above argument is only an argument against the purely ontological version of the doctrine that consciousness causes collapse, not against any explanatorily epiphenomenalistic doctrine involving purely psychical states causing collapse through their interaction with matter. For if there were purely psychical states that interacted with matter and played a part in wave function collapse, these states would be modelled in the simulation, and their causing collapse would be a natural law according to the logic of the simulation, not according to the nature of the universe in which the simulation takes place. This would not be absurd. It would be an illustration of the general doctrine described above that purely psychical states, if they exist, can only interact with physical states according to the metaphysical structure of representation, and not directly with the physical properties of the proximal material

state. This means that a natural principle about how purely psychical states should interact with matter is compatible with the possibility of skeptical scenarios. The absurdity only arises when one tries to claim that explanatorily epiphenomenal consciousness nevertheless has ontological causal efficacy.

Analogously to the train of argument in the above example with wave function collapse, in general any ontological doctrine that explanatorily irrelevant phenomenal consciousness has ontological causal efficacy on matter means that in a skeptical simulation scenario, the psychophysical laws of ontological causation in the more fundamental universe are aesthetically absurd. This means that all such ontological doctrines are incompatible with the possibility of reality being simulated on a computer. Given that such skeptical scenarios should be admitted as possible, this is a weakness of such ontological theories.

A summary of the above argument is as follows. If reality is being simulated, then the ontological theory of causation makes an absurd claim about the ontological laws of causation of the simulating universe. Therefore, such ontological theories are incompatible with reality being simulated. But this latter scenario should be admitted as a possibility, and it is a weakness of a theory if it is incompatible with the possibility. Therefore, it is a weakness of such ontological theories that they are incompatible with this possibility.

On a Weak Test for Ontological Hypotheses

The thought experiment of the last section considers the scenario in which the reality of experience is being simulated. This section considers the scenario in which the reality of experience is the simulating reality. Thus, it takes the thought experiment upwards rather than downwards in the level of simulation, so to speak. This yields, not merely a purely ontological thought experiment, but a method of conducting a very weak empirical test of any ontological hypothesis of mental causation. The method is as follows.

Suppose one has some ontological hypothesis about mental causation. For example, the hypothesis might be that consciousness causes wave function collapse. Suppose one creates a simulation of a physical reality in which there exists an automaton that, if it were realized physically, would have consciousness correlated with it. Suppose this consciousness would be of such a kind

that it would cause something to happen according to the ontological hypothesis, and a purely material automaton would supposedly behave differently. According to the principle of organizational invariance, this simulated automaton would have the same mental phenomena correlated with it that it would have if it were realized as an ordinary automaton in the simulating universe. According to panpsychism, this computer simulation would only need to simulate a relatively simple automaton for consciousness to arise. This means the test could be realized in practice.

There are two main kinds of experiments that can be run on such a simulation. One can attempt to run the simulation according to the simulated laws of physics assuming the simulated automaton is dual, or one can attempt to run the simulation according to the different laws of physics that would supposedly result from the absence of mental causal efficacy if the automaton were purely material. In either case, there are two main possibilities. Either the simulation could behave as it is programmed to behave, or it could exhibit altered behaviour corresponding to its variables changing so that the simulated reality follows the causation resulting from the alternate ontological assumption. Something else entirely could happen, but this possibility is not meaningful relative to the logical structure of this thought experiment. These two binary possibilities lead to four possible compound outcomes.

Suppose one attempts to run the simulation so the simulated automaton behaves like a real dual automaton, and in practice, the simulated automaton does behave like a real dual automaton. This outcome is consistent with the hypotheses of explanatory epiphenomenalism and of materialism about the simulation. It is consistent with both because the simulation is simply behaving as it is programmed to behave. This outcome is also consistent with the conjunction of the principle of organizational invariance and the ontological hypothesis of causation, because the simulation is behaving as it is supposed to behave if it has a mind correlated with it.

Suppose one attempts to run the simulation so the simulated automaton behaves as a physical purely material automaton would, which, according to the ontological hypothesis, would involve different simulated laws of physics. A physical purely material automaton is here understood to be a purely material automaton in the simulating universe, as opposed to in the simulated universe. Suppose the automaton then behaves as it should according to this ontological hypothesis. This outcome is consistent with the hypotheses of explan-

atory epiphenomenalism and of materialism for the same reason as in the first outcome. This outcome is in tension with the conjunction of the principle of organizational invariance and the ontological hypothesis, because the former principle implies the simulation has a mind, and the latter hypothesis implies the mind would cause the simulation to behave differently. Since the principle of organizational invariance is the more plausible of these two hypotheses, such an outcome would thus cast doubt on the ontological hypothesis. For in order to grant the ontological hypothesis, one would really be granting its reverse. One would be claiming, not that the existence of phenomenal consciousness causes a certain physical effect and that in its absence a different physical law would take effect, but rather that the programming of these different physical laws leads to the presence or absence of phenomenal consciousness.

The above two possibilities are likely the only ones that would be observed in practice. That is to say, the simulation would likely behave as it was programmed to behave according to the laws of physics, and this outcome would not be strange. Since this is by far the most likely outcome, and since it is almost a thought experiment that it could give any other outcome, this experiment, even if realizable in practice, is only a very weak test of any ontological theory of mental causal efficacy.

However, alternative possibilities could conceivably occur, and are worthwhile considering at least as thought experiments to clarify a priori theories of psychophysical causation. Thus, nevertheless suppose that the simulation behaved differently than it was programmed to behave, in a way that was inexplicable given the laws of physics. Practically, the likeliest explanation would be that the simulation was behaving according to the laws of physics, but that one had failed to model properly what this behaviour should be. But suppose hypothetically that one had ruled out practical error with certainty or reasonable confidence, and one was certain or reasonably confident that the simulation was behaving counter to the laws of physics. There are two main possibilities.

Suppose one attempts to run the simulation so the simulated automaton behaves like a physical dual automaton, and instead the simulated automaton behaves like a physical purely material automaton. This would be a bizarre and difficult-to-interpret outcome. Some kind of non-physical causal efficacy would be in effect, but it would be a causal efficacy consistent with the simu-

lated automaton being purely material, at least, according to the ontological hypothesis.

This outcome would be compatible with explanatory epiphenomenalism. According to this theory, one would be forced to posit the existence of purely psychical states responsible for this non-physical causal efficacy. The trouble here is that it might not be practically possible to simulate this psychophysical system on a physical computer, if simulating the physical state alone gives rise according to the laws of nature to purely psychical states, which then evolve and interact with the physical states in a way that is not merely a function of the physical states, or even of the physical states that endeavour to represent psychical states. That is to say, if simulating a physical automaton gives rise to a causally effective psychophysical system, it might not be practically possible to construct a physical simulation of a psychophysical system, because in trying to simulate the psychophysical system, one a fortiori simulates the physical aspect of it, which in turn by assumption gives rise to a causally effective psychophysical system, and the psychophysical states may interact with the physical simulation differently from how the physical simulation of the psychical states do. Further tests under different circumstances would be needed to elucidate further the structure of these states, but a complete simulation might not be possible. However, because of the a priori argument for explanatory epiphenomenalism, this would still not contradict that doctrine.

One could say this outcome would cast doubt on the principle of organizational invariance, because it would suggest the simulation was purely material, because the simulation behaves as a physical purely material automaton would behave. However, on closer investigation, this outcome is inconsistent with the simulated automaton being purely material, because the simulation is not behaving as a purely material automaton should behave. After all, a purely material automaton should simply behave as it is programmed to behave, and the simulation is not doing so. For this reason, this outcome would cast doubt on materialism, at least according to the laws of physics the experimenter was working with. It would require a psychophysical interactionism, which is at first appearance incompatible with materialism as traditionally understood. However, it would still be compatible with a materialism that granted the existence of purely psychical states due to empirical evidence, but claimed that psychical state automata do not have phenomenal consciousness, and that

what a psychical state phenomenal dualist claims is phenomenal consciousness is an illusion.

This outcome would cast doubt on the ontological hypothesis of causation. For such hypotheses posit that the existence of the mind leads to some kind of causation, and its non-existence would lead to indistinguishable behaviour from a purely material automaton, by definition of the notion of non-existence. But the simulation here is not behaving as it is purely materially programmed to behave. One way to try to save ontological causation would be to take this outcome as evidence for a different ontological hypothesis. This hypothesis would be that the simulated automaton has a mind, and that the causal efficacy of this mind affects the simulation so that the simulated automaton behaves as a physical purely automaton would behave if the simulated universe were realized physically. However, on closer inspection, this new ontological hypothesis is seen to be absurd, because it is not naturally compatible with the behaviour of real automata. For real automata in reality are dual and behave as dual automata, by definition of the parameters of the simulation, because the behaviour of dual automata in reality is what informs how the simulation was programmed to behave. In other words, the minds of these real automata do not cause them to behave as if they were purely material. This is a tautology. But note how it contrasts with the new ontological assumption about the simulation. The mind of the simulation has causal efficacy on matter, which, according to the metaphysical structure of the relationship between the mind and matter, causes the simulated automaton to behave as if it were a physical purely material automaton. In the case of the real automaton, which automaton could conceivably be isomorphic to the simulated automaton, the mind causes its automaton to behave as if it were dual. The conjunction of these two facts is unnatural and may be discarded as a possibility. Thus, there is no way to save ontological causation if such an experimental outcome is observed.

It is highly unlikely that such an experimental outcome would be observed in the first place. If it were observed, the best response would be to extend explanatory epiphenomenalism with a system of purely psychical states interacting with physical states, and with phenomenal consciousness correlated with them.

Suppose one attempts to run the simulation so the simulated automaton behaves as a real purely material automaton would, and the simulated auto-

maton behaves instead as a real dual automaton would. Thus, the simulation was behaving so that the simulated automaton was behaving in the way a normal physical automaton would behave, but which the simulation was trying to go against.

This would be a bizarre and difficult-to-interpret result. As noted above, practically speaking one would be more inclined to attribute it to experimental error. However, if this could be ruled out with reasonable confidence, the outcome might be interpreted as follows.

This is the characteristic outcome that would seem to support the ontological hypothesis. However, this would not be evidence that the ontological hypothesis itself applies to reality. It would only be evidence for the aesthetically absurd ontological hypothesis that reality is the sort of universe with the absurd psychophysical laws such that in the simulated inverse, the simulated people could truthfully claim that the ontological hypothesis is valid, relative to how they are able to reference matter. Thus, it would suggest that reality is the sort of universe that nature has determined may run a simulation in which it is correct for the simulated beings to claim that the ontological hypothesis holds. For example, if the ontological hypothesis is that consciousness causes collapse, this bizarre outcome of this experiment would mean that the simulated beings would be correct to claim that consciousness causes collapse. This would not entail that the ontological hypothesis holds for reality itself. For that is a question of the ontology of reality itself, not merely of a reality simulated by reality. But one could interpret the result to suggest that it is more plausible that the ontological hypothesis holds of reality, because it at least means that nature is such that the ontological hypothesis holds for some ontological level that minds can reference according to their metaphysical structure.

Note that this interpretation of this last outcome presupposes the principle of organizational invariance, or at least the hypothesis that the simulated automaton has phenomenal consciousness. Furthermore, note that though this peculiar outcome would give evidence for the hypothesis of causation, the most metaphysically rigorous way to assimilate this hypothesis into the psychophysical laws of reality is not ontologically, but structurally through hypothesizing the existence of purely psychical states that have phenomenal consciousness correlated with them. This latter framework is preferable for the a priori reasons given above in the demonstration of explanatory epiphenomenalism.

On Idealism

A detailed discussion of idealism is not within the scope of this chapter, whose purpose is to establish epiphenomenalism. Idealism is discussed in detail later in this work. However, within the scope of this chapter, it suffices to show that idealism is compatible with epiphenomenalism as these two doctrines are understood in this work.

Idealism is an ontological doctrine about the essence of the substance that underlies the causal structure of physics, whereas epiphenomenalism is a structural doctrine about the causal structure of psychophysics. The two are thus not in conflict, and in fact independent of one another a priori. Thus, it is not necessary to argue for or against idealism in order to establish epiphenomenalism.

On Neutral Monism

A detailed discussion of neutral monism is not within the scope of this chapter, whose purpose is to establish epiphenomenalism. Neutral monism is discussed in detail later in this work. However, within the scope of this chapter, it suffices to show that neutral monism is compatible with epiphenomenalism as these two doctrines are understood in this work.

Neutral monism in its full ontological or aesthetic strength is likely not a viable doctrine. This work endeavours to show this later. The essence of nature may be slanted towards the view of neutral monism rather than substance dualism, but at least a dualism of explanatory relevance is necessary. However, entering into these arguments in detail in this place would distract from the intended train of thought. Furthermore, it is not even necessary to argue against neutral monism in order to establish epiphenomenalism. It is only necessary to show that neutral monism, properly interpreted, is compatible with epiphenomenalism.

One could say that, according to neutral monism, it is an ontic frill of reality that, in some sense, mental phenomena are necessary for causation, because they underlie reality. Thus, they cannot merely be epiphenomena. There is aesthetic justice to this. Perhaps it is a misuse of language to call this state of affairs epiphenomenal. After all, ontology is little more than aesthetics and

arguing over how to use words. And if one wants to make this insistence on the use of the term *epiphenomenalism*, there is no need to argue over the use of words. But it seems this is no longer a debate about the nature of reality. As the author understands things, epiphenomenalism as defined in this work is compatible with neutral monism, and even most forms of what is typically called interactionism. This is because neutral monism and ontological interactionistic dualism are merely ontological frills one can add on to epiphenomenalism according to aesthetic taste.

The important point is that the ordinary causal structure that epiphenomenalism posits remains the same. Any supposedly causally necessary phenomenal properties are still explanatorily irrelevant for modelling the causal structure of the system. The phenomenal properties can still be subtracted from the system to yield identical behaviour.

Thus, even if mental phenomena are interpreted as being of the same substance as matter in some ontological sense, the phenomenal properties of matter are not physical properties, because the phenomenal properties may in principle be subtracted from the physical properties to yield an aphenomenal copy with the same physical properties.

A class of ontological doctrines encompassing neutral monism is discussed later in this work under the name *the mind-correlate doctrines*.

On Ontological Frills

Ontological distinctions like those discussed in this chapter are discussed in detail later in this work. And throughout this work, whenever the discussion is not ontological in the capacity of ontology, and when the author uses dualistic and epiphenomenalistic language because that is his ontological preference, a reader who finds this ontology distasteful to their aesthetic sensibilities will hopefully be able to make the necessary verbal substitutions to reformulate these doctrines in their desired ontological system, be this ontological interactionistic dualism, idealism, neutral monism, or something else entirely, while preserving the logical, structural, and cognitive meaning of these doctrines.

On Genius

An objector may claim that the existence of genius is evidence for the existence of mental causal efficacy. Works of genius are so sophisticated that perhaps no purely material process in the brain is sufficient to produce them. William James advocated this view, as did Roger Penrose.

A response to this is as follows. Perhaps purely psychical states exist and their behaviour is involved in some way in the mechanism underlying genius. This does not appear likely, but the explanatory epiphenomenalism of this work is compatible with the existence of purely psychical states, so this is no issue here.

An objector may claim that the existence of genius is evidence, not only for the existence of purely psychical states, but for the existence of ontic causal efficacy of phenomenal consciousness itself. For the behaviour of purely psychical states is within the realm of the laws of nature, whereas the pure freedom of consciousness itself is beyond them, and the latter alone is sufficient to explain the mystery of genius.

A response to this is as follows. Nature is reality, and thus the laws of nature are the laws of how things occur in reality. But genius occurs in reality. Therefore, describing the occurrence of genius must fall within the realm of the laws of nature. For even an ontological hypothesis about mental causal efficacy, even a non-deterministic one, is a hypothesis about a natural law, just a very different and likely chimeric sort of natural law from typical physical or even psychophysical laws. But this is only an argument about how to use words.

A more meaningful response to the objector that responds to the intended meaning of the objection and not just how the objection uses words is as follows. Since consciousness exists, genius has a moral significance beyond any scientific model of it. That is, the fact that the universe exists in the way that it does with genius in it rather than having existed in some other way without genius being in it is morally significant, and perhaps not an accident. If this is all the objector claims, then there is no disagreement between them and this work. But such a moral significance is sufficiently vague that it is compatible with epiphenomenalism, and it is thus no evidence against it. The objector must further claim that it is impossible or dubious to claim that genius could result from the activity of a purely material brain, or from the activ-

ity of some other automaton with phenomenal consciousness subtracted. But this further claim is false, for there is good reason to believe that this is possible, and indeed the only plausible option. Now, a detailed discussion of the nature and causes of genius would be more within the scope of a work of ethics than of metaphysics. But for the purpose of this work, suffice it to say that genius is best understood as resulting from a combination of study, the innate abilities of the brain, and the restraint of self-expression in the short term to create work of significance in the long term. These factors apply as well to the purely material functioning of the brain of an aphenomenal copy as they do to the functioning of a brain with phenomenal consciousness correlated with it, so they do not suggest that one should posit some further mysterious force responsible for genius. Therefore, these factors lead to genius as a purely material process and as an epiphenomenal mental process. That is to say, this doctrine applies to the behaviour of a purely material human being as well as it models the behaviour of a dual human being, or any other automaton with genius, because this doctrine in no way references mental phenomena or their behaviour as causally distinct from their material correlate. Thus, genius can in principle be modelled as resulting from the purely material configuration of the brain, so genius is no evidence for the existence of mental causal efficacy.

On the Testimony of Genius

Several great philosophical geniuses of the modern era did not believe genius is evidence for the existence of mental causal efficacy. In fact, the author has not encountered this argument in any writings that he has investigated written by the philosophers whom the test of time and the consensus of posterity witness to be the greatest geniuses. He has only encountered the argument in the writings of philosophers who are less uniformly celebrated, or whose fame has only had to endure for a shorter time, and who hold the greater philosophers in awe somewhat, and see their creative process as an opaque mystery. And if great philosophers with firsthand acquaintance with genius did not see anything in their own experience of this process to take it as evidence for the existence of mental causal efficacy on matter, then this suggests that genius is not evidence for mental causal efficacy, using the little data available on this rare process.

An objector may claim that this argument ignores important data. For some great philosophers have believed that mental causal efficacy exists. For example, René Descartes and George Berkeley believed this.

A response to this is as follows. This objection misses the point of the argument. The argument is not whether one can find respectable authority for one side or the other of the argument over whether mental causal efficacy exists. For one can find such authority for both. Ultimately, authority alone is not a good reason to believe anything in philosophy. The point of the argument is whether any of those with first-hand acquaintance with genius believed that genius was evidence for the existence of causal efficacy. And it does not seem that any did. At least, the two philosophers listed above as supposed counterexamples did not claim this. Rather, they claimed mental causal efficacy existed based on other arguments. Thus, this is not an argument by authority, but an argument by externally reported first-person data. The data are necessarily few because genius is a rare phenomenon.

An objector may claim that at least William James believed that genius was evidence for mental causal efficacy, and William James was a great philosopher.

A response to this is as follows. William James made this argument in the chapter The Automaton Theory of his work *The Principles of Psychology* by appeal to the genius of William Shakespeare. He argued that it was absurd or at least problematic to claim that the authorship of a play by Shakespeare could be entirely explained by the workings of his brain, without any causal efficacy ascribed to the mental phenomena accompanying this act.

William James discusses the genius of William Shakespeare with great respect, as if referring to an obscure process whose workings he does not understand. This is in accordance with, and in fact an example supporting the argument that it is only writers who have no first-hand experience with the level of genius about which they theorize who believe this genius is evidence that there is something mysterious in it requiring more than brain activity. Indeed, being somewhat facetious here, the author has not encountered any passage in the works of William Shakespeare himself where that writer claims that his or any other genius is evidence for the existence of causal efficacy of phenomenal consciousness on the material universe. However, this should not be surprising, since that writer did not much concern himself with metaphysical invest-

igations, and the doctrine of epiphenomenalism had not even been precisely formulated by his time.

An objector may claim that this sort of argument may lead to a situation analogous to the fable of the emperor who was tricked into pretending and even believing that he was being given clothes that only wise people could perceive, because he did not want to admit that he could perceive nothing of the sort. For if only the supposedly great philosophers do not believe that genius is evidence for the existence of mental causal efficacy, but the supposedly lesser ones may believe this, then one may perhaps merely advocate the former opinion in order to associate oneself with the former group.

A response to this is as follows. Suppose the argument went something like the following. *Any genius knows by introspecting on their own cognitive processes that genius is not evidence for mental causal efficacy, because if one is a genius, the fact is obvious.* In that case, the vanity of the gullible might tempt them to agree with such an arrogant sophism, saying, *Yes, I too see how obvious it is.* However, the argument that this work advocates is different, because it appeals to objective rather than subjective evidence. It would be absurd to claim that philosophers such as René Descartes, David Hume, or Immanuel Kant failed to claim that genius is evidence for mental causal efficacy merely to suggest implicitly by this omission that they were great philosophers. Thus, the objection does not apply to the empirical data to which the argument appeals. In other words, the argument appeals to objective examples in the past that most now agree were indeed examples of genius, and there is thus little temptation to self-referential arrogance, which would exist were the argument based on subjective data.

On Free Will

Free will is the ability of the mind to act in a way not governed by necessity or chance, but by choice.

It is not known if free will exists. The definition of free will is more vague than that of epiphenomenalism. According to some interpretations, the existence of free will is compatible with epiphenomenalism. Most of the arguments of this work make no assumptions about whether free will exists.

On Free Will and Mental Causal Efficacy

The existence of mental causal efficacy and the existence of free will are independent. A demonstration of this is as follows. Suppose the mind has various extra mental phenomena that have no correlate in the automaton, but the behaviour of all mental phenomena, even of the purely mental phenomena, is predetermined by God or by some non-anthropomorphic law of necessity. Then purely mental phenomena exist, but free will does not. Next, suppose that the body can be modelled as an automaton and the complete description of its behaviour relies in no way on the state of the mind, and that the mind can be modelled as a function of the body, but suppose further, if it is granted that this is possible or even meaningful, that according to the underlying nature of mental existence minds are free to make choices nevertheless, and that the purely material model of the behaviour of the body is merely an adaptive framework giving structure to the reality in which minds exist and act freely. It is not here claimed that this is the case, or that it is even necessarily possible. It is only claimed that the imprecision in the definition of free will means that it is conceivable in some way, according to some interpretations of the meaning of free will. In this case, free will exists, but purely mental phenomena do not. Next, suppose that purely mental phenomena exist and that it is through them that free will exists. Then both exist. It is possible that both could exist without purely mental phenomena being what enables free will. Lastly, suppose that purely mental phenomena do not exist, and that there is no deeper nature to mental existence other than to be correlated with an automaton. Then neither exists. Therefore, the existence of free will and the existence of purely mental phenomena are independent.

As for which of them, or both, or neither, exists in reality, this is left undecided by the above argument. Later in this work, using interpretive dualism, an argument is put forward that purely mental phenomena do not exist, and a weak argument suggesting that free will could exist.

If free will exists, it involves the alteration of habits. A demonstration of this is as follows. Free will, if it exists, is not a physical process that physical laws can describe. It is according to the human condition. Therefore, it must have moral significance. Therefore, it must concern virtue. But virtue is the regulation of behaviour over time using the faculty of reason. And this regula-

tion is effected through the gradual alteration of habits. Therefore, free will, if it exists, involves the alteration of habits. This needed to be demonstrated.

An objector may claim that the alteration of habits may be detectable in a short period of time, because there are sudden moments when a person decides to change the course of their life significantly, and subsequently does. Such results may be the result of free will if free will exists. This would contradict the claim that free will, if it exists, can only exert a negligible influence on behaviour during a short time.

A response to this is as follows. Such sudden resolutions precede, and are the superficial preface to, a longer span of time during which the habits are gradually altered. A person may experience a sudden feeling of inspiration to change their habits, but subsequently fail to do so, indicating that it is not the feeling of inspiration in itself that is essential here, but the gradual process of change as a whole. This shows that the feeling itself is not what causes the alteration of habits, but the long span of time during which the habits are gradually altered.

On Free Will, Determinism, and Causal Efficacy

Determinism is the doctrine that the state of the universe at one time determines its state at all other times.

Indeterminism is the negation of determinism.

The existence of indeterminism, free will, and the causal efficacy of purely mental phenomena are often associated with each other, but they are logically independent. For the existence of indeterminism concerns modelling the evolution over time of the universe, whereas free will has a vague definition that can be made compatible with many other things, and the existence of purely mental phenomena concerns the existence of a certain kind of substance. A more rigorous demonstration of their independence must show the possibility of all eight independent combinations of their existence or non-existence.

Such a demonstration is as follows. Each of the examples given is a sufficient condition for, though not always a necessary characterization of, the combination considered.

It is possible that none of indeterminism, free will, or the causal efficacy of purely mental phenomena exist if there is no choice in reality, and the mind is merely a correlate of the body, and the evolution of this state of affairs is determined in advance by any particular state of the universe.

It is possible that only indeterminism exists if the body and mind are as specified in the preceding example, but there exists some indeterminism of some kind in the laws of physics, for example, some fundamental randomness not merely the result of practical uncertainty.

It is possible that only free will exists if a definition of free will is adopted that is sufficiently weak for it to be compatible with determinism and the non-existence of purely mental phenomena. For the definition of free will is sufficiently vague that this can be done, for example, if free will is taken to be a moral freedom to actions giving significance to human experience independent of, and compatible with, any particular scientific model of the evolution of reality over time.

It is possible that only causal efficacy of purely mental phenomena exists if the evolution over time of these purely mental phenomena and of matter is predetermined and if there is no freedom, not even of the vague kind.

It is possible that only indeterminism and free will exist if there is some indeterminism in some physical process unrelated to the function of the body, and if free will is not effected through purely mental phenomena, but is only a moral significance to life.

It is possible that only free will and purely mental phenomena exist if the evolution over time of purely mental phenomena is predetermined, and there is moral significance to choice unrelated to the existing substance of purely mental phenomena that happen to have causal efficacy on matter in a deterministic way.

It is possible that only the causal efficacy of purely mental phenomena and indeterminism exist if the evolution over time of purely mental phenomena is random and reflects no choice.

It is possible that all three exist if free will is effected through the causal efficacy of purely mental phenomena, and that the effect of free will is what leads to indeterminism.

Such are the possibilities. However, this work is not equally sympathetic to all of these possibilities.

This work positively advocates epiphenomenalism as valid a priori. It claims that purely psychical states likely do not exist. It claims that purely mental phenomena likely do not exist because purely psychical states likely do not exist, and because if purely mental phenomena did exist, the only coherent option would be for them to be the correlates or intrinsic substance of purely psychical states. This work is relatively sympathetic to determinism. At least, any apparent indeterminism in reality may be explained as practical indeterminism resulting from the chaotic amplification of small differences in the initial state, perhaps even from mechanisms too fine to be modelled yet using current methods. This possibility of determinism applies even to quantum measurement. That is to say, the apparent indeterminism of quantum measurement could very well be the result of an underlying deterministic process. This work is relatively sympathetic to the existence of free will, though in a way that is mostly independent of how it is sympathetic to determinism. And in any case, the existence of free will must be compatible with epiphenomenalism, and is likely compatible with determinism. Free will at least exists in the sense of people being practically free in the sense of everyday life. And free will may very well exist in a more subtle and more substantial sense that is nevertheless compatible with epiphenomenalism and even determinism.

On Judgements about Free Will

Since the existence of free will is independent of the truth of epiphenomenalistic interactionism, it is possible that judgments about free will are mostly or entirely independent of free will itself, even if free will exists, though most plausibly in the case where it does not. It is possible that all four possibilities of the existence or non-existence of free will and the presence or absence of judgements affirming the presence of free will are possible, and there may not even exist any correlation or causal connection between the two.

On the Coherence of Purely Material Automata

Though the above discussion is sufficient to establish epiphenomenalism for the purposes of this discussion, as a concession suppose that one believes for

some reason that mental causal efficacy exists, and that one therefore doubts whether the behaviour of purely material automata would be coherent.

Thus, it may be objected by an interactionistic dualist who denies the validity of the arguments in this chapter, or by anyone who doubts or denies epiphenomenalism for some reason, that if epiphenomenalism is false, then even if the purely material universe is conceivable or possible, the behaviour of automata in the purely material universe is unnatural, and thus cannot be interpreted without positing the existence of hidden variables corresponding to the causal efficacy of the mind. This would mean that thought experiments about the purely material universe would have little significance.

A response to this is as follows. If mental causal efficacy exists but is not responsible for the reference of the mind, then its existence does not impede materialistically interpreting the reference of the mind. Therefore, the objector must claim that mental causal efficacy is responsible for the reference of the mind.

The response to this stronger claim is as follows. If mental causal efficacy exists, it exerts a negligible effect on behaviour over a short period of time, because it has never been detected empirically. However, the reference to the existence of the mind is a detectable behaviour that takes place over the course of a few seconds. Therefore, mental causal efficacy cannot be directly responsible for the reference of the mind. Therefore, this behaviour results from the material configuration of the nervous system. Therefore, there must exist a materialistic interpretation of the reference of the mind, even if mental causal efficacy exists. And since all individual metaphysical doctrines, once learned, can be recalled and understood in a short period of time, similarly, there must exist a materialistic interpretation of all of metaphysics, even if mental causal efficacy exists.

In response to this, the objector may refine their position as follows. Let it be granted that it is impossible for mental causal efficacy to directly generate a particular instance of the reference of the mind. But it could gradually create in the automaton the material structure required for the ability to reference the mind. If this process of generation were sufficiently slow, then it would not contradict any empirical evidence. And once mental causal efficacy had generated this structure, the exercise of the ability to reference the mind in a short time could be attributed to only material causes, but there would not exist a

materialistic interpretation of metaphysics, because the only reason the structure existed in the first place would be the causal efficacy of the mind.

The response to this refined objection is as follows. The position of the objector has become contrived. Objective considerations do not justify it. All that justifies it is the desire to object to the original argument.

Also note that the objector is claiming that there exists mental causal efficacy through which free will is not exerted. For all are forced by necessity to acknowledge that their mind exists, because it exists. That which one is forced to do by necessity is not a choice. Therefore, the reference of the mind is not a free choice. Specifically, it is not a free choice to reference that the mind exists as opposed to claiming that it does not, because no rational being can deny that they think, though it could be a free choice whether to reference its existence at any particular time, or to deny its existence irrationally. Therefore, the supposed mental causal efficacy through which a person supposedly references the existence of the mind is necessary, and not free. This is not impossible. But it makes the position of the objector more contrived, because it is natural to suppose that mental causal efficacy and free will are connected, if they exist.

Thus, even supposing as a concession that mental causal efficacy exists, though it has been shown that it does not, it is still compatible with the existence of a materialistic interpretation of dualistic behaviour.

On Correlation

Introduction

This chapter and the next chapter elaborate on the consequences of the last chapter. Thus, this chapter and the next build a general metaphysical theory of automata and their behaviour according to epiphenomenalism.

This chapter discusses how mental phenomena are correlated with automata. Of fundamental importance to this discussion is the notion of the material correlate of a mental phenomenon, which is the minimal material system that is sufficient to give rise to that mental phenomenon. The discussion of the material correlate is made from a philosophical rather than scientific perspective. The focus is on stating abstract doctrines justified by a priori reasoning rather than discussing the latest theories that try to map the anatomy of the material correlate precisely. This is because the doctrines of this work are not much concerned with the precise anatomy of the material correlate, and because attempts to determine this anatomy tend to become outdated as knowledge advances.

On the Ontological Assumptions of this Chapter

The doctrines of this chapter are implicitly phrased from the perspective of dualism. However, this should only be interpreted as a linguistic convention for ease of expression, and not entirely as a dogmatic stance on the nature of reality. This work is ultimately dualistic, and arguments for dualism have already been put forward. However, not all of them have yet been put forward, and ultimately it is impossible to conclusively establish dualism as opposed to idealism or neutral monism. Thus, though the doctrines of this chapter, the next, and of most of this work are phrased according to epiphenomenalistic dualism, with some superficial changes they may be recast to these other viable ontological frameworks. For doctrines about the correlation between the mind and the body do not need to presuppose dualism. If the mind is material, or if matter is phenomenal, and metaphysical correlation thus turns out to be identity, then the doctrines all hold in a degenerate or vacuous form. Whether they hold in a degenerate form or not remains to be determined with confidence. This work argues later that they do not hold in a merely degenerate form. But assuming or demonstrating this is not strictly necessary at this stage of the discussion.

On the Material Correlate

The material correlate of a mental phenomenon is the minimal material system that is sufficient to give rise to that mental phenomenon according to the laws of nature.

The material correlate, mentioned by itself rather than as being the material correlate of a particular mental phenomenon, should be interpreted to be the material correlate of the whole mind, unless otherwise specified.

The material integration is the complement of the material correlate in the body.

That is to say, the material integration is the part of the body that is not a part of the material correlate.

The body is not the minimal collection of matter correlated with the mind. That is, the body is not the material correlate, and the material integration is non-empty. Evidence for this is that peripheral parts of the body may be

removed without parts of the mind being removed. At least, any changes that these removals have on the mind only take place because the removals influence a part of the body more directly correlated with the mind. Thus, the correlation of the mind with the material integration takes place indirectly through the correlation of the material integration with the material correlate of the mind.

What is called in this work the material correlate may also be referenced using the plural term *material correlates*. The distinction is only in the use of language. The material correlate, considered as extended, may be called several material correlates, and the material correlates, considered as unified, may be called one material correlate. The singular term is preferred here over the plural because of the frequency with which the term is used in this work. It also avoids mistakenly giving the impression that the material correlate can necessarily be broken down into discrete material correlates that cannot be broken down further for the purpose of analyzing their structure with respect to innate intuition. For it is not known for certain whether this is the case.

On Divisions of Mental Phenomena

The above definition of the material correlate is more general than it may at first appear. For it may be applied to define the material correlate of some aspect of what is traditionally considered to be one mental phenomenon. For example, the pitch and loudness of a sound are two different aspects of the sound, and thus may with some justice, at least for the purpose of generalizing the definition, be called two different mental phenomena in their own right, and thus each may have its own material correlate.

An objector may claim that there is a distinction between different mental phenomena forming part of one mind, and one mental phenomenon, which may very well even be compound, having different aspects. Certain divisions may be called different mental phenomena, but the aspects ought not to be called mental phenomena.

A response to this is as follows. There is no need to argue about words. If the objector prefers such a use of language, then to generalize the definition of the material correlate, one need only say that the material correlate of an aspect of a mental phenomenon is the minimal part or aspect of the material

correlate of the mental phenomenon that is sufficient to give rise to that aspect of that mental phenomenon.

However, though there is no need to argue about words, if one were to argue about words, one might say the following. The distinction of the objector may be chimeric and overly biased towards the subjective intuitions of space and time. If the objector allows that two parts of the field of vision may be called two different mental phenomena, but does not allow that two aspects of a sound perception may be called two different mental phenomena, they are assuming that the only real way to divide mental phenomena into other mental phenomena is according to the intuition of space, and perhaps also of time. But subjective space and time are not more ontically fundamental than the other intuitions. They are only more practically fundamental, because objective space and time are more ontically fundamental than the material correlates of the other intuitions. And it is also easier to define divisions according to these intuitions, both when analyzing mental phenomena introspectively and when analyzing the material correlate scientifically. But strictly speaking, even for space, it is not possible for one part of the field of vision to exist without the other parts existing, because they are parts of one consciousness. If one part did not exist, that particular consciousness would not exist. Thus, if that particular consciousness exists, then all of its parts exist. It is only possible for a different mind to exist that had a mental phenomenon in many ways qualitatively identical to that part of the field of vision which was a part of a larger mind. And a similar remark may be made for time. Thus, even divisions based on subjective space and time only specify different aspects of one mind, not different mental phenomena. Thus, either one can divide mental phenomena into other mental phenomena according to any intuition, that is, any structural distinction, or one cannot divide them into mental phenomena according to any intuition, and one must instead speak about aspects of the whole mind in all cases. At least, this is how the author would use language. This is only an argument about how to use words, not about the nature of reality.

On Dual and Purely Material Automata

A dual automaton is an automaton that has a mind metaphysically correlated with it.

A purely material automaton is an automaton that does not have a mind metaphysically correlated with it.

The material correlate of a dual automaton is the minimal collection of matter required to model the structure of its mind.

A dual automaton is an automaton that has a mind correlated with it. Its mind is such that it is referenced by its innately dualistic behaviour. A purely material automaton has no mind correlated with it. The material correlate of the mind of a dual automaton is the minimum collection of matter correlated with the mind of the automaton. It may also be called the material correlate of the dual automaton. The dual automaton can be interpreted as the material system that has a mind correlated with it, or it can be interpreted as the dual system of both the material system and its correlated mind. The two meanings are only distinctions in a convention in the use of language, and either meaning can be used depending on what is convenient.

An automaton may be composed of automata. Automata that compose an automaton retain their independent existence as automata in their own right. If an automaton is composed of automata that have minds, the automaton is dual if and only if there exists a mind correlated with the whole automaton. The existence of the mind of an automaton composing the compound automaton is not sufficient for the compound automaton to be considered dual. Thus, for a given automaton, there is a compound or global material correlate, which is to be regarded as the material correlate for that automaton properly speaking, as opposed to any of the smaller material correlates of any of its constituent parts.

The material correlate of a dual automaton is the minimal collection of matter that is directly correlated with the mind of the automaton.

On the Principle of Organizational Invariance

The principle of organizational invariance is the doctrine that automata with functionally identical organization have qualitatively identical mental phenomena.

The principle of organizational invariance is not logically necessary a priori. It is possible that there could exist a universe where functional isomorphs had different mental phenomena. A well-known example of such a

possibility is of functional isomorphs with inverted colour perceptions. However, it is plausible that the principle of organizational invariance holds in reality. That is to say, this principle is likely one of the laws of nature, though not one of the theorems of logic.

The principle of organizational invariance shows that it would be an inaccurate characterization of the material correlate to say that it is the material system that is necessary and sufficient to give rise to the mental phenomenon. For, according to that principle, an exact copy of the automaton, or even a functional isomorph of that automaton, would give rise to a qualitatively identical mental phenomenon. Thus, the existence of one particular material correlate is not a necessary condition for the existence of its correlated consciousness.

These examples also show that the material correlate, even when defined properly, is not necessarily unique. For if there does exist an exact copy or functional isomorph of the automaton somewhere else, then it is not clear which material correlate should be called the material correlate. Perhaps in such a scenario two qualitatively identical but numerically distinct mental phenomena exist, and the correspondence may be defined by saying that one mental phenomenon is that of one material correlate and not of another because it is the one that would disappear if that material correlate stopped existing but not if the other stopped existing. However, such a theory may be chimeric. Perhaps all qualitatively identical mental phenomena are numerically identical. However, in the real world, instances of duplicates and functional isomorphs do not occur much or at all, for human minds or minds of reasonably large complexity at least, so it is usually valid to speak about the particular material correlate of a mental phenomenon in such discussions. It is dubious whether the non-uniqueness of the material correlate extends to cases other than functional isomorphs. In this work, the material correlate is assumed to be practically definite, and its possible non-uniqueness is only brought up if this is relevant to the discussion.

In general, the intrinsic properties of mental phenomena only specify the physical properties of their material correlate insofar as they specify its metaphysical structure. In particular, a mental phenomenon cannot specify the position, mass, charge, or any other of the fundamental physical properties of its material correlate. Nor can it have any property that is a physically simple function of any of these, for metaphysical structure is not a physically simple

function. Durations of time defined mentally cannot intrinsically specify any physical duration of time.

A demonstration of this is as follows. The claims of the last paragraph follow from the principle of organizational invariance and the possibility of skeptical scenarios. Suppose the physical structural organization of an automaton is realized in two different ways. Suppose the material correlate of one realization has a different mass from that of the second realization. For example, one is an ordinary living being, and another is a computer simulation. However, they are functionally isomorphic, and thus have qualitatively identical mental phenomena. But these qualitatively identical mental phenomena are correlated with systems of different mass. Therefore, the intrinsic properties of mental phenomena do not characterize the mass of the systems they are correlated with. The same applies to position, and any other physical attribute. Similarly, one physical realization of an automaton may proceed at a different rate than another, but if the two are isomorphic, they lead to qualitatively identical mental phenomena. Therefore, the perception of intervals of subjective time in memory does not characterize the interval of objective time taken up by their material correlate, if the apparent memories even represent past mental states at all. This needed to be demonstrated.

However, the principle of organization invariance is not even strictly necessary for the negative part of the above demonstration. Philosophical skepticism rejects that any one thing in reality can necessarily determine the state of any other thing. However, the principle is necessary for the positive part of the demonstration, which makes a claim concerning how mental phenomena do specify some aspect of the material correlate according to the laws of nature. In any case, the proposition holds.

All of the examples used in the above demonstration are skeptical scenarios in which the mind is deceived or ignorant about the physical structure of its material correlate. If it is assumed that the mind has accurate knowledge of its material correlate, then in the context of reality the mind represents the physical properties, though in itself it does not specify them, because representation can only be defined relative to the circumstances. In general, it is assumed in this work that the mind may have accurate knowledge about its material correlate, though the mind in itself does not characterize the physical structure of the material correlate.

On the Principle and Psychical State Automata

In the chapter On Epiphenomenalism it is hypothesized that it might be natural to call a causal state *purely psychical* if it has properties that are not typically physical. The discussion of the last chapter focuses on the case where the automaton lacks purely psychical states, which case is the typical focus of this work, because it is likely that purely psychical states as defined in this work do not exist. Furthermore, the principle of organizational invariance applies most naturally to automata without purely psychical states.

However, this principle also applies to those with such states insofar as it constrains the nature of their mental phenomena. For even purely psychical states are by definition the structure of mental phenomena, and if these mental phenomena must be qualitatively identical for functionally isomorphic realizations of an automaton, a fortiori their purely psychical states must be causally equivalent. This does not mean that the physical states give rise to the purely psychical states as epiphenomena. For this would contradict the definition of purely psychical states. But if the physical states in two different realizations are isomorphic over an extended period of time, this does constrain the interaction such that the purely psychical states must be the same in the two automata. It is possible that if the physical states were identical at one time, they could have different psychical states that led to different evolution over time. And if the two automata have the same psychical states, then their mental phenomena have the same structure. And the principle of organizational invariance posits the further doctrine that these mental phenomena are qualitatively identical, provided that functional organization is interpreted to encompass the organization of physical and purely psychical states.

The principle of psychical interaction is independent of the general principle of organizational invariance. The former concerns how different types of states can interact, and the latter concerns how phenomenal consciousness depends on the causal organization of reality.

On Correlation

Statistical correlation between two things is the state of affairs in which knowing something about one of them gives one information about the other.

Metaphysical correlation is the relation between a mental phenomenon and its material correlate.

Metaphysical correlation is a structural isomorphism set by a law of nature between the mind and its material correlate.

The metaphysical structure of the material correlate of a mind is the abstraction from its complete physical structure determined by embedding the structure of the mind in its physical structure according to the isomorphism of metaphysical correlation.

All metaphysical correlation is statistical correlation, but not all statistical correlation is metaphysical correlation.

Metaphysical correlation is always statistical correlation, because knowing something about one thing that is in metaphysical correlation with another provides much information about the other, because there is a structural isomorphism between them. The knowledge of the correlate that is gained is not complete, because metaphysical structure abstracts from some aspects of physical structure. For example, the mind, from knowing how it itself is, knows that the metaphysical structure of its metaphysical correlate is the same. But this does not mean the mind knows the full physical structure of the metaphysical correlate, because the metaphysical structure abstracts from some of the information of the complete physical structure. Furthermore, the mind does not know the substance underlying the physical structure of its metaphysical correlate. Conversely, knowing the complete physical structure of the metaphysical correlate of the mind, and even its essence, were this even possible, does not entail knowledge of what it is like to be the mind correlated with it, because situations are logically possible, though likely naturally impossible, in which functional isomorphs, and even physical copies, have structurally isomorphic but subjectively different qualia correlated with them.

Statistical correlation is not always metaphysical correlation. For example, the mind is statistically correlated with the contents of the stomach, because when the stomach is full of food, the mind tends to feel a sensation of fullness, and when the stomach is empty of food, the mind tends to feel hunger. However, the stomach is not a metaphysical correlate of the mind. The statistical correlation between the mind and the stomach takes place indirectly through the physical interaction between the stomach and the metaphysical correlate of the mind, and the metaphysical correlation between the metaphysical correlate of the mind and the mind itself.

In this work, when discussing the correlation between the mind and matter, the term *correlation* is understood to refer to metaphysical rather than merely statistical correlation unless otherwise specified.

On the Neural Correlate

The neural correlate is the part of the material correlate that is composed of nervous tissue.

The neural integration is the complement of the material correlate in the nervous system in the case where the automaton has a nervous system.

In general, the neural correlate, if it exists, is not necessarily the same as the material correlate. The neural correlate must be a subset of the material correlate. However, if the material correlate is not entirely composed of nervous tissue, then the neural correlate is a proper subset of the material correlate. This could be the case for a cybernetic organism whose mind was partly the metaphysical correlate of computer activity and partly that of neural activity. Note that this would not be the case for a cybernetic organism that only had computer systems directly interfacing with parts of its peripheral nervous system, with parts of its brain not metaphysically correlated with the mind, or even interfacing directly with its neural correlate but not in the right way. To form part of the material correlate, the computer states would need to be in a close and unified functional interface with the neural correlate globally available for reporting and control. This functional criterion is discussed in more detail later in this work. In most reasonable circumstances, and likely all currently existing circumstances, the neural correlate, if it exists, is the same as the material correlate.

In general, the neural integration, if it exists, is not the same as the material integration. It is always a subset of the material integration. The material integration encompasses the rest of the body apart from the material correlate, whereas the neural integration only encompasses the rest of the nervous system. They would only be the same if the complement of the material correlate in the body is entirely composed of nervous tissue. This would only be the case if the body of the automaton were defined to consist only of nervous tissue and perhaps non-neural correlates of the mind forming the complement of the neural correlate in the material correlate. This could be the natural defini-

tion of the body in a skeptical scenario in which a person is a nervous system in a vat being electronically stimulated to have illusory experiences. But it would be an arbitrary and unnatural definition of the body for all living animals in ordinary circumstances. Therefore, in most circumstances, and likely all currently existing circumstances, the neural integration, if it exists, is a proper subset of the material integration.

All the general doctrines on the material correlate and on metaphysical correlation, framed irrespective of whether the automaton has a nervous system, also apply to the neural correlate when the neural correlate is the same as the material correlate. However, further doctrines may be stated about the neural correlate based on the further assumption that it is composed of nervous tissue, or based on assumptions of what animal one is discussing.

The nervous system is in more direct correlation with the mind than is the rest of the body of an animal. That is, the neural correlate of an animal is a subset of its nervous system.

The neural correlate is properly contained in the nervous system. Evidence of this is that memories not actively being remembered are stored in the nervous system in some way, but are not part of the neural correlate. Therefore, they must be stored in the neural integration.

The neural correlate is a collection of events, not a collection of enduring matter. An atom that is in the neural correlate at one time may not be in it later. Evidence of this is that consciousness disappears mostly or entirely during dreamless sleep, while the atoms that were in the neural correlate remain in the brain. This evidence also establishes the same doctrine about the material correlate, for the neural correlate is also the material correlate in this scenario.

In the remainder of this work, the following conventions in the use of language are followed. If the neural correlate is discussed, it is assumed to be the same as the material correlate, unless otherwise indicated. If the neural integration is discussed, it is assumed to be a proper subset of the material integration, unless otherwise indicated. If a material correlate is discussed in the capacity of being the correlate of mental phenomena in general, it is referred to as a material correlate rather than as a neural correlate. The term *neural correlate* is used when the assumption that the thing being discussed is composed of nervous tissue is relevant to the discussion. The neural correlate may also be

referred to using the plural term *neural correlates*, but the singular term is preferred in this work.

On the Local Structure of the Neural Correlate

Inactive nervous tissue, or even nervous tissue that is active but not in the right way, has no mental phenomena correlated with it, at least, not of the kind present to the human mind. But nervous tissue active in the right way has mental phenomena correlated with it. Thus, the neural correlate of these mental phenomena is not the nervous tissue as a whole, but rather the firing of the nervous tissue insofar as it is correlated with these mental phenomena. If a nerve cell is part of the neural correlate at a given time, this does not mean that the whole nerve cell is contained in the neural correlate. It only means that the firing of the nerve cell insofar as it contributes to the logical architecture of the brain is contained in it. Thus, in the remainder of this work, if it is said that some macroscopic part of the nervous system is a part of the neural correlate, this is an imprecise shorthand for saying that the firing of the nerve cells insofar as it contributes to the logical structure of the brain in this part, insofar as this logical structure is isomorphic to the structure of the mind, is contained in the neural correlate.

On the Global Structure of the Neural Correlate

The peripheral nervous system is not a part of the neural correlate. A theoretical justification for this doctrine is that the peripheral nervous system carries raw stimulus and output before or after integration, respectively, and that this nervous activity is not sufficiently integrated to be the correlate of mental phenomena, which have structure and must be correlated with integrated representations. An empirical justification of this doctrine is the fact that individuals with limbs amputated can feel phantom sensations as if in the limbs that are no longer present, or that the proper brain stimulation can cause visual, auditory, or other sensations in individuals where the organs of sense that are usually the originators of these signals are absent. Both of these examples show that peripheral nervous activity is not the neural correlate of the mental phenomena.

The fact that the lower animals have minds, as most philosophers grant, suggests that the neural correlate is likely mostly a subset of the part of the human brain that is more ancient in origin and thus has some evolutionary counterpart in the largest number of animals. For if a lower animal has mental phenomena correlated with some part of its brain, and if this part is the evolutionary counterpart of some part of the human brain that is thus more ancient in origin, then it is probable that there are mental phenomena correlated with this part of the human brain. This doctrine is supported by the fact that all mental phenomena are, at their foundation, sensory. Even the mental phenomena of thought are but faint recombinations of the mental phenomena of past sensory perception. And since human beings share the faculty of sensory perception with the lower animals, in general, though not in all its specifics, the neural correlate of sensory perception is likely an evolutionarily ancient neural structure shared with the lower animals.

The evolutionary correspondence here is not neuron-for-neuron, as this would be too limiting, but rather structure-for-structure. There may be a part of the human brain that has an evolutionary counterpart in the brain of a lower animal, and neither this part of the human brain nor that part of the brain of the lower animal has mental phenomena correlated with it. The structure-for-structure evolutionary equivalence between the human brain and the brains of most animals may be so close that no part of the human brain would satisfy the criterion of being sufficiently evolutionarily modern for the above a priori classification and ensuing argument.

Suppose such evolutionarily more modern parts of the brain, according to that classification, do exist. Then the question remains whether some of the evolutionarily more modern parts of the human brain are correlated with at least some mental phenomena. They likely would, because the great complexity of the brain means any generic categorization of behaviour is imperfect. However, most of the mental phenomena of the human mind are likely correlated with the evolutionarily more ancient parts of the brain.

A possible objection to this is as follows. The human mental phenomena of abstract thought have no correspondent in the minds of the lower animals. Therefore, the neural correlate of these mental phenomena should be an evolutionarily more modern part of the brain.

A response to this is as follows. The neural substrate of higher intelligence is indeed an evolutionarily more modern part of the human brain, be-

cause it does not exist in the lower animals, because they do not have higher intelligence. But the neural substrate of higher intelligence is not a subset of the neural correlate. Higher intelligence is not due so much to the mental phenomena of abstract thought in themselves as representational signs as it is due to the rules according to which these mental phenomena are changed and combined. And these rules are determined by the neural integration. The neural correlate of a lower animal has less potential for being exercised for higher intelligence by a suitable neural integration than does that of a human being. However, this inadequacy is less extreme than it may appear. What is most lacking in the lower animal is not the raw potential for the material representational substrate in itself to be used as a workspace for higher intelligence. For any given picture in the mental screen of thought in a human being, a roughly equivalent picture could be formed for a moment in the mental screen of thought of a lower animal if the animal has a sufficiently large brain. One can even conceive of the mental screen of thought of a lower animal that cannot typically use language being used by a suitable neural integration implanted around it to form language intelligently. Evidence for this is the fact that a parrot can mimic human language, meaning that its mental phenomena are capable of representing the form, if not interpreting the meaning, of human language. What is lacking in the lower animal is the neural integration that changes and combines the screen of thought in the exercise of higher intelligence. This higher integration is found in the evolutionarily more modern parts of the human brain, which are thus not part of the neural correlate.

A possible objection is as follows. Even if it is granted that the lower animals have minds, this demonstrates nothing about the structure of these minds. But the above argument invoking the minds of the lower animals assumes things about the structure of their minds, for example, that their mental phenomena are, at their foundation, sensory. Thus, the argument is founded on an unjustified assumption.

A response to this is as follows. Once it is granted that the lower animals have minds of some kind, then by analogy between the similar structures and behaviours of humans and lower animals, further inferences may be reasonably made about the minds of the lower animals. For example, the analogy between the structure and function of the sensory organs of human beings and those of lower animals suggests an analogy between the mental phenomena of sensory perception originating from these organs. Thus, the doctrine that the

mental phenomena of the lower animals are sensory in nature and that they resemble the sensory mental phenomena of human beings is justified.

On the Physical Anatomy of the Neural Correlate

The preceding discussion of the macroscopic structure of the neural correlate is a priori. It does not give precise anatomical details, such as which parts of the brain the neural correlate is in. Indeed, this work has little to say about the physical anatomy of the neural correlate of the human mind. For the purpose of this work, suffice it to say that at present, there is some scientific consensus on some aspects of the macroscopic physical anatomy of the neural correlate, but that much obscurity remains.

On False Positives in Mapping Correlation

It is not possible to measure metaphysical correlation directly. The mind has innate intuition of its own mental phenomena, but it has no innate intuition of the material correlates of these mental phenomena, and thus no way to intuit metaphysical correlation innately. It is only possible to intuit metaphysical correlation through representation. The mind must measure through representation the statistical correlation between mental phenomena and physical states, and use this to build a physical model of the material correlate. Or more properly speaking, it must measure a statistical correlation between physical states and external reports of mental phenomena. This methodological limitation leads to a high risk of false positives in mapping the material correlate. Not all statistical correlation is metaphysical correlation, and statistical correlation can be quite close, to the point of deceiving the researcher into thinking it is metaphysical correlation, without truly being metaphysical correlation.

The earliest examples of this are from the ancient uncertainty about which organ of the body contained the material correlate of the mind.

In his work *On the Motion of Animals* and elsewhere, Aristotle expresses the opinion that the heart is the organ of the body most closely associated with the mind. He formed this false belief even though he had conducted anatomical studies of animals and empirical observations of how different stimuli affected their behaviour, and even in consequence of these observa-

tions, the results of some of which he mentions while justifying this belief. Furthermore, he does not even directly mention the brain in this work.

In the division On Man in the first book of his work *Against the Logicians*, Sextus Empiricus, writing around the turn of the third century of the Common Era, gives witness to the ancient uncertainty about the location of the material correlate of the mind, writing that different dogmatic philosophers believed that the mind was located in the brain, the *pia mater*, the heart, and the portals of the liver.

The reason why one can be deeply deceived about what organ is most closely associated with the mind, even after conducting detailed anatomical observations on the body and detailed observations on the correlation between the body and the mind, is that the close structural interconnection and functional interdependence between the parts of the body lead to a detailed and regular statistical correlation between their behaviour. For example, if the heart is affected in a certain way, this regularly leads to a certain kind of effect on the brain, and vice versa. The mind thus will observe that certain stimuli to the heart, or certain behaviours or circumstances of the heart, are regularly correlated with certain states of the mind. It is natural to take this as evidence of a direct metaphysical correlation between the heart and the mind. But in truth, as is now known, the correlation between the heart and the mind is only indirect. It takes place through interactions between the heart and the neural correlate, and the metaphysical correlation between the neural correlate and the mind.

It may be tempting to view this past ignorance about the correlate of the mind as merely a symptom of the limited scientific knowledge of the past. However, there remains a great risk of falling into the same kind of error in modern studies of the neural correlate, only at a finer grain of precision according to the more advanced empirical methods now at the disposal of researchers.

Suppose an empirical researcher observes that neural activity of a certain kind in a certain part of the brain is consistently and precisely correlated with a certain mental phenomenon. A natural conclusion would be for them to claim that this means that this neural activity is the metaphysical correlate of that mental phenomenon. However, such a claim would be too hasty and could be falsified by more sophisticated empirical studies later. For it is possible that

they are only observing a close but indirect statistical correlation, and not a true metaphysical correlation.

On the Necessary Imprecision of External Reports

Since mental phenomena can only be innately intuited by the mind to which they belong, all empirical studies of metaphysical correlation are dependent to some extent on external reporting by the individual studied. Suppose a certain kind of brain activity is measured objectively. Then this measurement can only give evidence that the brain activity is metaphysically correlated with a given mental phenomenon if the experimenter knows that such a mental phenomenon exists, which can only be recorded objectively through external reporting by the mind of its experiences. The problem with this is that active introspection is a different sort of mental state than normal mental existence. That is, reporting on the state of the mind alters the state of the mind from what it was before reporting, because the act of reporting is itself a different state of the mind. It is only possible to measure metaphysical correlation between material processes and externally reportable mental states. Any mental state whose peculiar character depends on its not being reported cannot be empirically mapped to a material correlate. For example, mental states that are forgotten before they can be reported cannot be reported, and thus cannot be empirically mapped to a material correlate.

The problem is deeper than it may at first appear. For there is some imprecision in saying that this problem means that only reportable mental states can be mapped, because it might incorrectly suggest that there was a purely unreported mental state that, by virtue of some accidental property of being reportable, was later reported exactly as it was before the act of reporting. But more properly speaking, it is only the mental state of actively reporting on mental states that can be reported directly, because the report is necessarily an expression of such a state, not of any other. The report is at best an imperfect representation of any other mental state that it purports to represent. Even putting the mind into a state of active introspection and reporting alters its state. Thus, all that is being directly reported is the state of a mind that is in such an introspective state. It is fundamentally impossible to obtain a direct report of

the nature of a mind that is not in a state of active introspection and reporting. Memories can represent such non-introspective states, and continuity principles suggest that such non-introspective states must be similar to introspective states in various ways, especially the ways that reason suggests are least affected by introspection, for example, relating to sensory perception rather than thought. And such principles may suffice for imprecise maps. But if science ever reached the point where it was possible to more precisely measure brain states such that one could measure neural differences as slight as those corresponding to the difference between a non-reported mental state and that mental state as altered by the act of reporting on it, then this limitation may become more problematic.

For example, suppose there were an advanced empirical study measuring the metaphysical correlation between brain states and mental states. The subject of the study is first asked to perform a task unrelated to philosophical introspection while their brain state is measured precisely. Afterwards, they are asked about the state of their mind at some point in the task. They might not remember very well what their state of mind was because they were not trying to remember it. This makes it difficult to map the neural data obtained to mental states. But suppose they are told at the outset that they should focus introspectively on their mental state throughout the task because they will be asked such a question at the end. Then this very instruction primes their mind in a certain way and alters their mental state throughout the task. Thus, at the end, they would be reporting on the state of mind of a person primed to introspectively focus on their mind throughout the task, rather than someone simply doing the task. Again, this makes it difficult to map neural data to non-introspective mental states. Such an imprecision would not be of great importance to a relatively simple map, such as that between brain activity and the form of the image on the mental screen of vision. But it could greatly impair the development of a more sophisticated neural model, such as one of introspective thought itself.

Since the mind is isomorphic to the material correlate, and since such experiments would give identical results if performed on a purely material automaton with no mind, the experiments do not truly measure the correlation between neural activity and mental states, but rather between neural activity and innate self-reporting of neural states. Such studies are static insofar as they attempt to freeze a particular cognitive state through innate reporting and to

measure its material correlate. The problem with static studies of this kind is that reporting is not static, and this introduces fundamental limitations on the precision with which such static measurements can be made.

This is analogous to the uncertainty principle in quantum mechanics, according to which measuring a system alters it. However, there is little reason to believe that the uncertainty arising from reporting on the state of the mind also altering its state is fundamentally connected to the uncertainty principle in quantum mechanics.

It is too early to know what the practical solution to this problem would be, or whether neuroscience will ever progress far enough that this fundamental problem will become a practical one. However, if it does, one possible solution to it could be a dynamic theory of cognitive states and a dynamic system of measuring the mind and its external reports. Instead of asking the subject to try to freeze one particular state introspectively and report on it alone, the study could involve measuring the brain activity of the subject over the course of much behaviour, both involving external reporting of cognitive states and non-introspective behaviour. Then, the interpretation of the data would rely on applying a dynamic model of material innate intuition to all brain states and assessing how well it predicts the observed behaviour. If it does in a wide variety of situations, then the state function of the mental states could be determined from the data, not as an explicit datum directly reported, but as a prediction of the model fixed indirectly by the data. This is analogous to how observables allow one to learn about the wave function of a system in quantum mechanics before measurement occurred. However, this is only speculation.

On the Focus of This Work

This work is not much concerned with the empirical details of the anatomy of the neural correlates of human beings or animals. The main doctrines of this work stand irrespective of what the anatomy of the neural correlate happens to be. Furthermore, even if some or all of the above a priori doctrines on the microscopic and macroscopic physical structure of the neural correlate are false, the remaining doctrines of this work still stand, because these later doctrines generally do not rely on the doctrines on the physical structure of the neural correlate, even its a priori structure. This work is more concerned with the

metaphysical structure of the neural correlate. That is to say, this work is more concerned with its a priori structure relative to how it is correlated with the mind and how it forms representations.

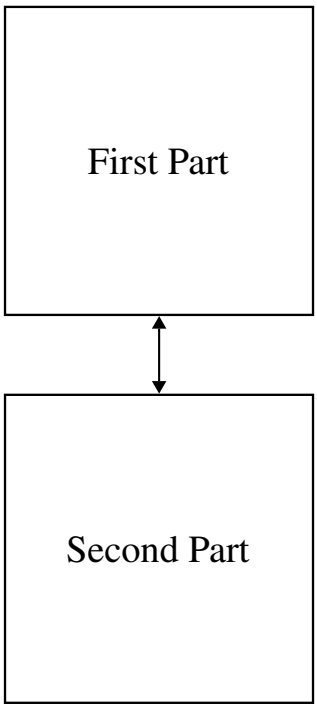

Figure. A partition of the material universe into two parts. The bidirectional arrow represents the interaction between the parts. The diagram does not necessarily represent spatial relationships. For example, each part need not be a square, or convex in space, as the squares in the diagram are. The squares simply represent arbitrary disjoint subsets of the material universe. Though the squares defining the parts are bounded in the diagram, the parts of the material universe that these squares represent need not be bounded in physical space, and indeed, at least one must be unbounded if the material universe is unbounded in space. This partition does not necessarily define an automaton.

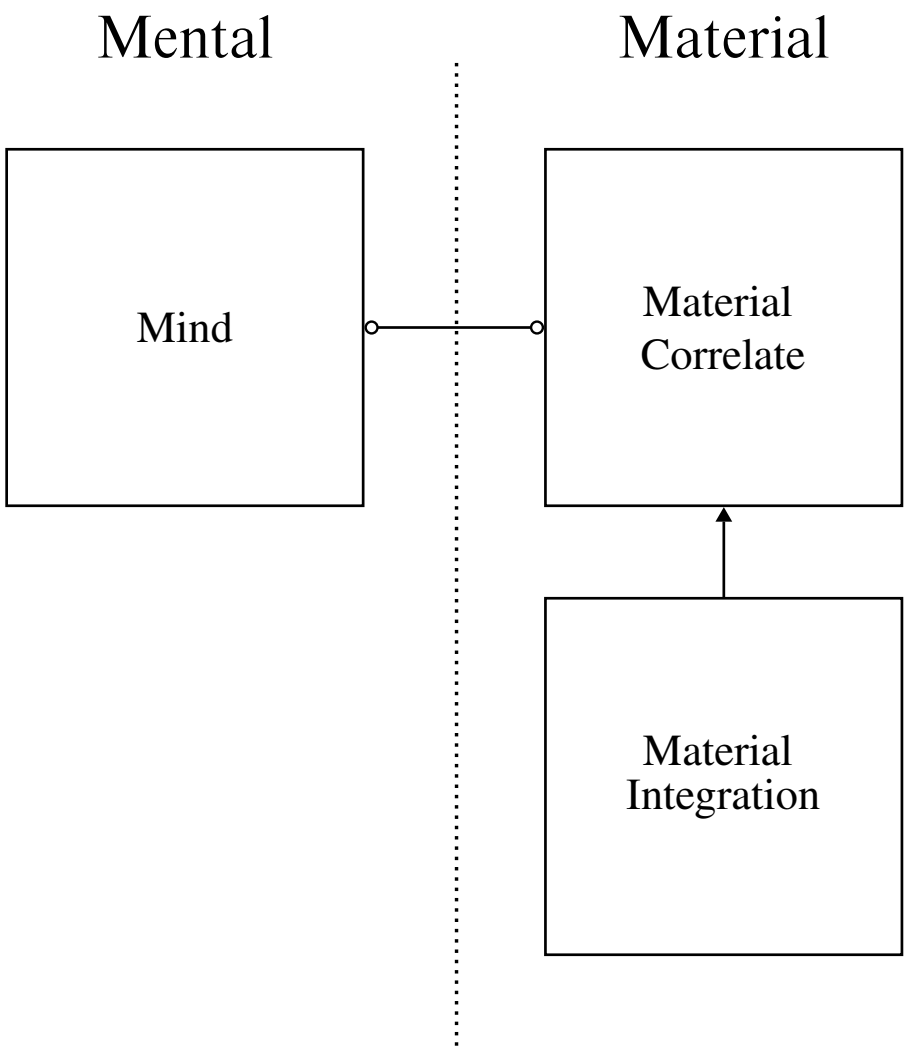
Mental
Material
Mind
Material
Correlate
Material
Integration

Figure. A dual automaton. Here, the material universe has been partitioned in a more structured way than in the preceding figure. The interaction between the two parts of the material universe is structured such that material innate intuition occurs, represented in this figure and in later figures of this work by a solid unidirectional arrow from the intuiting entity to the intuited entity. Material innate intuition is so complex that it cannot be described diagrammatically beyond noting that it occurs. It is not so much the strength of the interaction that makes it material innate intuition, but how structured it is. Material innate intuition is a highly structured interaction, whereas general interactions are unstructured. The material part of this diagram can be interpreted as again representing a partition of the material universe, in which case the material integration is the entire rest of the universe. For it is not wrong to say that the complement of the material correlate in the material universe is the material integration of the automaton. But though this is not wrong, it is unintuitive. It is more intuitive to set some bound for the automaton, usually determined by continuity. This bound further divides the material universe into three parts, the material correlate, the material integration of the automaton, and the part of the material universe beyond what is arbitrarily designated the material integration. This third part is not shown explicitly in this diagram. The dotted line represents the separation between the mental and material universes. The symbolic arrangement of the diagram represents this separation with spatial relationships, but the separation itself is not spatial. The line with circles at its ends represents the direct metaphysical correlation between the mind and its material correlate.

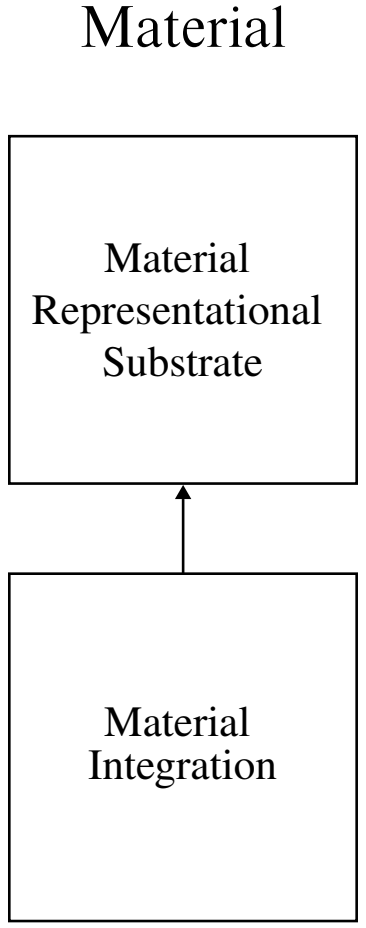

Figure. A purely material automaton. The physical structure and behaviour of this automaton are identical to those of the dual automaton that would ordinarily have a mind correlated with the material representational substrate, but this automaton does not have such a mind correlated with it. Such an automaton is likely not naturally possible in reality, but it is at least logically possible.

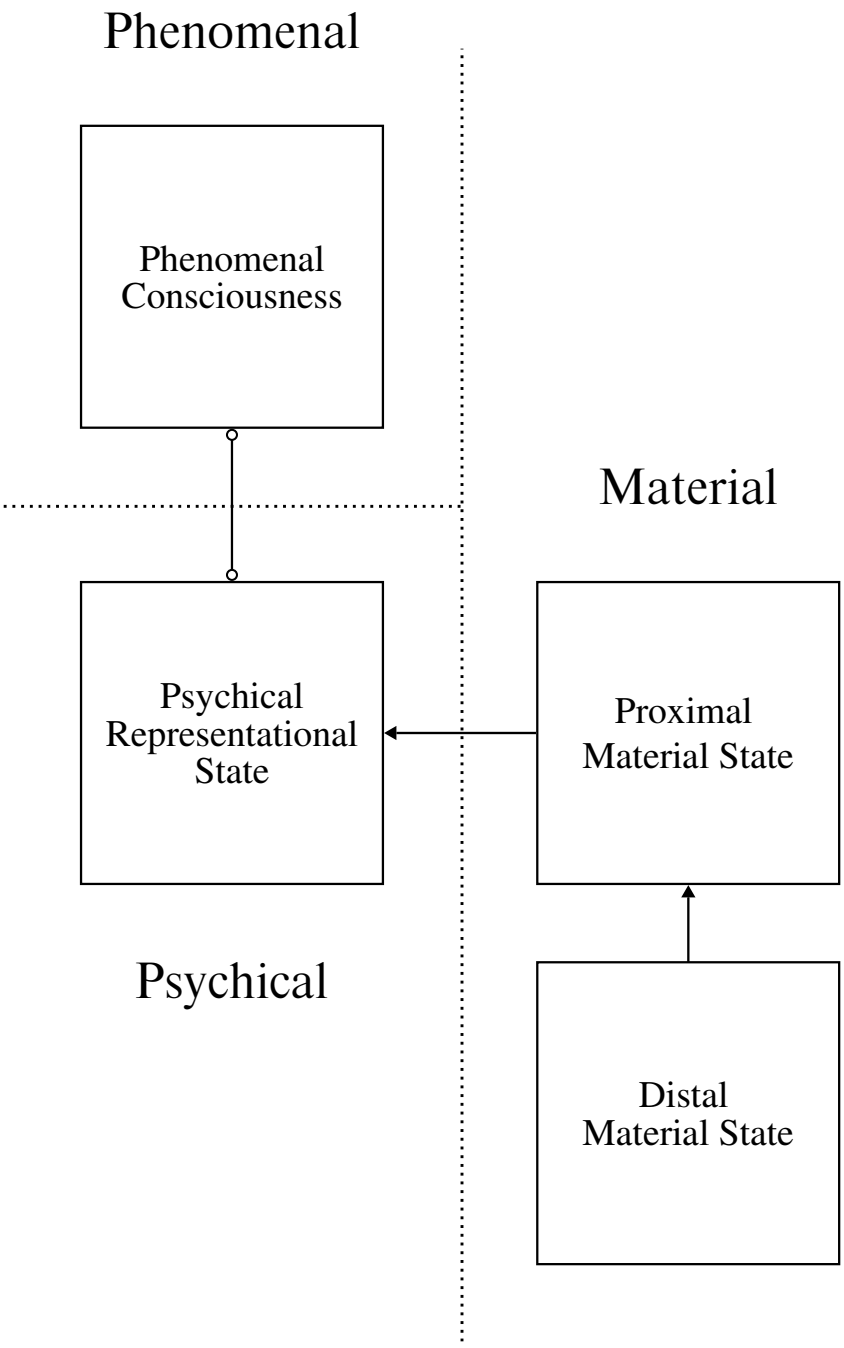

Figure. An automaton with a psychical state.

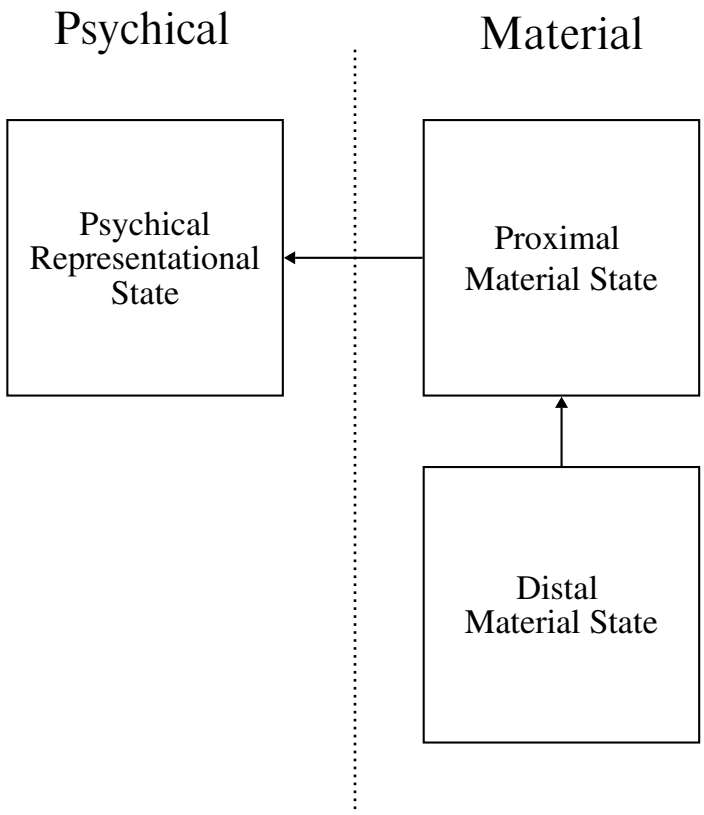

Figure. An aphenomenal automaton with a psychical state. According to the use of language where psychical states are material, this automaton may be called a purely material automaton. In any case, it is an aphenomenal copy, because it is a version of an automaton that typically has phenomenal consciousness associated with it. This diagram is not merely a relabelling of that of a dual automaton with the label *Psychical* replacing the label *Mental*. The difference is that the relation between the proximal material state and the psychical representational state is that of interaction and integration, rather than of epiphenomenal metaphysical correlation, as is that between phenomenal consciousness and the representational state, whether psychical or material. This difference means that the automaton innately behaves as though it has phenomenal consciousness correlated with its psychical representational state.

On Representation

Introduction

To make the doctrines of this work as general as possible, there is required an a priori metaphysics of automata in general that is independent of the particular automaton that may instantiate the discussion. The last chapter was general in some parts, but also discussed the particular structure of human beings or animals in other parts. Generally, in the remaining chapters of this work, the discussion applies to automata in general whenever possible, and only references the metaphysics of human beings or other particular automata when the extra assumptions that define these automata are used in the discussion.

This chapter discusses how automata and mental phenomena represent reality. It gives a metaphysical anatomy of the faculty of intuition. Of fundamental importance to this discussion is how representation is a tandem act of mental innate intuition, or consciousness, and material innate intuition, which is the material faculty that underlies the potential for representation.

The primary doctrines of this chapter are as follows. Mental innate intuition and material innate intuition are isomorphic and correlated. Mental intuition through representation is a subset of possible innate intuitions when the

innate intuition is performed in such circumstances that it represents something. Material intuition through representation is the material correlate of mental intuition. What is called the *mental object*, a particular mental phenomenon, is distinguished from what is called the *material object*, its material correlate. And if they represent something, this represented thing is called in this context the *represented object*. Various specific doctrines are elaborated using these notions, many of which are logical elaborations from the definitions themselves.

Most of the doctrines contained in this chapter are not completely new. However, the doctrines of this chapter are more original than those of the chapters On Knowledge and On Correlation. The doctrine that the minimal collection of matter directly correlated with the mind, the material correlate, happens to be identical to a particular collection of matter defined with certain functional properties, the material representational substrate, has already been put forward by many philosophers, though with different philosophers using different terms and with variations on the two definitions. Doctrines of this kind have dogmatic significance in showing the stance of the author on existing problems, but they do not have great significance for the progression of knowledge, because the doctrines are not new. This content forms the groundwork for the remainder of this work. A significant doctrine of this chapter is the notion of dualistic and materialistic behaviour, which in its definition alone is not entirely new, but which is elaborated later in this work and used to frame other original doctrines.

On Material Innate Intuition

Material innate intuition is the material faculty that underlies the potential for representation.

A material representational substrate is the maximal direct object of one material innate intuition.

Representational input is the part of the material representational substrate that is intuited passively as if from without.

Representational output is the part of the material representational substrate that is intuited actively as if from within.

The material representational substrate of an automaton is the collection of matter that is globally available to the automaton for influencing articulate behaviour through material innate intuition. Articulate behaviour need not necessarily be verbal.

The concepts of representation and material innate intuition are fundamental, but their absolute definitions are not yet clear. It might be possible to express the definition of a material representational substrate mathematically. Then, from this mathematical definition, various theorems about representation and automata could be logically derived. This work does not endeavour to describe such a mathematical theory. Attempts have been made to mathematically or scientifically describe such a theory, but it does not seem that any currently existing mathematical theory is yet a fundamental theory of representation.

Instead of attempting to define the material representational substrate mathematically, this work will attempt to describe and characterize it philosophically. Thus, the other doctrines of this work about automata will not be logically derived from the definition given in this chapter in the pure way of mathematics. Instead, they take their source from judgment.

On Automata

An automaton is a collection of matter containing a material representational substrate considered for the purpose of the discussion to be the body of the material representational substrate, as opposed to the external environment.

The material integration is the complement of the material representational substrate in the automaton.

Afferent integration takes raw input from the environment and integrates it into representational input.

Efferent articulation takes representational output and articulates it into raw output that acts on the environment.

The concept of an automaton is based on representation. In an automaton, peripheral structures integrate raw input from the surroundings into representational input. Representational input is globally accessible for articulate reference. The automaton articulately generates representational output in response to representational input. Representational output is globally access-

ible for articulate reference. Peripheral structures convert representational output into raw output. Raw output acts on the surroundings. Representational input and representational output together form the material representational substrate.

Some parts of this definition of an automaton are fundamental, and some are only practically useful concepts. The concept of a material representational substrate is fundamental. There exists in principle an absolute definition of what material systems constitute a material representational substrate and what do not. The distinction between representational input and representational output is not fundamental, and there are material representational states that are intermediate in character between the two. The definition of the external boundary of the automaton, that is to say the distinction between the peripheral structures of the automaton and the surroundings, is not fundamental. It is best set practically by considering continuity, and has no necessary definition. Thus, in the most precise sense, the material representational substrate is an automaton embedded in the body, and there is no fundamental distinction between the rest of the body and the rest of the material universe outside the body.

The automaton is the particular automaton, universally quantified, considered for the purposes of the discussion. The automaton materially and innately intuits the material representational substrate.

On Representation

A representational state is an object of material innate intuition.

A representational assertion is a material innate intuition when it occurs in such circumstances that it makes an assertion about reality.

Material intuition through representation is material innate intuition when it occurs in such circumstances that it represents something.

Representational assertions endeavour to represent the state of the automaton and the rest of reality.

All representational states are implicitly representational assertions. That is to say, all patterns of behaviour are in some sense implicit beliefs, and all beliefs take their source from patterns of behaviour. There is no clear-cut

distinction between them because human beings were designed by natural selection to survive and reproduce rather than to philosophize.

Representation is always with respect to the circumstances. Not all representational assertions are representations, because they may make false assertions that do not accurately represent anything.

An objector may claim that innate intuition is always intuition through representation, because all innate intuitions represent something, if only the direct object of that innate intuition, or the integration of information itself. If they are the same, then the distinction between a representational assertion and an intuition through representation made in this work is meaningless.

A response to this is as follows. As regards the claim that innate intuition represents the integration of information itself, the following may be said. This work does not positively affirm that mental innate intuition corresponds exactly to information integration as this term is sometimes interpreted in various more precise senses. However, information integration is at least attempting to define something similar. But if the two are the same, then innate intuition still does not necessarily represent the intuition of information. And in general, innate intuition does not necessarily represent itself. And innate intuition does not necessarily represent its direct object, or anything at all. For example, suppose one dreams of a tree that does not exist in reality, believing it represents a real tree because one is unaware that one is dreaming, and one has no other thoughts or beliefs at this time that would alter this representational assertion significantly. This mental innate intuition does not represent a real tree, because the dream is an illusion. It does not represent the integration of information itself, because one is not dreaming about reflecting abstractly on the integration of information. If one were, then it would represent this. It does not represent the mental phenomenon itself, because one is not dreaming about introspecting philosophically on the mental phenomenon itself. If one were, then it would represent this. The representational state and its representational assertion do not constitute intuition through representation in this case because the representational assertion is false.

On Dualistic and Materialistic Automata

A dualistic automaton believes that it is a dual automaton.

A materialistic automaton believes that it is a purely material automaton.

Dualistic behaviour of an automaton is behaviour that appears to reference the mind. A dualistic automaton behaves as if it has a mind correlated with it. In other words, a dualistic automaton has dualistic behaviour. A materialistic automaton does not have dualistic behaviour.

When a dual automaton states *I am*, this sentence references the mind of the automaton. This is because the second-order contemplation by the mind must represent the mind, because it is thus directly aware of its own existence. And since the mind and the automaton are one unified being with metaphysically correlated parts, both the mind and the body in this tandem act must represent the same thing. At least this is so if the behaviour of the material part of the automaton is to be interpreted within the context of the dual totality of the automaton, not in itself as matter.

A dualistic automaton likely believes that other automata like it are dual, unless it is a solipsist. A materialistic automaton likely believes that other automata like it are purely material.

An innately dualistic automaton has its material innate intuition structured such that its representations are all at least implicitly dualistic, irrespective of whether the higher-level judgements of the automaton are explicitly dualistic.

Innately dualistic behaviour is described in more detail later in this work.

On Two Definitions of the Material Correlate

Various philosophers have hypothesized based on empirical observation that the material correlate of a dual automaton is always its material representational substrate, and that mental innate intuition, defined metaphysically, is thus isomorphic to material innate intuition, defined functionally. Thus, it has been hypothesized that the material correlate of the mind coincides exactly with a material system defined in purely functional terms with no reference to the mind. This material system may be defined in terms of representation, or the integration of information, or the global accessibility of a state, or a similar notion. This hypothesis is elaborated and defended in this chapter and later in

this work. The material integration has been given its name because its role with respect to the material correlate is to integrate environmental stimuli and process representational states in the material correlate. This foreshadows the alternative functional definition of the material system that is the material correlate of the mind. As is shown later in this work, the representational function of the material correlate of the mind is interpretively connected to why this material correlate is the material correlate of anything at all. The connection is through the innately dualistic structure of representation.

There are thus three ways to characterize the material correlate or representational substrate, namely, correlatively, functionally, and anatomically. The correlative characterization uses the definition of the material correlate with respect to consciousness. The functional characterization uses the definition of the material substrate with respect to its representational structure as matter. Both of these definitions are a priori. The fact that they characterize the same thing is, as is shown later in this work, also capable of being demonstrated mostly a priori. The anatomical definition gives the physical structure of the material correlate or representational substrate as a set of events in spacetime, as intuited through representation scientifically, rather than its metaphysical structure, as intuited innately. This characterization is a posteriori, and this work is not much concerned with it.

The physical structure of the material correlate is its whole structure as matter. The metaphysical structure of the material correlate is its structure with respect to material innate intuition. The physical and metaphysical structures of the material correlate are different. The physical structure of the material correlate is more complex than its metaphysical structure. The mind has no physical structure. The metaphysical structure of the mind is its whole structure as a mental phenomenon. The metaphysical structure of the material correlate is a function of the physical structure of the nervous system. The metaphysical structure of the material correlate is isomorphic to the metaphysical structure of the mind.

On the Mental and Material Objects

The following metaphysical anatomy of the material correlate of a general dual automaton is a priori and makes no claims about the particular structure

of the automaton. It applies to the neural correlate of a human being as a particular consequence.

The mental object is the particular mental phenomenon, universally quantified, considered for the purposes of the discussion. The material object is the material correlate of the mental object. The mental object and the material object are distinct because mental phenomena and matter are distinct.

The mind innately intuits the mental object. To be innately intuited is the essence of the mental object. The complete structure of the mental object is its structure with respect to mental innate intuition. The innate intuition of the mental object characterizes what it is for the mental object to exist.

The material integration innately intuits the material object. It is not the essence of the material object to be innately intuited. The essence of the material object is its material essence as substance existing in itself. The complete structure of the material object is not its structure with respect to innate intuition. Its complete structure is its physical state. The innate intuition of the material object is a material process with certain functional properties. It is the process underlying the possibility of the state of the nervous system and of external reality being represented in articulate behaviour. Articulate behaviour is not necessarily verbal.

Innate intuition is a tandem mental and material act of the mind and the material integration on the mental object and the material object, respectively. Innate intuition of mental phenomena is always mental innate intuition, and innate intuition of a material representational substrate is always material innate intuition. The two forms of intuition, mental and material, can only be applied to their appropriate objects, mental and material, respectively, not to the other correlated object, material and mental, respectively. The material object, which underlies the mental object, only depends on innate intuition in itself, not on the circumstances in which innate intuition takes place.

The physical structure of the material object is its whole structure as matter. The metaphysical structure of the material object is its structure with respect to material innate intuition. The physical and metaphysical structures of the material object differ. The physical structure of the material object is more complex and fine-grained than its metaphysical structure. The mental object has no physical structure. The metaphysical structure of the mental object is its whole structure as a mental phenomenon.

The metaphysical structure of the material object is isomorphic to the metaphysical structure of the mental object. The physical structure of the material object has no complete parallel in the mental object, because the material object has a detailed structure as matter with much information that is irrelevant modulo the metaphysical structure of material innate intuition, whereas the mental object can only parallel its structure up to the scope and precision of innate intuition.

On the Represented Object

Material intuition through representation is when an automaton materially innately intuits the material representational substrate in such a way that the material representational substrate represents something.

The represented object is that which the mental object represents, if the mental object is being innately intuited in such circumstances that it represents something.

If the mental object represents something, then the material object also represents that thing. And if the material object represents something, then the mental object also represents that thing.

What the represented object is, if it is anything, depends on the circumstances. The represented object can be matter, a mental phenomenon, or a combination of matter and mental phenomena. If representation does not take place, then there is no represented object.

The material representational substrate usually represents something. Thus, material innate intuition is usually accompanied by material intuition through representation.

On the Relations Between the Types of Intuition

If the automaton is dual, then material intuition through representation is the material correlate of mental intuition through representation. In the particular case when the automaton is a human being, or when it has sensory perception as human beings do, and when sensory perception represents matter in the typical way, the represented matter causes the stimulus that the material integration integrates into the material correlate of the mental sensory perception.

The mind mentally intuits the represented object through representation. The material integration materially intuits the represented object through representation. The material integration intuits the represented object through representation, if and only if the material integration innately intuits the material object in such circumstances that this representation occurs, if and only if the material object represents the represented object, if and only if the mental object represents the represented object, if and only if the mind intuits the represented object through representation, if and only if the mind innately intuits the mental object in such circumstances that this representation occurs. The circumstances that determine that the material object represents the represented object are precisely the circumstances that determine that the mental object represents the represented object. Intuition through representation is a tandem act of the mind and the material integration, referencing the represented object, and acting directly on the mental and material objects, respectively. In principle, intuition through representation is either mental intuition through representation or material intuition through representation or both. In practice, intuition through representation is always a tandem act of the mind and the material integration, directly on the mental and material objects, respectively, to represent the represented object.

The represented object and the material object are the same if and only if the material object is intuited through representation, as occurs in analyses of the mind–body relationship, such as the preceding analysis. The represented and mental object are the same if and only if the mental object is being intuited through representation, as occurs during deliberate, second-order reflection on the mental object itself. Thus, the represented object is usually distinct from the mental and material objects. Usually, the represented object is structurally different from the mental and material objects, even when the mental and material objects endeavour to represent themselves, because of the limitations in the human mental faculties that make all representation imperfect. Usually, the mental and material objects do not represent the complete structure of the represented object, nor with complete accuracy. Thus, in the case above, where the represented and material objects were the same, the identity was only approximate.

Intuition in general is any of several faculties. Intuition can be material or mental, and innate or through representation. Material intuition is either material innate intuition or material intuition through representation.

The term *innate intuition* alone either refers to mental innate intuition, material innate intuition, or both. Strictly speaking, it could also refer to theotaxic innate intuition or innate intuition corresponding to a more exotic substance. This remark about more exotic substances also applies to other terms, such as *intuition through representation*, that have mental and material versions. Theotaxy and exotic substances are discussed later in this work.

The above discussion indicates why the uncommon term *mental innate intuition* is used in this work rather than the more common term *consciousness*. The term *intuition* makes it easier to precisely describe the parallel between mental innate intuition, mental intuition through representation, material innate intuition, and material intuition through representation.

On the Neural Special Cases

The neural representational substrate is the material representational substrate in the special case where the material representational substrate is composed of nervous tissue. The neural correlate is the part of the material representational substrate composed of nervous tissue in the special case where the automaton is dual. Conversely, the material representational substrate is the material correlate when the speaker wishes to remain agnostic about whether the mind exists, or wishes to deny it. Similarly, the neural representational substrate is the neural correlate when the speaker wishes to remain agnostic about whether the mind exists or wishes to deny it.

Neural intuition is material intuition in the special case when the material correlate is composed of nervous tissue.

Neural innate intuition is material innate intuition in the case where the material correlate and the parts of the material integration adjacent to the material correlate are composed of nervous tissue.

Neural intuition through representation is material intuition through representation in the case where the material correlate and the parts of the material integration adjacent to the material correlate, and the parts cognitively used in determining the representational state are composed of nervous tissue.

Material innate intuition is the faculty whereby an automaton articulately determines representational output using the state of its material substrate. Neural innate intuition is material innate intuition in the special case

where the material correlate and representation of the automaton are composed of nervous tissue. If the automaton is dual, then material innate intuition is the material correlate of mental innate intuition. Mental innate intuition likely has no subject, but the subject of material innate intuition is the material integration. The material integration materially intuits the material correlate innately. The neural integration is the part of the material integration that is composed of nervous tissue.

On Inactive Symbols

An inactive symbol is a material object with no innate intuition produced by an automaton and capable of causing that automaton and possibly other automata to form representational states when they intuit it according to a regular pattern of intuition analogous to the intuition used to create it.

An active symbol is a representational state of an automaton.

The innate intuition of an automaton is active. An inactive symbol has no global innate intuition. An inactive symbol is not an automaton.

An inactive symbol by itself does not represent. It only represents with respect to how an automaton intuits it. Thus, material innate intuition is the faculty that underlies the potential for representation even in such cases where the external symbol is an inanimate object and not part of the material representational substrate.

Written words are inactive symbols. The act of writing is a process that creates an inactive symbol. The act of reading is the process analogous to the act of writing in which the inactive symbol causes automata to form representational states. A book may be considered a collection of inactive symbols, or, if the collection is viewed as unified, a book may be considered one inactive symbol.

Spoken words are inactive symbols. Thus, the term *inactive* is used here with respect to innate intuition, not necessarily with respect to space. That is, an inactive symbol can be in a state of motion, that is, activity of some sort. Only, it must not be in a state of activity relative to innate intuition, that is, it must not be the representational state of an automaton. Indeed, even a book is not perfectly inactive, because its particles are moving about relative to one another. And the definition of whether something is in motion depends on the

reference frame used, and anything is in motion relative to some reference frame.

An objector may claim that an inactive symbol is a counterexample to the definition of material innate intuition given in this work. For it is commonly acknowledged that an inactive symbol can represent things. But an inactive symbol is not an automaton, and has no global faculty of material innate intuition. Therefore, in this case, the representation is not underlain by material innate intuition. Therefore, material innate intuition is not the faculty that underlies representation. This contradicts the definition given of material innate intuition.

A response to this is as follows. It is true that, in the common sense of the word *representation*, an inactive symbol by itself is a representation of things, be they things in the real world or mere ideas. However, in the more precise philosophical sense intended in this work, an inactive symbol does not represent anything by itself. An inactive symbol only represents anything with respect to how an automaton intuits it. Or, to speak still more precisely, an inactive symbol cannot actively represent anything ever, but it has the potential to cause an automaton that interprets it to form a representational state, and this representational state can represent things, as underlied by the material innate intuition of the automaton. And if the inactive symbol regularly causes various automata relevant to the discussion to represent the same thing, then this inactive symbol may be said to represent that thing. However, the fundamental definition of what the inactive symbol represents can only be given with respect to innate intuition by an automaton.

An objector may claim that if an inactive symbol has the potential to cause an automaton to form a representational state, then the inactive symbol underlies the potential for representation, and thus, by definition of material innate intuition, the inactive symbol has a faculty of material innate intuition, a contradiction, because it is generally agreed that this is not the case.

A response to this is as follows. This objection is based on a superficial application of certain patterns of words, not on the true nature of the concepts involved. An inactive symbol has the potential to cause an automaton to form a representational state, but an inactive symbol does not underlie the potential for representation in the particular way that characterizes material innate intuition. The words *potential* and *representation* are being used in both cases, but

they are being used in different ways, and the different meanings intended should not be confused.

An objector may claim that when a computer interprets a file to generate a human-readable display or a sound wave that a human can understand as spoken words, the computer is thus an automaton interpreting the inactive symbol of the file. Therefore, the variable in the computer used as the file is interpreted thus as an active representational state of an automaton. But this is a contradiction, because the computer does not interpret the meaning of the file, and is only generating an inactive symbol to show to a human being.

A response to this is as follows. This objection is based on an equivocation between two different senses of the word *interpret*. The computer is interpreting the file in the superficial sense of using the stored binary to generate a visual display or sound wave. But the computer is not interpreting the file in the sense of analyzing its meaning beyond this superficial structure. Thus, the variables used in generating the human-readable inactive symbol from the human-unreadable inactive symbol are innately intuited relative to the shallow structure of the operating system, but not relative to human understanding.

The reason why it is more precise to say that an inactive symbol causes an automaton to represent something, rather than that the inactive symbol itself represents something, is that the same inactive symbol can cause different automata to form different representational states and to represent different things. There is no objective criterion that picks out what the inactive symbol represents by itself, independent of the particular intuitions that different automata form from it.

The ability to interpret inactive symbols is grounded in experience. The same inactive symbol may be used to form a false or confused representational state by one automaton, if that automaton is incapable of properly interpreting the meaning of that symbol.

An objector may claim that if there exists a Platonic ideal of the meaning of terms, this implies that these terms have objective meanings in themselves, and that their meaning as interpreted by any automaton is only an imperfect shadow of this ideal meaning. But the author has claimed that the only meaning of terms is relative to particular automata. This is a contradiction.

A response to this is as follows. Certain words or symbols with deeper meanings may be misunderstood by those with faulty intellectual faculties. But ontologically speaking, there is not necessarily a truly existing ideal of

which particular uses of these terms are imperfect instantiations. It is possible that such ideals exist, because it is logically possible that reality has various ontic frills without causal efficacy. However, there is no necessary argument for the existence of such ontic frills. And if such ideals did exist, they would not have causal efficacy. And semiotically speaking, the only fundamental way to define representation is relative to automata.

However, this work does not positively claim that it is impossible for the Platonic ideals or ideals of some other kind to exist as ontic frills to reality without causal efficacy. The nature and possibility of ontic frills are discussed in more detail later in this work.

On the Conditions for these Doctrines

The collection of facts from the perspective of which it is determined whether or not a doctrine is a priori, and the reasoner from the perspective of which the reasoning takes place, depend on the circumstances of the discussion. In this context, general propositions about the organization of automata are a priori, and the existence of matter and the particular organizations of particular automata instantiating the discussion are a posteriori. If matter were not to exist, then the a priori arguments about the behaviour of automata hold vacuously.

On the Originality of these Doctrines

An objector may claim that this metaphysical anatomy of intuition and representation merely uses new terms and ways of expressing oneself to describe existing concepts. It is already known that mental and material acts are correlated and can represent things. The author is merely acknowledging this fact with new language.

A response to this is as follows. It is true that this discussion is not venturing into territory that no one has yet considered, so, in this sense, the concepts are not completely new. However, many existing treatments of the subject are flawed. Many philosophers have only attempted to describe this metaphysical anatomy of intuition to make a materialistic argument supposedly reducing the mind to material innate intuition, which is misguided. A physical theory of material intuition in the human brain or elsewhere cannot refute the

existence of the mind, nor can it be a substitute for a philosophical theory of intuition in general. And if the discussion is motivated by a flawed goal, then the discussion as a whole tends to be flawed.

Furthermore, even if the philosopher is a dualist, they usually discuss the above in the context of advocating a particular physical theory of material innate intuition. A philosophical theory of intuition in general should be built up separately from any scientific theory of the particular processes that instantiate these general processes in human beings or other particular automata. Attempting to do both at once, or simply doing the latter and neglecting the former, and thus failing to understand that these are two separate problems that both require a solution, confuses the discussion.

And even among those who give a philosophical treatment of intuition separate from any particular scientific model, many only give imprecise speculation in the context of claiming agnosticism on the exact details of the subject. Examples of this are to be seen in the many popular philosophy discussions that do little more than indicate that they do not understand the answer to the general problems they are discussing.

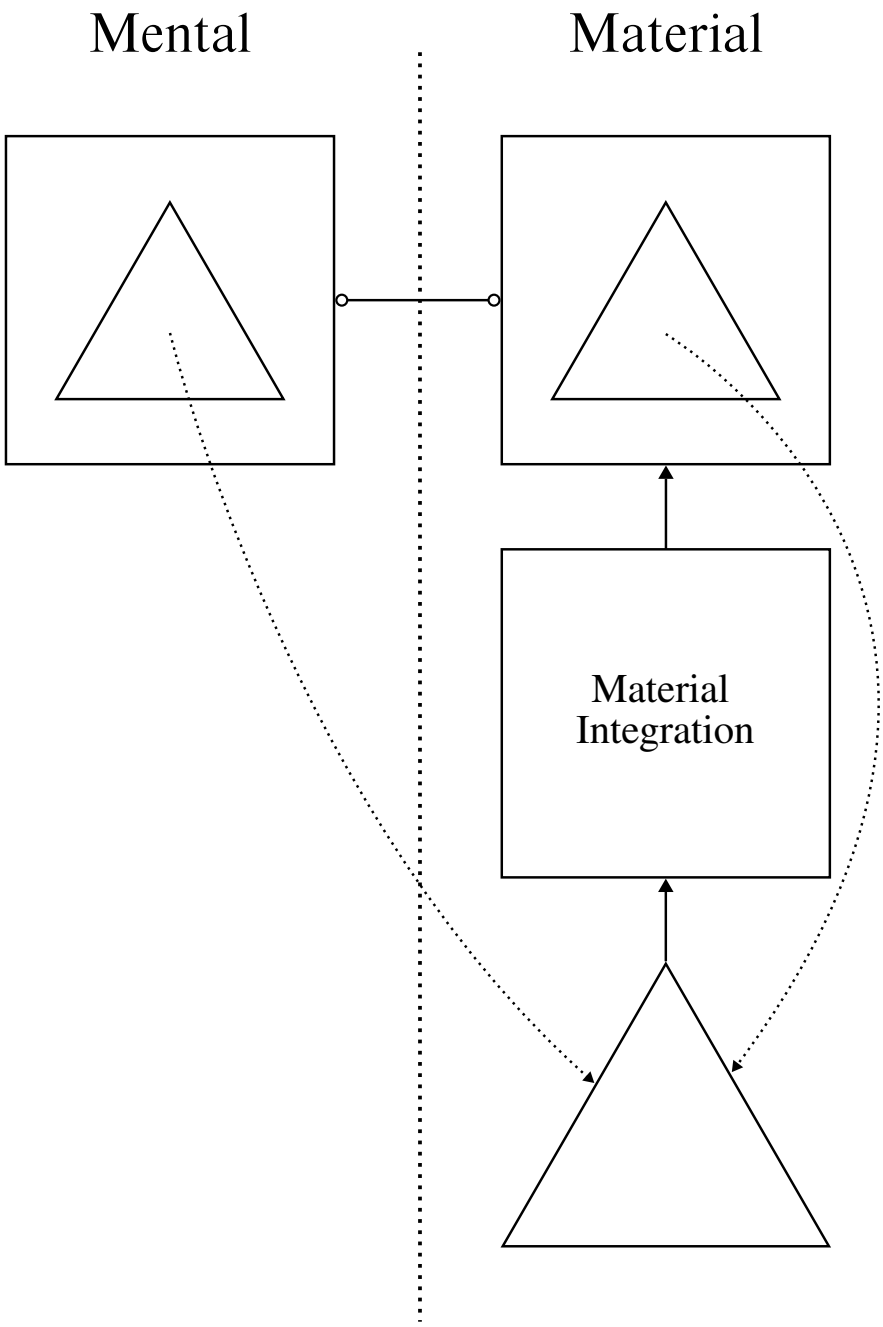
Mental
Material
Material
Integration

Figure. A dual automaton representing a material object. The top solid arrow represents material innate intuition, whereas the bottom solid arrow more generally represents the integration of information. Included in this latter integration of information is the material innate intuition of the units, such as nerve cells in afferent channels, though this is distinct from the main material innate intuition of the whole automaton. The two smaller triangles in the top two squares represent the active representational states in the mind and material correlate of the automaton, respectively. The represented object need not be triangular in shape in general. A triangle is merely used for illustrative purposes in the diagram. Furthermore, the active symbols that represent the material object need not be shaped like it. They are shaped like it in the diagram for illustrative purposes, and because this is the most straightforward kind of representation. But representation may occur without the representational state being shaped like the represented object. And this observation applies both to subjective and physical structure. Neither the subjective shape of the mental phenomena nor the physical shape of their material correlate needs to be shaped like the represented object. For example, the representational state could be the word *triangle*, both mental and physical in correlation, intuited in such circumstances that representation of the external triangle occurs. And the word *triangle* is not shaped like a triangle.

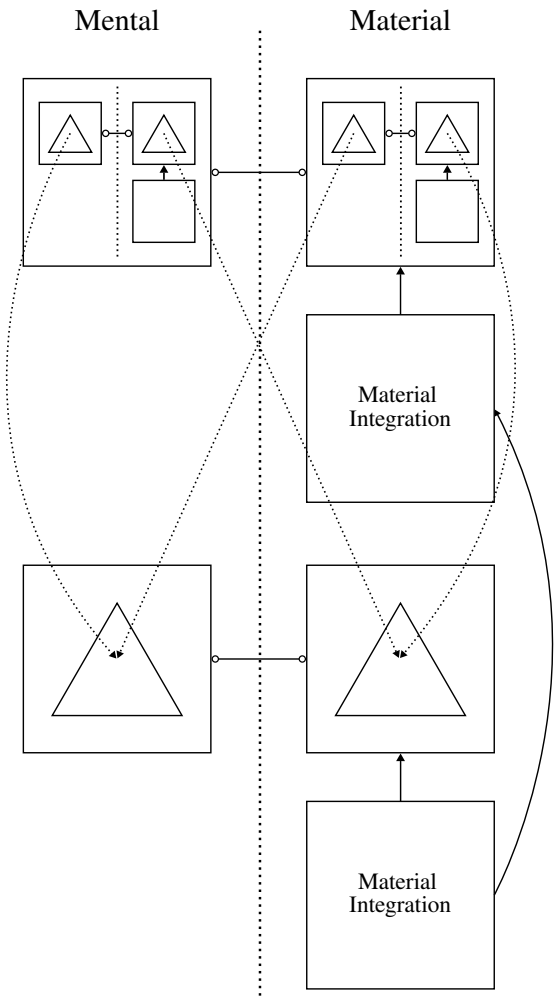

Figure. A dual automaton representing a dual automaton. This diagram pictures the represented automaton as being different from the representing automaton, for in general this may be so. But an automaton may also represent itself, in which case the dotted arrows representing representation in this diagram would loop back to the automaton itself, and the solid arrow from one material integration to another would have no clear correspondent, and would be replaced by the fact that an automaton is in constant interconnection and interaction with itself.

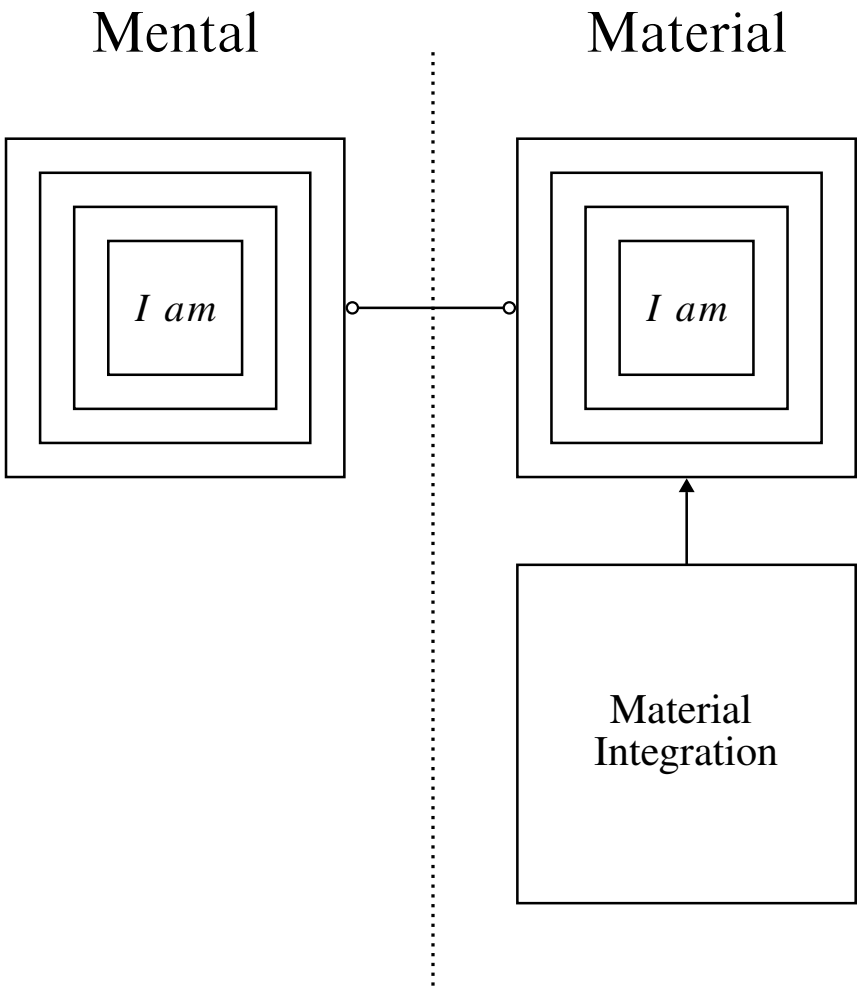

Figure. A dual automaton formulating the doctrine *I am.* It is engaging in second-order contemplation of innate intuition. Innate intuition is represented here by nested rectangles and the sentence, *I am.* It is not necessarily the case that the automaton is literally picturing nested rectangles to itself or saying the words, *I am.* When the mind of the dual automaton states *I am*, this sentence references the mind. When the body of the dual automaton states *I am*, this sentence references the mind, even though the material innate intuition of this automaton directly involves the material representational substrate.

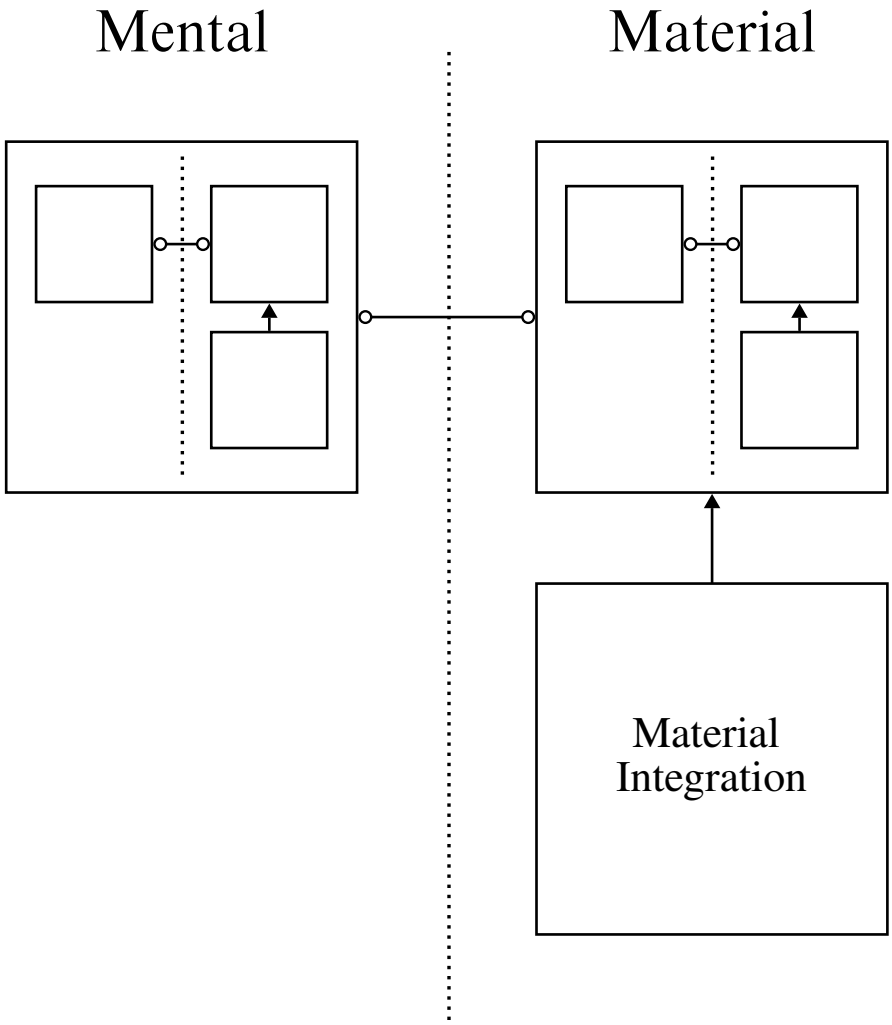

Figure. A dual automaton formulating dualism. This diagram is symbolic and does not indicate spatial relationships. Thus, the image in the square that represents the material correlate does not represent that this image is physically present somewhere in the material correlate, or even that the material correlate is representing the image itself. It only signifies that the material correlate is forming some symbol by which the automaton represents mind–body dualism to itself, whatever this symbol may be.

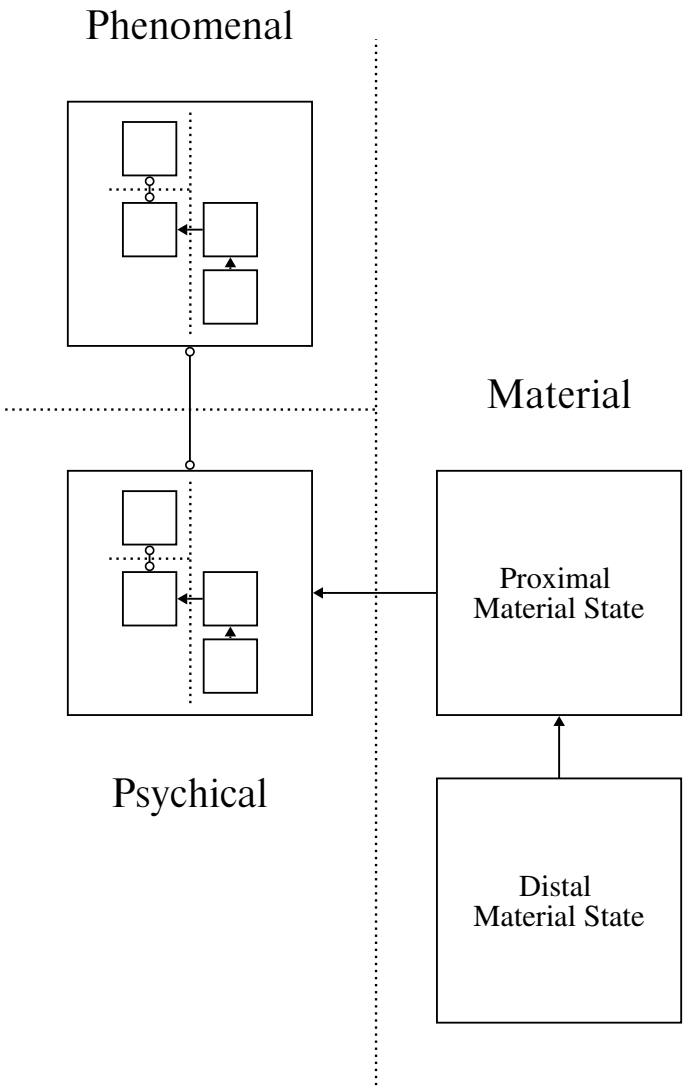

Figure. An automaton with a purely psychical state formulating dualism. Note that it not only claims there is a structural duality between interacting psychical and material states. It further claims that there is something that it is like to be a psychical state.

On the Paradox of Dualism

Introduction

The doctrines presented so far in this work may be plausible individually, but taken together they lead to a paradox, which is the paradox of dualism in the purely material universe. This chapter shows how the possibility of purely material automata and the truth of epiphenomenalism, though plausible individually, together lead to this paradox. For purely material automata nevertheless tend to claim that they are not purely material. Mental phenomena, being epiphenomenal, cannot be responsible for this. And at first appearance, it does not seem that anything else could be responsible for this, making such purely material expressions of dualism paradoxical.

This chapter and the next describe and defend the significance of this paradox, challenging the previously established dogmatic stance of epiphenomenalism and the theory of metaphysical correlation and representation based on this dogmatic stance. This chapter derives the paradox, characterizes its finer points, and describes its significance and repercussions. It begins with deriving this paradox from premises that have already been established in the chapters On Dualism and On Epiphenomenalism. Then, it clarifies what aspect

of the paradox this work focuses on, among other finer points. It concludes with some illustrations, which are not strictly necessary to describe it, but which may be useful for some.

Some of the content of this chapter reformulates doctrines already described in the chapters On Dualism and On Epiphenomenalism, but with a focus that helps bring out the more paradoxical consequences of these doctrines.

On the Purely Material Universe

Minds are correlated with, but distinct from, matter. Therefore, the material universe may be conceived to exist without minds existing with no logical contradiction.

The purely material universe is the hypothetical alternate universe in which the material universe exists in exactly the way that it exists in reality, but in which minds do not exist.

Reality is not the purely material universe, but there is no logical contradiction in the definition of the purely material universe. The purely material universe could exist.

For its behaviour to be explanatorily coherent, it requires the hypothesis that phenomenal consciousness is explanatorily irrelevant. However, this hypothesis has been demonstrated a priori in the chapter On Epiphenomenalism. Thus, so long as one accepts that argument, the purely material universe is well-defined.

In the purely material universe, the purely material automata would tend to believe they were dual, because the automata in reality tend to believe they are dual. They would make the argument, *I think, therefore I am*, and believe it, because the automata in reality do so. But these aphenomenal copies would thus be mistaken in making this argument. So how does the mind from the perspective of which reasoning takes place know that it is not a purely material automaton doing this? Because this mind mentally innately intuits the essence of mental phenomena, what it is for them to be, which is separate from any of its behaviour associated with this innate intuition. But a purely material automaton would say the same thing, but be mistaken, and have no mind doing what it claimed to be doing. But the mind still knows that it exists. But a purely material automaton would still be mistaken in making exactly that ar-

gument. This dualistic–materialistic argument can be followed indefinitely. And later in this work, a more developed interpretive version of the argument, which shows something about the argument that is not apparent at the level of the current discussion.

On Purely Material Automata

The purely material copy of a real material system is the identical copy of that system in the purely material universe.

A real person is a human mind and its correlated body. The person is the particular person, universally quantified, considered for the purposes of the discussion. The body is the body of the person. The automatic body is the purely material correlate of the body. The person is not a purely material automaton. The real body and the purely material body are automata. The real body is a dual automaton. The automatic body is a purely material automaton. The body has a mental correlate. The automatic body has no mental correlate.

The mind is the mind of the person. The mind exists. The person knows that the mind exists. The person references the mind. The person asserts that the mind exists. The behaviour of the person is a tandem act of the mind and the body. The mind asserts that the mind exists. The real body asserts that the mind exists. The purely material body behaves in the same way as the real body. The automatic body states that the mind exists. The automatic body is in the purely material universe. The mind is not in the purely material universe. Suppose the person studies metaphysics and states complex doctrines about the mind. Then, the automatic body studies metaphysics and states complex doctrines about the mind.

In reality, the mind exists and the material universe exists, and the two are distinct. Since they are distinct, one can be conceived as existing without the other without contradiction. The way of thinking employed here is due to David Hume. As a thought experiment, hypothetically conceive the material universe as existing exactly in the way it does in reality, without any minds existing. This alternate universe is called the *purely material universe* in this work. The purely material universe is identical in all respects to the material universe in reality, but the people in the purely material universe are purely material automata. Note that this thought experiment uses the assumption that

mental causal efficacy does not exist. If it exists, then it becomes doubtful whether this thought experiment is meaningful. For, to model the behaviour of the body, it would then become necessary to take into account the state of the mind, and so the body could not be accurately and naturally modelled as a purely material automaton. This difficulty can be surmounted, and the materialistic interpretation of metaphysics can be defined assuming the existence of purely mental phenomena, but to enter into this consideration further in this place would only obscure the exposition, so the present discussion assumes that mental causal efficacy does not exist. Thus, in the purely material universe, all human behaviour may be explained precisely as the reaction of the internal configuration of the body to external stimuli, and more abstractly as resulting from the general organization of the body.

All of the automata in the purely material universe are purely material automata. The mind from the perspective of which reasoning takes place thus knows that it is not in the purely material universe, because it knows that it exists as a mind, through the argument *I think, therefore I am*. However, there are intermediate possibilities between the purely material universe and the universe where all automata are dual. For example, some automata would be dual and some purely material, if there were a certain degree of complexity that an automaton must have for it to be dual, and this degree of complexity was not satisfied by all automata. Or perhaps some could be dual and some purely material according to some arbitrary criterion. Thus, it is possible to consider purely material automata without necessarily considering the purely material universe in its entirety.

On the Paradox of Dualism

The material universe of reality is a closed system. Thus, the purely material universe is a coherent system. Thus, the laws of physics in the purely material are natural, in that there are no arbitrary correction factors for the causal efficacy of the purely mental phenomena of reality, as there are no purely mental phenomena in reality.

It is only possible for an entity to reference something separate from it if it comes into interaction with it. But the automatic body does not interact

with the mind. Therefore, the automatic body cannot reference the mind, or assert that the mind exists, or reference complex doctrines about the mind.

The automatic body appears to reference something outside of its reality. The automatic body makes a false or meaningless assertion when it asserts that the mind exists. The assertions of the automatic body do not mean what they appear to mean. The term *the mind* does not mean the same thing in the purely material universe as it does in reality. In the purely material universe, the term *the mind* references something material, or it references nothing.

The purely material universe is identical in all respects to the material universe of reality, except that in reality, the mind exists. *In reality, the mind exists.* The author has written these words, and they represent his informed belief, still further, the truth. Writing them and believing them is a type of human behaviour. Since the author in reality has stated that the mind exists, and since all his actions are a tandem action of his mind and his body, in particular, his body has written those words stating that fact, and his brain has undergone the cogitations involved in rationally forming the belief in this fact. Since the purely material universe is in all respects identical to the material universe of reality, in the purely material universe, the purely material body identical to the dual body of the author, hereafter called his purely material body, has written those words, and indeed all of the words of this work so far written, and his purely material brain has undergone the cogitations involved in forming this belief. His purely material body is coherently talking about the mind. Why is it doing this? What is it talking about? It must be explained why it is an emergent property of the laws of physics that automata exhibit dualistic behaviour.

On Two Aspects of the Paradox

The paradox of dualism is not new. David Chalmers discusses the paradox under the term *the paradox of phenomenal judgement* in his work *The Conscious Mind*. Indeed, his discussion of the problem in that work is the most well-known formulation of the problem in the philosophical literature, though others had already formulated it before him, as he himself witnesses in that work.

However, his discussion in that work focuses on a different aspect of the problem than that most relevant to the present discussion. He uses that term to

discuss two related but distinct problems, or aspects of the problem. The first problem may be called *the paradox of dualism in the purely material universe*, and the second may be called *the paradox of dualism in the dual universe*.

The paradox of dualism in the purely material universe concerns why automata are innately dualistic in the causally closed purely material universe.

The paradox of dualism in the dual universe concerns how it is possible for the mind to reference itself and know that it exists if the material behaviour underlying this knowledge can be explained without reference to the existence of the mind.

In the account given by Chalmers, and in the relatively sparse philosophical literature surrounding the paradox, most of which is derivative of his thinking, the term *the paradox of phenomenal judgement* is mostly associated with the paradox of dualism in the dual universe rather than in the purely material universe. And in his later work, when he takes up the discussion of this problem again, he also focuses on the paradox in the dual universe. Arguments on this aspect of the paradox often border on chimeric, because they do not make empirically testable claims, and moreover do not even analyze the structure of innately dualistic behaviour, and thus do not contribute much to understanding the structure of the brain or of automata. This work focuses instead on the paradox of dualism in the purely material universe. Solving this paradox is the key to understanding the paradox in the dual universe.

On the Paradox in the Dual Universe

An objector may claim that if the mind exists, it cannot reference itself if the material universe is a causally closed system. Something in a causally closed system cannot reference something outside of that system, because the second thing has no causal efficacy on the first thing, and they thus never come into interaction. Therefore, the body, an automaton in the material universe, a causally closed system, cannot reference the mind, which is outside of the material universe. And if the body cannot reference the mind, then by a similar argument, neither can it reference the minds of other bodies. Thus, the body cannot reference any minds. Next, note that the system consisting of the material universe together with the mind is a closed system. For the other minds of reality, which are the only things outside of this system, have no causal efficacy on the

material universe because the material universe is a causally closed system, and neither do they have causal efficacy directly on the mind itself because that would constitute telepathy, which all scientific investigation so far has shown to be impossible. Therefore, the mind, which is in the closed system consisting of itself and the material universe, cannot reference other minds, which are outside of that system. And neither can the mind reference itself. For if the mind references a thing, then since the mind is a correlate of the body with no causal efficacy, the body must reference that thing in correlation for the reference by the mind to be possible. But the body cannot reference the mind. Therefore, the mind cannot reference the mind. Therefore, nothing in the universe can reference the mind.

A response to this is as follows. The conclusion of this argument is self-contradictory. For the sentence, *Nothing in the universe can reference the mind*, either references the mind, or it does not. If it references the mind, then it contradicts its claim. If it does not, then it is false, because it does reference the mind, insofar as it makes a claim about the mind. Either way, the conclusion is false.

But as a concession, suppose that the conclusion is somehow not self-contradictory. Then the argument is still invalid, because it makes use of a false premise. The false premise of the argument is the hypothesis that a thing in a causally closed system cannot reference a thing outside of that system. This is not true in general. It is true in the material universe alone. For in the material universe, if two things are correlated by nature, it can only be through causal interaction between them, because there is no purely one-way causal efficacy in the material universe, this being the third law of mechanics discovered by Isaac Newton. However, this law of mechanics does not apply to the relationship between the mind and the body, which is one-way. The body determines the mind, and is correlated with it, but is not in causal interaction with it. Thus, the body can reference the mind because it determines it and is correlated with it, even though the material universe is a causally closed system from the physical perspective involving interaction.

But suppose, as a concession, that something in a causally closed system cannot reference something that it determines outside of the closed system, even though in truth it can. Then it would be true that the body cannot reference the mind or other minds, and that the mind cannot reference other minds, as the argument of the objection endeavours to show. But it would still

not be true that the mind cannot reference the mind, as the objector claims. For the argument uses a second intermediate hypothesis that is problematic. In particular, it assumes that something that is determined by and correlated with a closed system can only reference what the closed system can reference. This hypothesis is problematic because it is not obvious that it is true, and because the only case to which it applies is the case of the mind being determined by and correlated with the body. Thus, the objector is doing something akin to assuming their desired conclusion as an unproven hypothesis. And there are independent reasons to doubt this hypothesis. For since it is the essence of the mind to innately intuit its mental phenomena, referencing itself is what it is for the mind to be, which is in addition to any structural properties determined by the body.

On Dualism and Psychical State Automata

An objector may claim that an automaton with a purely psychical state would only formulate a duality between this psychical state and its material integration, but it would not believe in any so-called phenomenal consciousness, or in the possibility of an aphenomenal scenario. Thus, the possibility of psychical state automata solves the paradox of dualism in the purely material universe.

A response to this is as follows. There is, on the contrary, good reason to believe that an automaton with a psychical state would claim that it could experience something that it was like to be it, and that it would thus think of its psychical state, as it interacted with the rest of the causally effective universe, in the same way that normal humans think about their material correlate. Such automata would still wonder why there was something it was like to be that psychical state, and they could conceive of an aphenomenal scenario different from their reality where there was nothing it was like to be it. This is because the physical and psychical, together, may be viewed as an extension of physics. They still involve states interacting according to laws. This causally effective state as a whole is still an automaton, even if a part of this state does not have attributes such as position or momentum. Thus, if physical automata have consciousness and believe they have consciousness and believe aphenomenal copies can exist, then these other automata in a universe where

the laws of causation are different would form similar beliefs, because belief is a product of structure.

Even supposing as a concession that an automaton with a purely psychical state would not believe in any phenomenal consciousness, this cannot be a solution to the paradox of dualism in the purely material or aphenomenal universe, because human beings in reality believe there is something it is like to be them, so they could not, at least according to the very doctrine the objector is advocating, be psychical state automata. Thus, it is still necessary to explain how dualism arises in physical automata without purely psychical states. In any case, the fact that human beings innately believe there is something it is like to be them indicates that it is necessary to explain why the causal organization of automata leads them to believe this, when this causation may be conceived as it is with phenomenal consciousness subtracted.

On the Analogy of an Alternate Universe

The following analogy gives an imperfect illustration of the paradox of dualism in the purely material universe. Suppose scientists open a portal to another dimension that had previously been causally closed from the dimension of the scientists. They find that the alternate dimension is almost exactly like reality, except in the alternate dimension, most people already correctly believed without external evidence that there existed another dimension like theirs with which they were correlated. Moreover, suppose that the residents of this alternate universe had detailed beliefs concerning the state of the universe of the scientists, and that these beliefs mostly happened to be true and the only false ones were from individual confusion, not systematic global error. This would be strange and difficult to interpret.

The analogy is as follows. The dimension of the scientists is like the mental universe. The alternate dimension is like the material universe. The fact that the alternate dimension was causally closed is like the fact that the material universe is causally closed, and that the thought experiment of the purely material universe is thus explanatorily coherent. The fact that the people in the alternate dimension reference the dimension of the scientists is like the purely material body referencing the mind. The fact that this is surprising is like the paradox of dualism.

The scientists naturally suspect that someone else from their dimension has already made a portal and instructed the people of the alternate dimension that the dimension of the scientists exists. This is like the paradox of dualism leading one to believe that the mind has causal efficacy on the body. But suppose the scientists conduct a thorough check, both of their own dimension and of the alternate dimension, and become thoroughly convinced that this had not occurred, and that the two dimensions had hitherto been causally separate. This is like modern scientific evidence and a priori arguments strongly suggesting that mental causal efficacy does not exist. The scientists are thus forced to conclude that it is an emergent property of the laws of physics of the alternate dimension that its residents come to believe in the existence of the dimension of the scientists. Likewise, this work concludes that it is an emergent property of the laws of physics that automata, or at least human beings, are innately dualistic. The resulting challenge for the scientists is to model how and why this occurs. Likewise, the challenge of this work is to model how and why automata are innately dualistic. The scientists may wonder whether the residents of the alternate dimension were truly referencing their dimension if there was no causal connection between the two dimensions. Likewise, the philosopher may wonder if the innately dualistic behaviour of the body is truly referencing the mind, if the existence of phenomenal consciousness is explanatorily epiphenomenal.

This analogy is for illustrative purposes only. In any cases where its logical structure differs from the logical structure of the true thought experiment and paradox of dualism, the analogy should be discarded in favour of abstract reasoning with the original concepts.

On Referring to the Paradox

The term *phenomenal* denotes that the essence of a thing is subjective rather than objective. After all, it is used thus in the term *phenomenal consciousness*. But as used in the term *the paradox of phenomenal judgment*, it has a different meaning. For it is not the fact that the judgement is phenomenal as opposed to material that is important. Indeed, the paradox persists, and is most striking, when it applies to a purely material automaton, whose judgements are thus not phenomenal, but purely material. The term *the paradox of phenomenalistic*

judgement would be more etymologically just, though the term *phenomenalistic* does not have the intended meaning, because it generally means the doctrine that everything is phenomenal, rather than the doctrine of dualism, that some are phenomenal and some material. But this observation naturally leads to the conclusion that the term *the paradox of dualistic judgement* is more apt, which may be more concisely and accurately expressed by saying *the paradox of dualism*. After all, dualism is a dualistic judgement, and all dualistic judgements are expressions of dualism. And this latter term is the term that this work advocates. However, this criticism verges into pedantry, and the term *the paradox of phenomenal judgement* is not an incorrect use of language.

This paradox is closely related to what is called *the meta-problem of consciousness*.

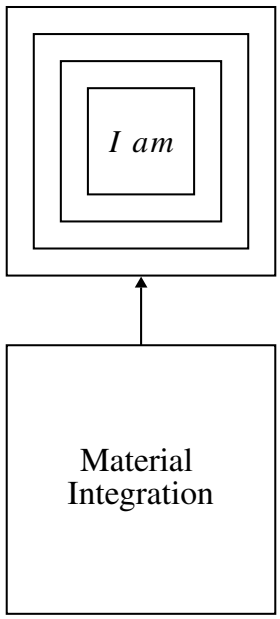

Figure. A purely material automaton formulating the doctrine *I am*. Since the mind does not exist in this case, this contemplation cannot reference the mind. The sentence *I am* in the material representational substrate of the dual automaton referenced the mind, so it appears paradoxical that the same sentence and behaviour in the material representational substrate of the purely material automaton cannot reference this. And if not the mind, what does it reference? Perhaps it references the material representational substrate itself. But if it does so here, why does it not reference it in the dual automaton? Or does it reference the material representational substrate in the dual automaton, too?

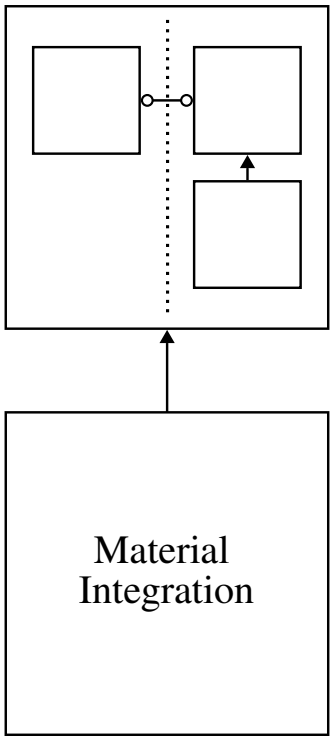

Figure. A purely material automaton formulating dualism. It believes it is conscious, and that this consciousness is distinct from its material representational substrate. But it is mistaken in this belief, because it is purely material. Why does the purely material automaton behave this way? Is the automaton referencing anything at all when it behaves in this way? What does its representational symbol of what it thinks is consciousness represent, if it does not represent consciousness? This expresses the paradox of dualism in the purely material universe.

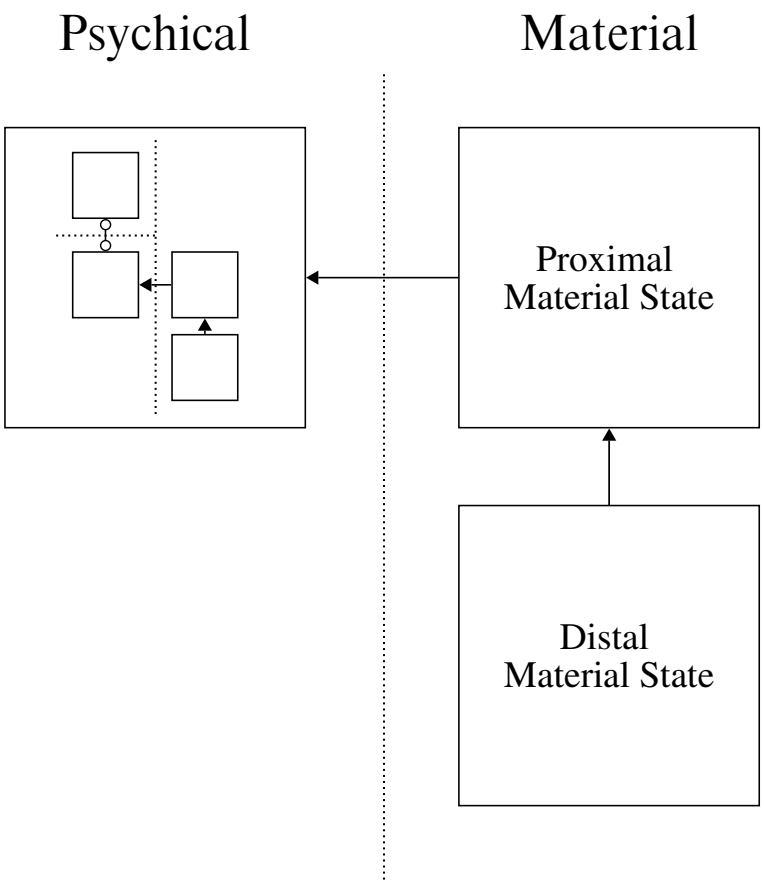

Figure. An aphenomenal automaton with a purely psychical state formulating dualism. It thus falsely claims that it has phenomenal consciousness. The fact that this would occur if aphenomenal copies with purely psychical states existed demonstrates that the paradox of dualism presents an interpretive challenge even if the universe has purely psychical states.

On Responses to the Paradox

Introduction

This chapter endeavours to show that no existing responses to the paradox of dualism are sufficient, and that further doctrines are required. That is, the paradox demands the development of a materialistic interpretation of how automata come to formulate dualism.

This chapter discusses various attempts to solve the paradox of dualism. Some of the attempts are implicit in that the doctrines in question were not explicitly formulated with respect to the paradox, but can be interpreted to implicitly provide an attempt at solving it. Indeed, most of the attempts are of this kind, because the paradox of dualism has not been widely discussed, even by those attempting to construct theories of consciousness. The attempts discussed in this chapter include doctrines the author considered in his initial attempts to solve the paradox but eventually discarded as inadequate on his way to discovering the materialistic interpretation advocated in this work, as well as attempts by other philosophers, both explicit and implicit. All the possible solutions considered in this chapter are inaccurate or incomplete in some way.

On Skepticism, Agnosticism, and Indifference

A possible response to the paradox is not to try to solve it. This can be done either negatively, in an agnostic suspension of assertion or still further complete ignorance or indifference concerning the paradox, or positively, in an assertion that the paradox is unsolvable. The positive claim that the paradox is unsolvable is a form of mysterianism.

Agnosticism is now perhaps the most common response to the paradox of dualism, insofar as the paradox is even considered, which is rarely. In some, this agnosticism may be complete. Especially in a popular context, it may be considered philosophical enough to understand that the thought experiment of the purely material universe leads to a paradox, and a solution to it is considered to be beyond the scope of the discussion. But agnosticism on this problem need not be complete, and may be coupled with some other possible solution. Thus, the following attempts to solve the paradox, including explaining innately dualistic behaviour as an accident, as resulting from evolution, as a social construct, or as natural design, all involve some agnosticism to fill in the gaps where the explanatory mechanism of the proposed interpretation fails to account for all.

For example, none of the scientific or mathematical theories of consciousness that have emerged in recent decades, such as integrated information theory and global neuronal workspace theory, seem to have much to say about the paradox of dualism in the purely material universe. It is possible that some proponents of these theories have used them to speculate in some way on the paradox of dualism, but the author is not aware of any such attempt. So far as the author understands these theories, their main doctrines do not discuss this paradox, nor do their main proponents endeavour to use these theories to explain how innately dualistic behaviour comes about. This is a considerable weakness in these theories, and indicates that they are, at best, studies of empirical correlation guided by theoretical principles, and not fundamental theories of the structure of representation and of consciousness. For any fundamental theory of representation and of consciousness should explain why material innate intuition is structured such that automata talk about having minds.

On the Two Aspects of the Paradox

There are two possible kinds of solutions to the paradox of dualism, divided according to whether one is considering the paradox in the dual universe or the paradox in the purely material universe. The paradox in the dual universe is not of great importance in itself, and has already been solved. Any incompleteness in understanding of it comes from an incomplete understanding of the paradox in the purely material universe. The paradox in the purely material universe has two possible solutions. The first is that this paradox shows that mental causal efficacy exists, and that its agency leads to the reference of the mind. The second possibility is that there exists a materialistic interpretation of metaphysics. Most of this chapter is dedicated to discussing attempts to give a materialistic interpretation of metaphysics.

On the Paradox in the Dual Universe

The sparse philosophical literature on the paradox of dualism usually considers it in the dual universe, namely, whether it is possible for the mind to know that it exists if epiphenomenalism is true.

The solution to the paradox of dualism in the dual universe is as follows. Since experience is immediately present to the mind rather than mediated by an indirect chain of causation, knowledge of its existence need not be dependent on its having causal efficacy, in contrast to most other kinds of knowledge, which are mediated by causation, and which are thus dependent on the thing being known having causal efficacy. Since most other kinds of knowledge are dependent on causation, some philosophers have erroneously claimed that knowledge is dependent on causation, and that if causation from the thing known is not present, knowledge is not present. But this is an unjustified principle based on undue extrapolation from a limited subset of possible kinds of knowledge. This unjustified principle discarded, one sees that the mind can know that it exists. Indeed, this is the most certain kind of knowledge that the mind has, and it is prior to any knowledge about causal efficacy.

This solution is not new. The author is here only paraphrasing what others, for example Chalmers, have said. One problem with this solution, however, is that it does not solve the most important part of the paradox, namely,

the paradox of dualism in the purely material universe. Chalmers has also tried to solve this latter aspect of the paradox, and his attempt is discussed later in this chapter. However, other writers, after showing that the mind can know that it exists, proceed to claim that they have solved the paradox of phenomenal judgement, as it is called in this context. This may be true in a limited sense if the paradox is defined only to involve the question of whether the mind can know it exists. But this is an incomplete notion of the paradox, and thus, more properly speaking, their claim that they have solved the paradox is false. Furthermore, though the above solution to the paradox in the dual universe is true, its truth is superficial, because it ignores the structure of why the automaton comes to form the belief in question. Indeed, though the paradox in the dual universe has been superficially solved, solving the paradox in the purely material universe further removes doubt about it in the dual universe, by describing more clearly the structure underlying dualistic behaviour in the dual universe.

On Mental Causal Efficacy

If one moves past the paradox in the dual universe to the paradox in the purely material universe, there are two directions one can take. The first is to claim that the paradox shows that the mind has causal efficacy. The second is to claim that there exists a materialistic interpretation of innately dualistic behaviour. But in the chapter On Epiphenomenalism it has already been shown that causal efficacy does not exist. Indeed, phenomenal consciousness, whose essence as appearance is distinct from any psychical structure that underlies it, cannot have causal efficacy.

The most important consequence of this doctrine is that the paradox of dualism cannot be resolved by claiming that it shows that mental causal efficacy must exist. Even if purely psychical states exist with causal efficacy on the physical, the question remains of why an automaton with such psychical states claims that there is something it is like to be it, in addition to and separate from the causal structure of those psychical states, if that phenomenal consciousness has no causal efficacy on the psychical or physical states that are talking about it.

This doctrine has already been demonstrated. In case doubt remains, a description of it more closely tailored to the discussion at hand is as follows.

Suppose the material universe, considered as a concession to be only physical, not both physical and psychical, is not causally closed. The laws of physics are unnatural by themselves, and there are in addition to them correction factors accounting for the causal efficacy of the mind. These correction factors may be called purely psychical states. The psychical correction factors represent the structure of the mind. The automatic body could reference these psychical correction factors, because the psychical structure has causal efficacy on the body. But an aphenomenal copy is still conceivable in this situation, and such an aphenomenal copy may still be called a purely material automaton provided matter is defined suitably. It is conceivable that these psychical structures could have causal efficacy, without there being anything that it is like to be these psychical structures or to be correlated with them. That is, the psychical structures could be purely structural. Suppose this is so. But the aphenomenal copy with these psychical structures would not only talk about the physical structure of matter, and not only talk about these psychical structures through representation, but it would also talk about what it was like to be these psychical structures, innately, as if there were phenomenal consciousness that also existed. But the aphenomenal copy would be mistaken in making this phenomenal judgement, because no such phenomenal consciousness would exist. Therefore, the paradox of dualism remains. Instead of being in the purely material universe, it is here in the purely structural universe, or in the universe where physical matter exists and non-physical psychical structure exists, but where there is nothing that it is like to be this psychical structure phenomenally. But such a universe could be called the purely material universe, and such an aphenomenal copy called a purely material automaton, without too much imprecision. For the psychical structure can be called psychical matter. Such psychical matter, in the aphenomenal copy, does not have phenomenal properties. It merely underlies causal efficacy, as does physical matter. Any phenomenal correlates of any kind of matter, physical, psychical, or otherwise, may be separated in principle from the causal structure that underlies them, and thus an aphenomenal copy is always conceivable. Thus, provided that the mind is defined to be merely phenomenal consciousness as separate from any possible psychical structure, the mind necessarily cannot have causal efficacy.

However, with a materialistic interpretation of metaphysics still undescribed, it may be tempting to nevertheless claim that the existence of mental causal efficacy is the most plausible solution to the paradox, and that the arguments given that mental causal efficacy does not exist must have some hidden flaw. Therefore, the strongest justification for the existence of a materialistic interpretation is its exposition, which is contained in some of the following chapters of this work. However, for the sake of completeness, before it is put forward, the remaining sections of this chapter describe further attempts to solve the paradox, which this work ultimately rejects as incorrect or incomplete.

On Materialism

In itself, materialism is not an adequate response to the paradox of dualism in the purely material universe, because in itself it fails to explain that for which this paradox demands an explanation. The paradox demands an explanation for why automata believe in dualism in the purely material universe. The materialist claims reality is the purely material universe. Whether or not this claim is true, the claim in itself does not explain why dualism arises.

Now, various materialists, while endeavouring to argue that reality is purely material, have adduced various theories of why people nevertheless believe in dualism. Indeed, whenever a materialist tries to argue that the mind is an illusion and that matter is all that exists, they are implicitly making an attempt at solving the paradox of dualism, even if they are not referencing it explicitly. For such an argument must implicitly or explicitly assert that there is some mechanism by which the brain comes to believe in the existence of the supposedly illusory mind, and the materialist by virtue of their position must claim that this mechanism is purely material.

However, all these theories are flawed, and they must be flawed, no matter how subtly they are constructed, because they are built on flawed foundations. As is argued later, even the materialistic interpreter believes that it is more rational for a person to believe in dualism than in materialism. Thus, for a materialist to found their materialistic interpretation on the belief in materialism puts an error at the foundation of their interpretation, and this error propagates throughout the whole interpretation, and means that it does not

even correctly model the behaviour of the purely material universe. And in order for the materialist to describe their interpretation without realizing this error, the error must be concealed by sophistry.

Materialists who are trapped in dogmatic entrenchment will find the doctrines of this work destructive to their position in the capacity of their giving a materialistic model, because the materialistic model put forward in this work is the logical conclusion of their imperfect attempts at explaining why the mind is an illusion, and thus refutes them or renders them obsolete. They will paradoxically find the doctrines of this work constructive for their position in the capacity of their being materialists as opposed to dualists, provided they are not so trapped in dogmatic entrenchment as to reject even the materialistic model of this work in itself, because they will be able to make a confused argument why the argument of this work that dualism is true is a false argument, and then take the materialistic model given in this work and use it for their own confused purposes, as an enhancement of their original confused materialistic model. Materialists who are not trapped in dogmatic entrenchment will find the doctrines of this work destructive for their position, both in the capacity of giving a materialistic model of dualistic behaviour and in the capacity of their believing in materialism as opposed to dualism, but constructive for their discovery of truth.

On Idealism

The paradox of dualism is just as much of a paradox for an idealist as it is for a dualist. For an idealist still grants that the body has a structure, and that the behaviour of the mind and the body can be modelled as resulting from the structure of the body alone.

On Dualism as an Accident

One attempt at giving a materialistic interpretation of metaphysics is to say that dualistic behaviour and the mind arise as an accident. This can be advocated by a materialist, or it can be advocated by a dualist who believes that the mind does exist in the way the body is accidentally configured to reference it. This system may be put forth as follows. To represent reality, the human

nervous system integrates sensory stimuli and forms imperfect representations of it. To interact with reality, it forms imperfect representations of its state and its outgoing volitions, enabling high-level and global analysis of such states. Since this representation differs from what is represented, because the human intellectual faculties are imperfect, it is naturally referenced as something different from what is being represented. It is called the mind, whereas what is represented is called matter. This is an accident of the imperfections of the human mental faculties. And the mind does exist in reality in the way the nervous system is accidentally configured to reference it. It only makes sense that it would exist thus. As for any elaborate system endeavouring to explain this any further, it could be true, and perhaps not. But it is a general principle to accept the simplest model that describes all that is observed, and this principle casts doubt on any attempt to elaborate a materialistic interpretation of metaphysics beyond the rough sketch given here. A priori, the nervous system could exhibit dualistic behaviour, or it could not. In reality, it does, and this is an accident.

A response to this is as follows. The process that the rudimentary material interpreter describes, whereby the nervous system comes to reference the mind, is not strictly incorrect. But it is an unrigorous and incomplete description of the full materialistic interpretation. The interpreter says that an explanation should be as simple as possible, but their explanation is too simple, because it breaks down when more precise questions are asked. For example, the rudimentary materialistic interpretation explains why the material representational substrate is considered to be different from external represented matter, but fails to explain why the material representational substrate intuited innately, and called the mind, is considered by the dualistic automaton to be different from the material representational substrate when it is intuited through representation like external represented matter, and called the material representational substrate. The full materialistic interpretation put forth in the second part of this work explains this. The rudimentary material interpreter claims that the human mental faculties could be configured not to reference the mind, perhaps if they were more sophisticated, but fails to give a precise model of an innately materialistic automaton. As is be shown later in this work, the nervous system must exhibit innately dualistic behaviour. It is not an accident that the nervous system references the mind.

On Dualism and Reproductive Fitness

A more precise material model of why innately dualistic behaviour comes about invokes evolutionary theory. It was already mentioned in the derivation of the paradox of dualistic behaviour, and it may be elaborated as follows. Given that the mind exists, and given that people were designed by evolution, it is reasonable to suppose that the mind serves some purpose evolutionarily. Its benefit may be due to mental causal efficacy on matter. William James believed this. Or perhaps evolution has merely favoured the formation of automata that behave as if they have minds, and in addition, minds happen to exist in the way automata have evolved to refer to them. It only makes sense that minds should exist thus, if they exist at all.

A response to this is as follows. This is a stronger attempt at giving a material model of innately dualistic behaviour than the last one, because it is more precise. However, it is similarly flawed because it claims that innately dualistic behaviour is a result of evolution, and that if it had been more favourable for innately materialistic automata to evolve, they would have evolved. But this is not the case. The fact that all automata are innately dualistic is a necessary limitation on the effectiveness of representation and a constraint within which evolution must act. This is shown in the third part of this work.

An evolutionary analysis further illustrates the paradox of dualism. According to the theory of evolution discovered by Charles Darwin, the general structure of living beings is a consequence of natural selection, the preservation of favoured variations in the struggle for survival. To a first approximation, any given aspect of structure is either the direct consequence of selective pressure, or an accidental epiphenomenon not under much selective pressure, or a consequence of another fundamental principle other than evolution that is a constraint within which evolution must work. This three-part distinction could be refined, but this would be beyond the purposes of this discussion. An aspect of structure is not a material subset of the body in itself, but also the consideration of that subset being thus rather than otherwise in some particular way. For example, a hand in itself is not an aspect of structure, but the fact that the human hand has five fingers rather than some other number is an aspect of structure. An example of an aspect of structure being the consequence of selective pressure is human beings having more developed brains than their immediate relatives among the great apes. An example of an aspect of structure

being an accident is how all vertebrates share a common basic bone pattern, for example having five fingers, or at least the bone vestiges of five fingers, on each hand. This is an accident of history, likely because the common ancestor of all vertebrates had this bone structure. An example of an aspect of structure being a consequence of a more fundamental principle is the fact that the metabolism of all living beings must work within the laws of thermodynamics. They did not evolve to follow these laws, with systems that did not follow them tending to produce less reproductively fit offspring, for it was never possible for systems not following them to exist in the first place. Nor is it an accident that they follow them, because all systems must follow them. Rather, evolution had to work within the constraints of these laws from the start.

What aspect of the structure of the brain causes the brain to talk about what seems to be the mind? In other words, what aspect of the structure of the human brain leads to innately dualistic behaviour as opposed to innately materialistic behaviour? Is it under selective pressure, or is it an accidental epiphenomenon of selective pressure on other structures, or is it the consequence of another fundamental principle? If it is an epiphenomenon, and if automata could have been innately materialistic with no sense of mental self, then there still remains the task of explaining the accidental mechanism by which it evolved as an epiphenomenon. But this aspect of structure is deeply related to the configuration of the intellectual faculties, and is not arbitrary like the number of fingers on a hand, and so is likely not an accidental epiphenomenon. Something so important as the mental sense of self is likely under selective pressure, if not necessarily the direct consequence of natural selection. If it is under selective pressure, the task remains of explaining how such a structure arose under natural selection. Either it is a direct consequence of natural selection, or it is the consequence of a constraint within which natural selection must work.

Suppose it is a direct consequence of natural selection. Then this aspect of the structure of the body serves a purpose significant for survival and reproduction. It is subject to selective pressure, and is as it is for the sake of reproductive fitness, at least up to the degree of perfection and precision possible for a biological system subject to the complexities of reality. This structure should be as it is for the sake of reproductive fitness. Does building an innately dualistic automaton with a sense of mental self lead to a biological system more reproductively fit than one that is innately materialistic? If so, how and

why is this so? How is it beneficial for reproduction and survival to talk about what seems to be the mind? And since it is complex, it likely evolved gradually, with each step determined by selective pressure. How does selection lead to the development of this structure gradually over time? It should be possible to identify how this aspect of structure evolved gradually, with each step determined by selective pressure. And if, as has been assumed, purely mental phenomena do not exist, then the development over time of the material universe can be perfectly modelled as that of an isolated system. Within this isolated system, why would there evolve biological automata that reference that which exists outside the isolated system, and which therefore exerts no influence on their development at any time? How has the purely material process of evolution by natural selection led to the development of a race that has produced the science of metaphysics?

Or perhaps the structure underlying innately dualistic behaviour is neither an accident nor an adaptation, but the consequence of some other, more basic principle of how any automata in the material universe can form representations. If so, what is this principle, and why must evolution work within it? Does it apply to all automata, or only to systems designed by evolution, or only to biological systems of a certain chemical makeup, or perhaps only to human beings? If it applies to all automata, how does it arise as an emergent property of the laws of physics? How could it be an emergent property of the laws of physics in the causally closed material universe that all automata talk about something outside that causally closed system?

A summary of the argument of this section is as follows. It is a pattern of human behaviour to reference the mind. Therefore, some aspect of the structure of the body must determine this behaviour. This aspect of structure is either a direct consequence of natural selection, an accidental epiphenomenon, or a consequence of some other principle. In all three cases, the paradox of dualism remains. It is unclear how systems in the causally closed material universe come to talk about something outside this causally closed system.

The correct answer is the third possibility. This is demonstrated later in this work. Innately dualistic behaviour is not selected for by evolution, nor is it an accidental epiphenomenon. Rather, it is a necessary consequence of a fundamental principle of how any material system whatsoever, be it organic or inorganic, developed by evolution or not, can form representations. The fact that all automata are innately dualistic is a constraint within which evolution must

act, rather than a consequence of evolution. But to demonstrate this, first the materialistic interpretation of metaphysics must be defined.

On the Mind as a Social Construct

A more precise attempt at defining a materialistic interpretation through evolutionary theory is to say that innately dualistic behaviour is a social construct.

This rudimentary materialistic interpretation may be put forth as follows. The mind does not exist. The dualist is not mistakenly referencing any particular matter when they talk about the mind. They are talking about nothing. This is so because it is an elaborate social custom among human beings to act as though they possess minds. This custom is so deeply ingrained that it is not learned by experience, but it is inherent in the structure of the brain from birth. The nervous system, instead of acting like it is an automaton directed by itself, which it would do if it had infallible self-awareness, pretends there exists an immaterial mind with experiences, emotions, and thoughts, which is the supposed owner and director of the body. The body is an automatic puppet deluding itself that it has a puppet master. Why do human beings act this way? The simplest reason is inertia. If all other human beings act this way, then there is selective pressure to conform to the custom to fit in. But this inertial explanation does not explain how the custom originated in the first place. The original cause of the custom is that fabricating an illusory self facilitates social behaviour, especially hypocrisy and deception, both of others and oneself. To survive in human society and to succeed, it is necessary to put on a mask and act in certain appropriate ways, without showing everything one thinks. Human social interaction is a hypocrisy. Therefore, it is useful to fabricate a self as a mask for interaction. This is the original reason why human automata began referring to what appears to be the mind. It was a useful abstraction for deceiving oneself and others. And the rudimentary beginning of this behaviour was likely long before the human species evolved. For even animals act according to a rudimentary sense of self. Philosophers and scientists have subsequently studied the precise structure of how this social construct is fabricated, but all they are studying is the construction of a deceptive illusion.

The simplest response to this is as follows. If this alternate materialistic interpretation truly claims that phenomenal consciousness does not exist, then

the proof due to Descartes refutes it. This demonstration that the mind exists is undeniable, and immediately establishes that any doctrine that claims otherwise is false.

However, the advocate of the mind as a social construct may refine their position, as is done for the true materialistic interpretation later in the development of interpretive dualism, by saying that the mind does exist in the way that automata are configured to reference it as a construct delusionally. After all, it makes sense that reality would be configured thus.

The response to this refined theory of the mind as a social construct is as follows. This model, being a refinement of the preceding evolutionary model, shares the flaw of that model outlined above.

More specifically to this refined materialistic model, note that social behaviour, insofar as it is social behaviour, does not require, and is not facilitated by, most kinds of mental phenomena, or dualistic judgements concerning mental phenomena. To see this, note that mental phenomena are either sensory perception or thought. Sensory perception in itself has little to do with social behaviour in the capacity of being social behaviour. The purpose of sensory perception is to represent the physical structure of the external world. If the mind were a social construct, then sight, hearing, and such basic sensory perception would likely not be a part of the mind. Only emotions and other kinds of abstract feelings are part of social behaviour in the capacity of being social behaviour, if they are even categorized as sensory perception rather than as thought. Next, consider thought. Thought is either memory, volition, imagination, or dream. Again, memory serves a basic functional purpose that has little to do with social behaviour in the capacity of being social behaviour. Dreams are more obscure in this connection. Likely, only imagination and volition are the instruments of social behaviour in the capacity of being social behaviour. This suggests that the structure of the mind is derived from something other than social behaviour, because basic sensory perception and memories are part of the mind.

But as a concession, suppose that these redundant mental phenomena exist to make social behaviour complete, and are necessarily a part of the mind as a social construct. Yet social behaviour still does not require, and is not facilitated by, the construction of the self in the way it exists in reality. For the interpreter of the mind as a social construct must claim that human beings could be less innately dualistic than they are in reality, and that the reason they

are more innately dualistic in reality is that such innately dualistic behaviour facilitates social behaviour. Thus, suppose people could have evolved to be less innately dualistic than they are in reality, that is to say, that the nervous system intuited innately with such sophistication that it felt less duality between that which it intuited innately and that which it intuited through representation. The interpreter of the mind as a social construct must claim that this alternate state of affairs has not come about because such more advanced human beings would be at a social disadvantage, and less able to succeed in society, and thus that innately dualistic behaviour gives an evolutionary advantage by facilitating social behaviour. But this is a contradiction. For a less innately dualistic automaton, if it could exist, would be more sophisticated than a more innately dualistic automaton, and would be more capable of more refined social behaviour, being more aware of its inner workings. Humans are innately dualistic automata because of innate limitations on the strength of their mental faculties. This limitation does not serve a useful purpose.

An objector may claim that unthinking hypocrisy is often more useful than deliberate, rational hypocrisy.

A response to this is as follows. Though there are always exceptions, in general, greater strength of the intellectual faculties is more advantageous. The rapid increase in brain size in the recent evolutionary history of the human race demonstrates that there has been selective pressure for stronger intellectual faculties.

On the Genealogy of Dualism of David Hume

Metaphysics had not advanced sufficiently by the time of David Hume for it to have been possible for him to give a response to the paradox of dualism in the purely material universe as it is now understood. However, in the section Of Scepticism with Regard to the Senses in the part Of the Sceptical and Other Systems of Philosophy in his work *A Treatise of Human Nature*, he speculates on the causes for why people believe in dualism, and his remarks may be interpreted within the context of the present discussion.

Hume states that nature is responsible for the human belief in the existence of external objects despite the impossibility of proving this for certain. Then, using a combination of philosophical and naturalistic reasoning, he goes

on to give a detailed description of why people come to believe in dualism. His description goes approximately as follows. He begins with explaining why people initially believe in naïve monism, then how they may realize it is false, and then why philosophers posit a dualistic system according to which mental phenomena imperfectly represent, but are distinct from, matter.

This discussion shows remarkable prescience. The focus on naturalistically describing why people come to believe in dualism, as opposed to focusing on arguing over whether the belief in dualism is right, is particularly valuable. However, in itself it does not solve the paradox of dualism. This is for a few reasons.

Hume does not explicitly formulate the discussion as applying to the behaviour of a purely material automaton. On the contrary, he formulates it for the dual case, and explains how the mind comes to believe in the existence of matter. Hume does not explicitly advocate interactionism, and on the contrary, his doctrines appear to be more in line with epiphenomenalism, but since materialism, and thus its dualistic counterpart epiphenomenalism, had not been clearly formulated by his time, his discussion is in itself ambiguous as to if or how it could be made to apply to a purely material automaton.

Furthermore, Hume seems to advocate that innate dualism evolved because it is useful, as the sentence in the above quotation referencing nature suggests. Now, in his time the theory of evolution had not yet been discovered. But earlier theories concerning natural design often lead to similar conclusions as evolutionary theory, and they do so here. Thus, in the context of the modern understanding of evolutionary theory, his remarks may be interpreted to claim that innately dualistic behaviour evolved because it is conducive to reproductive fitness. For if people stopped trying to survive because they were convinced of skepticism concerning external objects, they would die. Now, this explanation is in one sense true. The particular kind of innately dualistic behaviour that has evolved did evolve because it is reproductively fit. Evolution is responsible for people not innately being philosophical skeptics. But, as this work endeavours to show, evolution is not responsible for people, or any automata, being innately dualistic. Rather, the innate dualism of automata is an a priori law of design within which evolution must work. But this is discussed in more detail later in this work.

In addition, Hume phrases his discussion without taking into account the notion of a material correlate of consciousness, and his discussion is thus

not precise enough to be interpreted in itself as a genealogy of modern dualism as opposed to modern materialism. For a materialist may grant that his genealogy correctly explains why people believe that their internal representational states differ from external objects, but this in itself does not establish that these internal representational states are not a part of the brain. Ultimately, Hume only gives a refutation of naïve monism, which is in a sense naïve materialism, but not of more modern materialism, because more modern materialism had not been formulated when he was writing.

On Dualistic Behaviour as Natural Design

Perhaps the best previously published response to the paradox of dualism in the purely material universe is to be found in the work of David Chalmers. In the section On Explaining Phenomenal Judgments in the chapter The Paradox of Phenomenal Judgement in his work *The Conscious Mind*, and later in the section Some Supporting Arguments in the chapter Consciousness and Information: Some Speculation, Chalmers speculates why an automaton would come to exhibit what he calls *phenomenal judgments* according to its material structure alone, irrespective of whether it is dual or purely material. He communicates this explanation mostly using examples, as an aside that is not the primary focus of his work.

An abstract summary of the materialistic interpretation of Chalmers using the terminology of this work is as follows. It contains some elaborations to foreshadow the further doctrines of this work.

Suppose one is designing an automaton to distinguish representational states of different kinds. It is more efficient to program it to justify its judgement that two states are different by having it somewhat irrationally claim that it merely intuits them thus, than to program it to justify its judgement with a scientific description of the process of material innate intuition being judged. If all that is necessary is to distinguish the states, then all the justification for that distinction should be an abstraction of the distinction itself. This is not a real justification in the scientific sense, but then again, it is all the justification the automaton needs, because the goal of an automaton is usually to fulfill some task, rather than to model its own behaviour scientifically. This efficient way of distinguishing representational states will naturally lead to a concept of

quality and appearance, that is, phenomenal judgement. By judging that a state is different from another because it is innately intuited thus, rather than describing that innate intuition scientifically, the automaton already comes to form an inner world that is analogous in many ways to the mind, at least epistemically, if not ontically. And one can imagine how this principle of design would also be adopted by evolution, because evolution is a great seeker of efficiency. And thus, one can imagine how evolution would design the brain following this principle. Suppose the brain is asked, *Why is this visual stimulus different from this one?* It will naturally reply, *Because I intuit them thus as different.* Further justification for the innate intuition referencing the structure of the brain itself would be inefficient and likely selected against by evolution.

This is the materialistic interpretation of innately dualistic behaviour as efficient design. Biologically, this materialistic interpretation is more specific than the interpretation of innately dualistic behaviour as a consequence of evolution, because it makes a more precise claim about how exactly innately dualistic behaviour is selected for by evolution. But this interpretation also applies to automata that are not designed by evolution, so it is not merely one particular version of the evolutionary interpretation.

The materialistic interpretation of innately dualistic behaviour as efficient design strikes quite close to the truth, but it has a number of weaknesses that mean it is incomplete.

This is an admirable train of thought that shows great prescience. However, it is only an uncertain and imprecise sketch, and by itself it is not a sufficient response to the paradox. It is not elaborated in detail, and it is mostly communicated with particular examples, rather than as an abstract principle of design. Thus, though many of the claims of this attempt at a materialistic interpretation of metaphysics are true, they only strike at a portion of the truth, and leave much missing. Chalmers himself admits this readily.

This criticism applies as well to the later work of Chalmers along similar lines relating to what he calls *the meta-problem of consciousness*, which is similar to what he earlier called *the paradox of phenomenal judgment*. He does not give a satisfactory answer to either of these problems, nor does he claim to do so.

The imprecision of this interpretation limits it to being first-order. It does not give a precise second-order materialistic interpretation of how the automaton still affirms that the mind knows that it still exists even though there

exists a materialistic interpretation of its innately dualistic behaviour, nor a third-order interpretation of why the automaton still affirms that the mind knows that it still exists after understanding the second-order materialistic interpretation, and so on. More descriptive precision is necessary to descend into recursive interpretation rigorously. This descriptive precision requires moving past a heuristic explanation of how innately dualistic behaviour comes about, and towards the materialistic interpretation of that innately dualistic behaviour itself.

The notions of efficiency, usefulness, and reasonableness, though they appear on the surface sensible to use in this connection, do not accurately describe why innately dualistic behaviour comes about. This work endeavours to show that the correct notion is that of necessity. It is not merely efficient, useful, or reasonable for an automaton to be innately dualistic. It is necessary for it to exist thus. Even if with no concern for efficiency one tried as hard as one could to create an automaton that was innately materialistic, one would only ever succeed in building an automaton that was very representationally materialistic, but all of its representational materialism would still be built on an implicit but all-encompassing substrate of innately dualistic material innate intuition.