CONSCIOUSNESS

A DUALIST PHILOSOPHY

HANE HTUT MAUNG

CONSCIOUSNESS

A DUALIST PHILOSOPHY

CONSCIOUSNESS

A DUALIST PHILOSOPHY

HANE HTUT MAUNG

LULU PRESS

Published by Lulu Press
627 Davis Drive, Suite 300, Morrisville, North Carolina 27560
www.lulu.com

This is a comprehensively edited reissue of a philosophical work that was originally published in 2006, when it appeared as *Consciousness*. The present book is the definitive form of that philosophical work, the official title of which is now *Consciousness: A Dualist Philosophy*.

First published 2006
Edited and reissued 2024

This book can be cited as: Maung, H. H. (2006). *Consciousness: A Dualist Philosophy*. Morrisville: Lulu Press, edited reissue 2024.

ISBN: 978-1-4466-8033-9 (previously 978-1-84728-758-8)
DOI: 10.5281/zenodo.10214048

For my parents

Contents

A note on the 2006 original publication

This book was published mainly as a personal project to explicate some of my philosophical thinking on the topic of consciousness and is comprised of original content along with work based on a Bachelor of Arts dissertation which I completed at the Department of History and Philosophy of Science at the University of Cambridge in 2005. I am deeply grateful to my academic supervisor, Professor Peter Lipton, for his invaluable instruction while I was completing the dissertation and for his encouraging endorsement of the philosophical value of my work while I was advancing the main thesis further through this book.

A note on the 2024 edited reissue

Professor Peter Lipton passed away on 25th November 2007, a year after this book was initially published. With this comprehensively edited reissue of the book, I wish to pay tribute to Professor Peter Lipton, whose influence will forever be acknowledged and appreciated.

I
Subjective Existence

> "… it must finally be established that this pronouncement 'I am, I exist' is necessarily true …"
>
> — René Descartes

Understanding consciousness

This book is a philosophical enquiry into consciousness. While consciousness is undoubtedly the phenomenon with which we are most acquainted, attempts to understand it have tended to be unsatisfactory. Much of this reflects the widespread unclarity in the literature regarding the concept of consciousness. This is understandable, for consciousness seems to elude our empirical methods of investigation. However, as I am a philosopher approaching the subject, I hope to show herein that a clearer account of consciousness is philosophically important and that philosophical analysis can contribute much of value to our knowledge of the phenomenon. Given that philosophy is a discipline that aims to attain truth and understanding through conceptual clarity, it is reasonable to think that it is important to attain clarity regarding the meaning of the concept of consciousness in order to attain a true understanding of the essential nature of consciousness. In this book, I shall propose a dualist philosophy of consciousness which acknowledges that consciousness is fundamental to the very understanding of existence.

As noted by David Chalmers (1996), consciousness often gets conflated with various psychological and neurophysiological features. These features are often spuriously labelled "consciousness", but they do not correspond to consciousness in the relevant philosophical sense. Rather, they are capacities and processes that are involved in the production of behaviour. While these features get explained in detail, consciousness itself gets overlooked. I describe some of these features below.

Awareness: This is perhaps the feature that is most commonly conflated with consciousness. It refers to the ability of an individual to access information from its environment and its capacity to use this information for further processing, usually leading to a change in bodily state or the generation of behaviour. The term is very general,

and so some of the other features described below can be considered to comprise different sorts of awareness.

Introspection: This refers to a particular kind of awareness. Specifically, it is one's reflexive awareness of one's own psychological state. An example is one's ability to recognise and evaluate a cognitive or affective state, such as worry or anxiety, and to use this awareness to modify one's subsequent cognition and behaviour in a given situation.

Reportability: This refers to one's ability to communicate, to others, the informational contents of one's awareness. This can be analysed in terms of the ability to introspect coupled with the capacity for language.

Self-awareness: As the name suggests, this refers to one's awareness of oneself. Specifically, it is one's capacity to acknowledge oneself as a unique agent who is distinct from others. In principle, this can partly be explained with reference to a cognitive model that has access to some sort of representation of oneself as an agent.

Perception: This refers to the act through which information, acquired through awareness, is further processed into a comprehensible representation of the object that is being perceived. An example is the location of a sound source by the auditory system, a capacity which can partly be explained in terms of neural coding mechanisms and the integration of bilateral input from the cochleae.

Cognition: This is a general term that refers to the capacity to receive, process, store, and use information. In an organism, it is implemented by the nervous system in conjunction with the environment wherein the nervous system is embedded.

Volition: This refers to the behaviour that results from the act of cognition or deliberation. Voluntary acts are those acts that are performed intentionally. That is to say, they are acts that are motivated and endorsed by some aspect of reflexive thought.

Wakefulness: As the term suggests, this refers to the state of being awake. Such a state can be defined in terms of one's responsiveness to stimuli and capacity to process such stimuli. Its explanation may involve reference to the level of activity in the individual's brainstem reticular activating system. People who are comatose, anaesthetised, or syncopal are often erroneously said to have lost "consciousness", but it is, in fact, wakefulness that has been lost. Likewise, different neurophysiological states are often erroneously said to be associated with different "levels of consciousness", but these actually refer to different levels of wakefulness.

Attention: This is the selective focus of awareness on a particular stimulus, and is partly explainable in terms of the levels of activity in different parts of the nervous system.

Knowledge: This refers to the capacity to believe and recall a fact. From a philosophical perspective, knowledge is factive. That is to say, it involves a true belief. From a psychological perspective, knowledge requires the capacities for cognition and intentionality.

Intentionality: This relates to the notion of representation and refers to the way that the contents of awareness tend to be about things in the world. That is to say, the contents of awareness have propositional content. They are representations of the features that are being perceived. An explanation of this may involve an analysis of how a system organises information from a stimulus into a comprehensible form that is representative of the stimulus, as well as an analysis of how social and linguistic norms and practices influence the meanings and uses of expressions.

Further to the above, the term "consciousness" has also been used in political theory. For example, the notion of "false consciousness" does not refer to consciousness in the philosophical sense being used here, but is an expression that was coined by Friedrich Engels (1893) to describe how a subordinate class are misled by the ideology of a ruling class in a capitalist society. While this is a useful concept in political theory, it does not pertain to the concept of consciousness that is relevant to the current philosophical analysis.

Importantly, the psychological features described above are structural and dynamical features. They pertain to the processes involved in behaviour. Given that they are structural and dynamical features, they can be explained in terms of how the organisation of a system allows it to respond to the environment, process information, and execute behaviour in appropriate ways. For example, reportability can be explained in terms of the causal dynamics that link the reception of a stimulus to the behavioural act of reporting that the stimulus has been perceived. Similarly, perception can be explained in terms of the mechanisms through which the nervous system codes and integrates inputs in such a way that their resulting states can have appropriate roles in directing cognitive processes.

Although the above psychological features are highly complex, they pose few metaphysical problems. This is not to say that there cannot be worthwhile philosophical discussion about these features. Physical explanations of such features may be possible in principle, but current accounts of these features are hazy, and so philosophical analysis can contribute much of value by attaining greater conceptual

clarity. For example, epistemology has much to say about what constitutes knowledge. Likewise, the studies of intentionality can provide much valuable insight into what it is for a state to be representational. Nonetheless, these psychological processes are not ontologically baffling and there is no reason why they could not be explained in physical terms.

The conflation of consciousness with the aforementioned psychological features is evident in the various accounts which claim to explain "consciousness", but which actually only explain these psychological properties while leaving consciousness unexplained. For example, Daniel Dennett (1991) gives an account of one's to report one's own mental states. This is really an account of introspection and reportability, but Dennett inaccurately claims that it is an account of "consciousness". Similarly, Paul Churchland's model (1995) is inaccurately presented as a tentative account of "consciousness", but it is really an account of certain perceptual and cognitive processes. The above reflects a common misunderstanding of what consciousness actually means. While the above psychological features are inextricably linked to the concept of the mind, they omit something from the picture, namely the subjective quality of experience. Hence, I argue that the conflation of consciousness with a third-person psychological feature amounts to a false definition of "consciousness" which fails to capture the first-person subjectivity that is essential to consciousness.

In addition to the structural and dynamical properties of the mind, there is also a subjective aspect, which pertains to the experience of a mental state from a first-person viewpoint. When I look at a red object, a bustle of neural activity and information processing occurs, but this activity is also accompanied by my having the subjective experience of red. As Thomas Nagel (1974) notes, there is "something it is like" to have a conscious experience. It is this subjective quality of experience that is relevant to my analysis of consciousness. The psychological properties discussed above may be contingently correlated with experience, but they fail to capture the first-person subjective quality of experience, and so they are not as directly relevant to the philosophical analysis of consciousness.

Some further clarification is required here. Phenomenal qualities, or qualia, are obviously relevant to the study of consciousness and many contemporary philosophical analyses of the mind recognise that they are important. However, it would be a mistake to conflate qualia with consciousness, as this would erroneously imply that there is nothing more to "consciousness" over and above these experiential

qualities. It is false to claim that qualia exhaustively comprise consciousness, for consciousness and qualia are different features. Qualia are the phenomenal qualities which are experienced, whereas consciousness is the experiencer of these phenomenal qualities.

From my acquaintance with myself as a conscious subject, it is clear to me that I am not a collection of qualities, but rather that I am the experiencer of these qualities. Indeed, it is necessarily true that an experience entails an experiencer. This is a conceptual truth that holds in virtue of the fact that being experienced by an experiencer is what makes something an experience, for an experience is subjective by definition. It is also an ontological truth that holds in virtue of the fact that the existence of an experiencer is a necessary condition for anything to manifest as an experience.

While qualia depend on consciousness, consciousness is a distinct feature from the qualia. Given all the facts about qualia, the first-person individuation of consciousness remains a further fact to consider. For example, the qualitative character of a given phenomenal quality may capture what that quality is like, but it does not capture the fact that this quality is individuated to *me* and not to *you*. The fact that *my* consciousness rather than *your* consciousness experiences the phenomenal quality is a further fact beyond the qualitative character of the quality. Also, as illustrated by Ibn Sīnā's (1027) thought experiment of a floating man who has no sensory input, consciousness still exists even when there are no qualia. Without such qualia, consciousness obtains as a pure first-person existence and maintains the potential to experience qualia that are individuated to this first-person existence. And so, it must be taken as true that consciousness is a separate entity from the qualia that it experiences. This distinction between qualia and consciousness roughly equates to George Berkeley's (1710) distinction between "any one of my ideas" and the "thing entirely distinct from them, wherein they exist", as well as to Erwin Schrödinger's (1944) distinction between "a collection of single data" and "the canvas *upon which* they are collected".

Therefore, in the relevant philosophical sense, I propose that the true definition of consciousness is first-person subjective existence. This is a definition of consciousness that is often accepted in philosophy, as it equates broadly with Chalmers' (1996) "subjective quality of experience", as well as with Sāṃkhya philosophy's notion of *puruṣa*. Importantly, it captures the subjectivity of Nagel's (1974) "something it is like". Nonetheless, this definition involves some reflexivity, because one's understanding of it hinges on one's first-

person acquaintance with consciousness. Still, it is a meaningful definition which accurately denotes the feature that is essential to consciousness. It clarifies that consciousness refers to the existence of first-person subjectivity wherein experiential qualities manifest. This distinguishes consciousness from the structural and dynamical properties involved in cognition, as well as from the particular qualities it experiences. In the philosophical sense that is relevant here, the true meaning of consciousness is the first-person existence with which one is experientially acquainted as a subject.

The uniqueness of consciousness

As noted above, the subject of consciousness is frustratingly intangible to our conventional methods of definition. Accordingly, Stuart Sutherland's entry on consciousness in the *International Dictionary of Psychology* (1989) states that "the term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means". Indeed, the definition I give above, while informative, is also reflexive, as first-person acquaintance with consciousness is required to grasp what it means.

It may be tempting to suppose that the apparent intangibility of consciousness to our conventional methods of definition suggests that the dispute over consciousness is merely linguistic and does not reflect anything real, but I argue that this sort of semantic deflationism is false with respect to consciousness. First, the reality of something is not necessarily dependent on one's ability to define it. After all, prelinguistic infants interact with features in their environments without being able to define them, yet these features are real. Second, consciousness can be understood in another way. Indeed, while there is some reflexivity involved when defining consciousness, it is nonetheless true that the definition of consciousness as first-person subjective existence precisely denotes the correct meaning of consciousness and accurately refers to the actual feature that is essential to consciousness. Hence, the concept of consciousness is clear, even though first-person acquaintance with consciousness is required to understand it.

In fact, I argue that the intangibility of consciousness to our conventional methods of definition is not a barrier to understanding consciousness, but reveals something about its nature. The reflexivity of consciousness reveals an important truth about consciousness, namely that consciousness is an ontologically unique

phenomenon that can only be understood through itself. Hence, any attempt to describe consciousness in terms of other properties would be unsuccessful, because consciousness is unlike anything else we know. Consciousness is a *sui generis* phenomenon of its own kind.

Many of the features in the world with which we are familiar have a third-person ontology. Objects are experienced as other and are taken to be part of the objective world. The term "objective" is often associated with the term "physical", but objective is a broader concept. What defines something as objective is its third-person ontology, whereas what defines something as physical is its structure and dynamics. Physical features are necessarily objective, because structural and dynamical facts are third-person facts, but not everything that is objective is necessarily physical. Abstract concepts, such as numbers, may not be physical but nonetheless have a third-person ontology, and so are objective. We can also conceive of fictional substances, such as ectoplasm, which are not physical because they do not interact with the structure and dynamics of the world, but are objective due to their third-person ontology.

The notion of the objective world, as I am using it here, does not encompass the totality of existence. The objective world encompasses the set of things that have a third-person ontology, but this does not exhaust everything that exists. Rather, the totality of existence comprises everything that exists, which includes the objective world and the subjective existence that is consciousness.

As noted by John Searle (1992), consciousness has a first-person ontology. This makes consciousness a different kind from the objective world, which has a third-person ontology. Consciousness is not an object that is experienced, but it is the subject that experiences. Hence, consciousness is essentially subjective. I have hitherto been describing consciousness as first-person existence, but it is more accurate to say that *my* consciousness is *my* first-person existence. My consciousness is the "I" that I *am*.

The distinction between first-person subjectivity and third-person objectivity does not amount to the mere perspectivalist claim that the first-person and the third-person represent different perspectives. Such a perspectivalist claim may partly recall Albert Einstein's (1916) notion of a body of reference, which is a coordinate system that standardises measurements of spatiotemporal events relative to a chosen point. Rather, the distinction amounts to a substantial ontological claim about first-person subjectivity as the mode of existence which is the *sine qua non* of experience. As Dan Zahavi (1999) notes, experience does not occur in a third-person objective

space, but necessarily presents in the first-person existence of a given subject. This indicates that first-person subjectivity is not an abstract body of reference relative to which the objective world is observed, but is the distinct form of existence which is essential to experience. Accordingly, the complete perspectival facts about a chosen perspective would still fail to account for consciousness. Over and above the perspectival facts, whether such a perspective is accompanied by a given first-person subjective existence remains a further fact. And so, an exclusively perspectivalist analysis is false with regard to consciousness, because it fails to account for first-person subjectivity as a distinct mode of existence. The distinction between first-person subjectivity and third-person objectivity is not merely perspectival, but is a genuine ontological distinction, insofar as third-person objectivity is nonexperiential and first-person subjectivity is experiential.

As noted earlier, consciousness is not comprised of qualities, but is the pure first-person existence wherein such qualities manifest. Insofar as consciousness is just a pure existence, it is true that consciousness exists as a mereologically simple unit. The range of qualia experienced in that first-person existence may vary from moment to moment, but the very presence of that individuated first-person existence is an all-or-none issue. Likewise, the different neurophysiological states of an organism may be associated with different levels of wakefulness and different capacities for awareness, but to talk of "levels of consciousness" is to commit a category mistake, because the issue of whether or not the first-person existence of consciousness is present is an all-or-none issue. Thus, the suggestion that the presence of consciousness is a matter of degree is false. In virtue of its being mereological simple, it true that the presence of consciousness is an all-or-none phenomenon.

It is the first-person ontology of consciousness that makes it so unique and intangible to our conventional methods of definition, which rely heavily on our ability to operationalise the feature being defined in the third-person. We can fairly straightforwardly define tables and chairs, because we can refer to them in the third-person as objects and describe them in terms of third-person criteria. These criteria may be functional such as when one defines a chair as an item of furniture designed for sitting, they may be reductive such as when one defines water in terms of H_2O molecules, they may mereological such as when one defines protons and neutrons as constituents of atomic nuclei, and so on. Of course, these are not exclusive definitions and the criteria are neither necessary nor

sufficient, but they nonetheless help to describe the objects in terms of features that are accessible in the third-person. However, given its first-person ontology, consciousness eludes this approach to definition. It cannot be characterised in the third-person as an object, because it is essentially subjective. Hence, first-person acquaintance with consciousness is required in order to understand consciousness.

It is plausible that the intangibility of consciousness to our conventional methods of definition has contributed to its conflation with various other features. However, given that consciousness is fundamentally subjective, any attempt to objectify or operationalise it will ultimately fail to characterise it. As I discussed earlier, several accounts falsely characterise consciousness as something it is not. The psychological and neurophysiological features with which it is commonly conflated are objective, and so they can be characterised adequately by our conventional methods of definition. Due to this accessibility, they appear to provide convenient descriptions of "consciousness" but, as I have noted, these definitions are inaccurate and fail to acknowledge consciousness for what it actually is.

What I have discussed in this section illustrates the uniqueness of consciousness. Many of the features with which we are familiar have a third-person ontology, but consciousness has a first-person ontology. Consciousness is not like anything in the objective world, but is a *sui generis* phenomenon of its own kind. Accordingly, the definition of consciousness which I presented earlier is truly an essentialist definition, as it denotes the first-person subjectivity that is the essence of consciousness. Moreover, first-person acquaintance with consciousness is required to grasp this definition.

The existence of consciousness

The subjectivity of consciousness presents us with a unique epistemic situation. Subjective experience cannot be demonstrated empirically, yet we still know that it exists. Indeed, the existence of consciousness is not something that can be confirmed in the same way as, for example, confirming that of a snow crystal. A snow crystal is an object with a third-person ontology, and so can be accessed empirically. Consciousness, however, has a first-person ontology, and so cannot be accessed in this way. I can experience your behaviour, but I cannot experience your consciousness. You can experience my behaviour, but you cannot experience my consciousness. Despite all this, the existence of my consciousness is

as certain to me just as the existence of your consciousness is certain to you and I cannot doubt the existence of your consciousness just as you cannot doubt the existence of my consciousness.

In virtue of its first-person ontology, the existence of consciousness is necessarily proven to me by the very fact that I *am*. My consciousness is my first-person subjective existence, and so it is necessary to me that it exists. Thus, realism about consciousness is necessarily true, because consciousness is the very existence through which reality is discerned. I know that consciousness exists by being a conscious subject.

It could even be claimed that I am more certain about the existence of consciousness than about the existence of anything else. This recalls René Descartes' *Meditations on First Philosophy* (1641), where he argues that one can doubt the reality of the external world based on the possibility that it may be no more than a fiction of one's mind or an illusion imposed onto one's mind by a deceiving daemon, but one cannot possibly doubt one's own existence as a thinking being, because the fact that one is doubting entails that one exists. Hence, I have direct knowledge of my experience, but the external world is only known to me indirectly through the subjective experience of it in my consciousness. According, the nature of the external world is open to speculation and doubt. For example, it is conceivable that it may be no more than a fabrication of my mind, the imagination of a deceiving daemon, or the work of a villainous scientist who is stimulating various parts of my brain. However, while I can doubt the reality of the external world and consider it to be a mere appearance, I cannot doubt the existence of my consciousness, because my consciousness is the first-person existence that is necessary to experience this appearance. Therefore, it is true that consciousness exists.

The above indicates that the causal criterion for reality, which claims that for something to be real is for it to have causal efficacy, is false with regard to consciousness. I know that consciousness is real through my first-person acquaintance with it, independent of any causal efficacy. Indeed, the reality of consciousness is ontologically more fundamental than the reality of anything with causal efficacy, because such causal efficacy is only discerned through consciousness. Thus, the causal criterion for reality is false, because the very possibility of causal efficacy presupposes the prior existence of consciousness wherein such causal efficacy can manifest.

There has been some attempt to deny that there is anything more to experience over and above individual qualities, but I argue that

this is mistaken. Notably, David Hume in *A Treatise of Human Nature* (1740), argued that one can experience a bundle of perceptions, but one cannot experience a substance that one can define as one's "self". From this, he concludes that all there is to the mind is this bundle of perceptions. However, bundle theory fails to account for this first-person individuation of experience. The fact that qualities are manifesting at all necessitates an existence wherein they manifest. Moreover, these qualities do not manifest in some third-person objective space but present to a first-person subjective viewpoint, and so this existence is subjective. As noted earlier, over and above the facts about the qualities that comprise the bundle, the fact that the bundle is experienced by *my* consciousness and not by *your* consciousness remains a further fact. Therefore, bundle theory is false. Beyond the bundle of perceptions, there exists a separate consciousness wherein this bundle manifests. Indeed, as Immanuel Kant recognised in *A Critique of Pure Reason* (1781), the existence of consciousness is a transcendental condition of possibility for experience. Consciousness is not experienced as an object, but this is because consciousness is the subject that is doing the experiencing.

There has also been some attempt, as seen in the work of Dennett (1991), to claim that consciousness comprises an illusion, but I argue that this is mistaken and even incoherent. Consciousness is impossible to negate, because the very existence of consciousness is necessary for the discernment of what is real and what is illusory. An illusion is itself an experiential state, and so it presupposes the existence of consciousness wherein it can manifest. Therefore, this sort of illusionist eliminativism is necessarily false.

This indicates that the existence of consciousness is a necessary truth. Indeed, the suggestion that consciousness does not exist is necessarily false, because the existence of consciousness is necessary for the discernment of what exists and what does not, insofar as such discernment is only done through consciousness. Its nonexistence is impossible, because it would preclude the very discernment of what does and does not exist, which would negate the very possibility of nonexistence. It is only because consciousness exists that the discernment of existence is possible. Thus, ontological nihilism is false regarding consciousness. Every possibility regarding what exists and what does not presupposes the prior existence of consciousness as a necessary condition. This can also be expressed analytically through the fact that existence necessarily exists, for existence is *what is*. My consciousness is my first-person existence, and so my consciousness necessarily exists to me.

The above highlights the inescapability of subjectivity. Consciousness is my first-person existence, and so it is the necessary foundation for my access to reality. Existence is known from the first-person viewpoint of consciousness, and so what is known to be real is grounded in what manifests in consciousness as experience. Indeed, the idea of a purely objective view of reality is untenable. Consciousness is the first-person existence through which reality is known. The suggestion that one's experiential viewpoint could "step out of" its first-person mode into a third-person mode is false and even incoherent, because an experiential viewpoint is essentially first-person by definition. Such exclusion of the first-person would involve, as Nagel (1986) suggests, a "view from nowhere", which is not like anything. A view is necessarily someone's view. Hence, when one conceives of any sort of reality, one is necessarily doing so from the first-person viewpoint that equates to one's existence. One cannot possibly step out of one's own existence.

Importantly, this does not imply skepticism about the external world. Although we only access our experiences, the objective world is what brings about these experiences in our consciousnesses. Accordingly, it is reasonable to infer that it subsists on its own. The objective world is what is experienced, qualia are the qualities as which it is experienced, and consciousness is the experiencer of these qualities. We can, therefore, say that idealistic monism is false, because it fails to account for what occasions the experiences that manifest in our consciousnesses. Without the mental, the physical has no reality, and without the physical, the mental has no content.

Nonetheless, while it does subsist on its own, the above suggests that the objective world has no quality on its own. This somewhat recalls Kant's (1781) proposal that we do not access objects as they are in themselves, but only the appearances of these objects in our minds. That is to say, knowledge of the world outside experience is not real, but ideal. However, I go further than this and make not just an epistemic claim, but an ontological claim. As noted earlier, any conception of reality is necessarily from a first-person viewpoint. And so, the objective world is only made manifest when realised as subjective experience in the first-person existence of consciousness. Through experience, it acquires reality, or the quality of being like something experientially. Hence, without experience, the objective world is not like anything. It has no reality on its own, because it has no first-person existence wherein it can manifest. Rather, the objective world subsists as a mere potential which is only realised and given quality through experience by consciousness.

II
The Self

In spite of the increasing scientific understanding of the world around us, there remains the enduring philosophical mystery of the nature of the subjective self that experiences this world. What is this "I" that I am? There have been many attempts to answer to this deceptively simple question that appeal to aspects of our psychology and neurophysiology, but I argue that any such objective account will fall short of its goal. The reason is that the self is essentially subjective. It has a first-person ontology, and so any attempt to objectify it in terms of third-person properties will be unsuccessful.

Despite being unable to account for the self, objective analyses can provide insight into the psychological feature of self-awareness. This refers to one's capacity to think about oneself as an individual who is distinct from others. However, this is different from providing an account of the self, which would involve providing an account of the haecceity, or the essence, of the "I" that I *am*. My proposal, in this chapter, is that in order to truly understand the self, we need to appeal to the phenomenon of consciousness.

It is the irreducible subjectivity of the self that makes it so inaccessible to objective analysis. Our interactions with each other are limited to features that are objectively accessible, such as our bodies and our behaviour. I can access your body and your behaviour, but I cannot access your subjectivity in the manner in which I can access mine. Conversely, you can access my body and my behaviour, but you cannot access my subjectivity in the manner in which you can access yours. Hence, we are inclined to characterise others through descriptions of and judgements about the outward features that we experience, such as their appearances and their behavioural dispositions.

Many attempted accounts of selfhood and personal identity appeal to the aforementioned outward features. For example, the bodily theory suggests that personal identity is constituted by bodily identity. The activity pertaining to a person is centred in a limited region of space, namely the person's body. Not only does this body have a clear boundary, but the what occurs within this boundary appears to be correlated with the person's sense of agency. Therefore, a person's body gives us a useful referent for the personal identity of that person. We consider everything that is within it as

part of that person, and everything that is external to it not, so that the person's hands, for example, are considered to belong to the person, whereas the air surrounding the person is not. Furthermore, differences in the bodily appearances of persons, notably their facial appearances, help us to distinguish one from another.

However, the bodily theory of personal identity is problematic if one assumes a reductive definition of what makes a later body the "same body" as an earlier body. If we say that a later body and an earlier body are the same if they contain exactly the same bits of matter, then the conservation of one's body fails to account for one's personal identity. As noted by Richard Swinburne (1984), the body is continually changing. Matter is exchanged through the processes of nutrition, metabolism, and excretion, and old cells are continually replaced by new cells. Some neurones may persist for several years, but on a smaller scale the atoms and molecules are continually being exchanged and replaced. It follows that a reductive definition of bodily identity fails to account for personal identity, because one's body is always materially different at any two moments in time.

A more reasonable alternative to the reductive definition of bodily identity given above is to consider the overall organisation of the body. Although bodily matter is continually being replaced, this replacement is gradual and does not dramatically alter the body's overall organisational structure. Its macroscopic anatomy and its physiological processes are largely maintained despite the gradual changes in matter. Under such a holistic view, a later body may be considered the same body as an earlier body if it shares the same overall organisation. If we assume this holistic definition of bodily identity, then the bodily theory of personal identity becomes more palatable, because it accommodates the fact that the body is continually changing.

However, the holistic version of the bodily theory of personal identity is also problematic. First, a single body is not always associated with a single person. Notably, a pair of cojoined twins are two distinct people who share a single body. Second, although this version of the bodily theory of personal identity can account for gradual changes that do not alter the overall organisation of the body, cases that involve quick changes which alter the overall organisation of the body are harder to accommodate. Also, our judgements about personal identity may depend on which part of the body is being altered. For example, if one receives a kidney transplant, then one's overall anatomy and physiology would change, but one would still be the same self. Also, if a transgender woman has hormone

treatment and gender reassignment surgery to change her biological characteristics to accord with those that are socially associated with her authentic female gender, then this would amount to a significant change in her overall bodily organisation, but she would still be the same self. However, if a person's brain is transplanted into another body, we might then identify the person with the new body, rather than with the old body. We might say that the person has had a body transplant, rather than saying that the person has had a brain retrieval.

What the above suggests is that the overall organisation of the body, as postulated by the bodily theory, is insufficient to account for personal identity. Perhaps the identity of the brain may be more relevant. Under the brain theory of personal identity, a person is to be considered the same person as an earlier person if that person's brain is the same as that of the earlier person. Again, this does not mean that for a brain to be the same as an earlier brain it must be constituted from exactly the same matter. Like the rest of the body, the matter in the brain is continually changing, and so one must consider its overall organisation when referring to its identity.

The brain theory of personal identity may seem appealing, insofar as the brain is the bodily structure which is involved in the production of a person's behaviour. As noted earlier, our interactions with others rely heavily on their behaviours, and so it is not unreasonable to identify them with the structures that produce these behaviours, namely their brains. However, the following thought experiments show that the brain theory of personal identity is problematic. In the first thought experiment, a villainous neurosurgeon removes the cerebral cortex from body *A*, transplants the left hemisphere into body *B*, and transplants the right hemisphere into body *C*. Which of the resulting persons, if any, is to be considered the "same person" as the person associated with body *A* before the operation? In the second thought experiment, the villainous neurosurgeon removes the left hemisphere from body *D*, removes the right hemisphere from body *E*, and transplants the resulting structure onto the brainstem of a body *F*. With which of the two persons previously associated with body *D* and *E*, if any, is the resulting person associated with body *F* to be considered the "same person"? While these scenarios are not naturally possible, they do present conceptual problems for the brain theory of personal identity.

An attempt at an answer is attempted by Derek Parfit (1971), who suggests that in the first thought experiment, each of the resulting personalities associated with bodies *B* and *C* cannot be considered

the same person as the original person previously associated with body *A*, but both are constituted by parts of the original person associated with body *A*. In the second thought experiment, the resulting person associated with body *F* cannot be considered the same person as either of the original persons previously associated with bodies *D* and *E*, but is constituted by both of them. And so, according to Parfit's "complex view", personal identity, although determined by the identity of the brain, is suggested to be a matter of degree. A person is the same person as an earlier person only to the same degree that their brains are the same.

Indeed, we may not need to appeal to such hypothetical thought experiments to find situations to which Parfit's ideas could possibly be applied. Historically, patients with severe epilepsy were treated by surgical removal of the nerve bundle that usually connects the cerebral hemispheres, or the *corpus callosum*. The experiments by Roger Sperry (1969) suggested that although these patients could behave in coordinated ways after the procedure, their cerebral hemispheres could work independently when given specific tasks. If we analyse the personal identity of a person who has undergone a corpus callosotomy under Parfit's theory, then each cerebral hemisphere, insofar as it is capable of some degree of independent activity, can be suggested to correspond to a different "personality" and, although neither hemisphere can be considered the same "personality" as the original person before the corpus callosotomy, both hemispheres together belong to the original person.

As noted earlier, the brain theory of personal identity seems appealing, because the brain has a central role in controlling how an individual thinks, speaks, and acts. That is to say, the brain is associated with enabling the features that comprise one's personality. If this is the reason why we the brain theory of personal identity seems appealing, then this suggests that it is one's personality that is relevant to one's personal identity. Indeed, our interactions with an individual depend heavily on the individual's behaviour. Given that the individual's personality, or the enduring pattern of the individual's emotional, psychological, and dispositional traits, heavily influences the individual's behaviour, we are inclined think about the individual's personal identity in terms of the individual's personality. Furthermore, due to the complex array of social and cultural factors that causally contribute to personality development, the personalities of different individuals are so complex and diverse that each individual's personality is effectively unique. Hence, we tend to associate particular individuals with particular personalities.

The personality theory of personal identity differs from the bodily theory and brain theory insofar as it does not explicitly associate personal identity with a physical feature of the body, but rather associates it with a psychological feature that pertains to one's behaviour. Historically, this psychological feature was sometimes thought to be the attribute of an immaterial substance, commonly referred to as the soul. Among those associated with this view was René Descartes (1641), who suggested that one's personal identity is determined by one's soul, or "*res cogitans*", which was characterised as being composed of one's consciousness in conjunction with one's capacity for thought.

Another psychological theory of personal identity is John Locke's (1689) theory based on memory. This is loosely linked to the personality theory of personal identity, insofar one's memories of autobiographical events clearly influence how one's personality develops. According to Locke, an individual can be considered to be the same personality as an earlier personality if the individual remembers having done certain things which were, in fact, done by the earlier personality. This suggests that one's personal identity is secured by one's memories of one's past autobiographical events. At initial glance, this appears to be a reasonable suggestion. After all, my memories of my own childhood are a great part of what lead me to believe that I am the same person as the child in my mother's photograph album. As Parfit suggests, one's memories provide psychological continuity that links different periods in one's life.

However, the memory theory of personal identity is also problematic. Consider the following thought experiment. A villainous psychologist, through some hitherto undiscovered method, erases a person's memory and replaces it with a copy of a different person's memory. After the procedure, the two individuals would have exactly the same memories as each other. Nonetheless, despite this, the two people would not be the same person, but would still be two different people. In fact, such a hypothetical thought experiment is not needed to make the point. First, consider the example of retrograde amnesia. If a person suffers from retrograde amnesia following a traumatic injury, then we would still consider that person to be the same person as the person before the accident. Second, consider the example of confabulation due to dementia. If a person develops confabulatory memories, then that person would still be considered the same person as the person before those confabulatory memories were acquired. Therefore, memory is not an adequate criterion for the definition of personal identity.

A skeptical approach to personal identity is suggested by David Hume (1740). According to Hume, one's sense of personal identity is fictitious. He argues that upon introspection, one will notice that one's mental state is no more than a bundle of different perceptions. Furthermore, for Hume, memory does not provide identity between a present perception and a past perception, but rather reflects a causal link between them. That is to say, a present perception is not the same as a past perception, but is merely caused by it. Hume's bundle theory, therefore, suggests that there is no stable feature beyond the bundle of perceptions that can ground identity.

There is much in common between Hume's bundle theory of personal identity and the more recent theory developed by Daniel Dennett (1991). Like Hume, Dennett rejects the idea that one's personal identity is a substantial feature and instead suggests that it is nothing but a useful fiction that emerges from the collection of mental events. According to Dennett, one's collective mental events, which include one's perceptions, memories, beliefs, and ideas, organise themselves in such a way that spins out a fictitious "centre of narrative gravity". This, Dennett suggests, is not something concrete that can be identified in the same way that one's brain can, but is a fiction produced by one's psychological activity.

So far, I have presented an overview of some attempted theories of personal identity. The very fact that we are able to identify ourselves and distinguish ourselves from others by appealing to various material and psychological features suggests that we possess self-awareness. This refers to one's ability to think about oneself as a distinct agent. As David Chalmers (1996) suggests, self-awareness might be explained by a system's having access to some sort of representation of itself. This representation could contain information about the system's internal states, which would enable the system to engage in introspection. The representation could also provide the system with some sort of model of its own structure, which would allow it to distinguish itself from other systems. As noted in chapter one, the notion of self-awareness is commonly conflated with the notion of consciousness. Several attempted models of "consciousness", such as those by David Armstrong (1968), Douglas Hofstadter (1979), and Daniel Dennett (1991), are not accounts of consciousness at all, but are accounts of self-awareness. Relatedly, insofar as the theories of personal identity discussed above only appeal to the physical and psychological features that are cognised through self-awareness, I argue that these theories of personal identity ultimately fail to account for the self.

The reason why the above theories of personal identity are false with respect to the self is that they fail to capture the subjectivity that is essential to selfhood. It is a necessarily true, by definition, that my self is what I *am*. This is essentially a fact about my first-person identity. Accordingly, it is necessarily true that the self has a first-person ontology. In view of its first-person ontology, the self cannot be characterised by the theories discussed above, because these theories are based on features that have a third-person ontology. Some of the features are material, such as the body and the brain, while other features are psychological, such as one's memories and perceptions. Nonetheless, these features are objective, and so cannot possibly constitute the self. As noted above, the self is essentially first-person. The suggestion that the self could have a third-person ontology is necessarily false, because my self is what I am, which is essentially a fact about my first-person existence.

As Immanuel Kant noted in *A Critique of Pure Reason* (1781), the existence of subjectivity is a transcendental condition of possibility for experience. An experience does not occur in some neutral third-person space, but is individuated to a given first-person experiencer. Hence, Hume's (1740) skepticism about personal identity is false, because it fails to account for this first-person individuation of experience. For example, over and above the facts about the qualities that comprise a bundle or perceptions, the fact that the bundle is experienced by *my* consciousness and not by *your* consciousness remains a further fact. Moreover, the reason why the self cannot be perceived as an object of perception is that the self is the subject that is doing the perceiving. The self is the first-person subjective existence that experiences the bundle of perceptions.

This idea that third-person features cannot comprise the self brings to mind the doctrine of *anattā* in Buddhism, which suggests that one's bodily features and mental contents are impermanent, fluctuating, and dependent on the conditions of their arising, and so cannot constitute one's unconditioned self. As noted by Ṭhānissaro Bhikkhu (1993), this doctrine should not be interpreted as the metaphysical denial of the self, but instead should be interpreted as the pragmatic claim that one's attachment to conditioned features that are mistakenly regarded as self is unhelpful. The metaphysical reading of *anattā* is false, because it fails to account for the first-person individuation that is a *condicio sine qua non* for experience. Indeed, this is acknowledged in Sāṃkhya philosophy and Jaina philosophy, which note that experiences are individuated to distinct subjective existences. Even when a person has attained the

unconditioned state of *nibbāna*, the concept of selfhood as first-person individuation is necessary to acknowledge that there are other subjects who, in virtue of their individuated subjectivities, are experientially distinct from that person and who may or may not have attained such a state. By contrast, a pragmatic reading is fully compatible with the metaphysical acceptance of the self in the current discussion. I know for certain that my self exists, for it is what I *am*. However, my self is something distinct from the structural and dynamical constituents that make up my body, my personality, my memories, and my perceptions. My self is subjective, so it cannot be identified with objective features, such as my body and my brain. Furthermore, my self has a first-person ontology, and so it cannot be identified with the bundle of memories or perceptions that I experience. Rather, my self is the first-person experiencer of this bundle of memories and perceptions.

Therefore, a true account of personal identity acknowledges that my self is my individuated first-person subjective existence. Under this account, it is true that consciousness is the self. This definition of the self as consciousness is a definition that is often accepted in philosophy, such as in the phenomenological work of Edmund Husserl (1921–1928) and Dan Zahavi (1999), as well as in the tradition of Sāṃkhya philosophy. I may be associated with my body and my brain, but I cannot be identified with them. I have a body, a brain, a personality, and memories, but I am my consciousness.

This account also proves that the self exists. Given that selfhood is the first-person individuation that is essential to consciousness, the fact that there is such first-person individuation of consciousness entails the existence of the self. Indeed, it is necessarily true that the self exists, because my self is my consciousness, which is the first-person existence that I necessarily am.

Accordingly, the claim that "I" is a nonreferring term is false. It was famously suggested by G. E. M. Anscombe (1975) that the use of "I" in the description of a mental state could be akin to the use of "it" in the description of a state of affairs such as "it is raining". However, I argue that this a false analogy, because it fails to account for the first-person individuation of experience. While "it is raining" obtains in a third-person objective space, a mental state is necessarily individuated to a given first-person subject. Hence, "I" denotes the specific first-person subjective existence wherein the mental state is experienced. That is to say, the true referent of "I" is consciousness.

From my direct acquaintance with myself, it is clear to me that the "I" that I am exists as a discretely individuated first-person unit

with a unique ipseity. It is in virtue of its unique ipseity that my self is different from the selves of others. This indicates that it is true that my personal identity is essentially determined by the unique first-person individuation of my consciousness. Each self is a distinct first-person subjective existence. And so, it is true that selves exist as ontologically discrete units which are essentially separate from one another in virtue of their unique ipseities.

Such individuation is an essential feature of the first-person ontology of consciousness. It is what accounts for the givenness of an experience to a specific subjective viewpoint. Accordingly, it is false to suppose that such first-person individuation of the self can be reduced to some third-person objective feature. As noted above, my bodily and behavioural properties continually change, but my consciousness necessarily remains individuated to me. Furthermore, the complete physical facts about *this* body and *that* body fail to account for why the experience associated with *this* body is individuated to *me* rather than to *you* and why the experience associated with *that* body is individuated to *you* rather than to *me*. Therefore, the first-person individuation of the self is an essential fact about consciousness that is separate from the third-person facts about any physical or psychological properties.

This shows that haecceitism is true with respect to consciousness. Famously, Max Black (1952) showed that Gottfried Wilhelm von Leibniz's (1686) principle of the identity of indiscernibles is false by conceiving of two indistinguishable spheres that swap their locations. Similarly, the scenario where *my* consciousness is associated with this body is different from the scenario where *your* consciousness is associated with this body. Of course, the haecceity of a sphere may seem mysterious. By contrast, the haecceity of a given consciousness is determined by something ontologically substantial, namely its first-person individuation. Thus, it is true that each consciousness is essentially unique. There is an ontologically substantial difference between *my* consciousness and *your* consciousness.

Given that the first-person individuation which essentially determines the unique identity of a self is discrete, the claim that selves could undergo fission is false. Likewise, the claim that selves could undergo fusion is false. Each self is a discrete existence that is essentially individuated from other selves by its unique first-person ipseity. Hence, it is true that selves cannot undergo fission. Likewise, it is true that selves cannot undergo fusion.

What can this tell us about the results of Sperry's experiments on people who have undergone corpus callosotomies, or about the

thought experiments involving the villainous neurosurgeon and the villainous psychologist? According to Swinburne (1984), what happens to one's subjective self in such a scenario is not entailed by the neurological and psychological facts about the scenario. At most, neurological and psychological continuity provides fallible evidence for the continuity of selfhood. Hence, what happens to the self remains underdetermined by the experimental findings.

Nonetheless, in light of the first-person individuation of consciousness, we do know that the person's self necessarily maintains its first-person identity. Each self exists as a distinct first-person unit whose identity is discretely determined by its unique ipseity. Thus, Parfit's (1971) claim that personal identity is a matter of degree is false. Given the first-person individuation of selfhood, it is true that personal identity is an all-or-none phenomenon.

The outcomes of the above experiments may depend on how the psychophysical laws that obtain between the mental and the physical happen to operate in our world. While this is speculative, the psychophysical interface between consciousness and the body may be stronger in one part of the brain over another. Suppose, for example, that the psychophysical interface with consciousness is more strongly associated with a certain part of the brainstem. In a person with an intact corpus callosum, there is a single subjective self whose experiences are correlated with the events in both cerebral hemispheres, as these are both connected to the brainstem. The above would also account for why a pair of conjoined twins who share a body are two distinct subjective selves, as there are two brainstems that form interfaces with two different consciousnesses. It also accounts for why a person with chimerism, where the body is produced by the aggregation of cells with different genotypes, is a single subjective self, as there is a single brainstem that forms an interface with a single consciousness.

In a person who has undergone a corpus callosotomy, the self's experiences may still be correlated with the events in both cerebral hemispheres, as these are both still connected to the brainstem. However, given the loss of connection between the two hemispheres, the contents of awareness may present differently from how they had previously presented. Hence, the person who has undergone a corpus callosotomy remains associated with a single subjective self, although awareness, cognition, and behaviour may be partly altered.

In the thought experiment involving the surgical excision of the two cerebral hemispheres and the consequent transplanting of each hemisphere into different bodies, the donor's subjective self may

remain associated with the original body *A*, assuming that this body still contains the donor's brainstem. However, since the cerebral hemispheres have been removed, the donor's cognitive capacities will be significantly affected. Meanwhile, after being grafted onto the brainstems of the two recipient bodies *B* and *C*, the two hemispheres may respectively become associated with the subjective selves that are respectively associated with the recipient bodies *B* and *C*. Since these recipients had each received a hemisphere from the donor, it is reasonable to assume that they would acquire some of the cognitive capacities from the donor. And so, the three bodies, namely the donor's body *A* and the two recipients' bodies *B* and *C*, continue to be associated with their three respective selves, but their psychological capacities are altered. That is to say, the self that was previously associated with body *A* remains associated with body *A*, the self that was previously associated with body *B* remains associated with body *B*, and the self that was previously associated with body *C* remains associated with body *C*.

In the thought experiment involving the removal of opposite cerebral hemispheres from two different bodies *D* and *E*, and the consequent combination of the two detached hemispheres in a new body *F*, each of the donors' subjective selves may remain associated with each of the donors' respective bodies, assuming that the bodies still contain the donors' brainstems. However, due to the loss of a hemisphere, each of the donors would likely exhibit changes in their cognitive and behavioural capacities. Meanwhile, the new recombinant brain, after the transplant, may become associated with the subjective self that is associated with the brainstem of the recipient's body *F*, and since this recipient had received a hemisphere from each donor, the recipient might perhaps acquire some of the donors' cognitive capacities. Again, the three bodies, namely the recipient's body *F* and the two donors' bodies *D* and *E*, continue to be associated with their respective subjective selves, but their psychological capacities and features are altered. That is to say, the self that was previously associated with body *D* remains associated with body *D*, the self that was previously associated with body *E* remains associated with body *E*, and the self that was previously associated with body *F* remains associated with body *F*.

The thought experiment involving the erasure by a villainous psychologist of a person's memories and their consequent replacement with another person's memories is easier to explain. There simply remain two different subjective selves, namely the self associated with the memory donor and the self associated with the

memory recipient, who happen to share similar memories after the procedure. In the example of a person with retrograde amnesia and the example of a person with confabulatory memories, we can accept that despite the disruptions in their memories, the people are still respectively associated with the same subjective selves. This may also apply to the example of multiple personality disorder. In such a case, is reasonable to suppose that it is the same subjective self who is experiencing the multiple personalities.

What I have presented herein is a subjectivist account of the self. Under this account, the self necessarily has a first-person ontology by definition. Accordingly, it is true that my consciousness is my self. While I may be contingently associated with a body, a brain, a personality, and memories, it is true that I am my consciousness.

This accounts for how I remain the same self despite the changes in my physiological and psychological properties. While my physiological and psychological properties continually change, the suggestion that my consciousness could change is necessarily false, because my consciousness is essentially individuated to me. In virtue of this first-person individuation, it is necessarily true that my consciousness remains the same. And so, the changes in my physiological and psychological properties occur over the constant first-person existence of my consciousness.

The above raises the issue of how there can be interactions between selves. I noted earlier that our interactions with others rely on their objective aspects, such as their bodies and behaviours, and that the subjectivities of others are experientially inaccessible to us. However, this does not mean that our subjectivities are not involved in our interactions with one another. In chapter six, I provide a more detailed account of why our subjectivities are central to our interactions. The subjective experiences of others may be inaccessible to us, but we can and do correctly acknowledge others as subjects who have experiences. As noted by Jean-Paul Sartre (1943), “I experience the inapprehensible subjectivity of the other directly and with my being”. Therefore, although our interactions with one another largely involve our objective aspects, we are intimately aware of one another as subjective experiencers. Moreover, such intersubjective appreciation could be considered foundational to an egalitarian moral philosophy, insofar as it suggests that we are all coequals as subjective experiencers, regardless of the contingent differences between us. We fundamentally acknowledge one another as selves.

III
Falsifying Physicalism

Science is the empirical study of the objective world that is experienced by us. More specifically, it is an empirical study of the physical features of this world, or its structure and dynamics. Among the major activities of science is the construction of theories to describe and explain the structural and dynamical features that we observe in the world. These theories are also valued on how well they predict further features and how well they inform practical interventions. Science is sometimes assumed to amount to physicalism, but this is inaccurate. While the former is the empirical study of the features of the world that are physical, the latter is the claim that everything is physical. The former does not entail the latter, for it is entirely possible to accept everything that science reveals about the physical features of the world, while accepting that there is something over and beyond these physical features that cannot be grasped with the empirical methodology of science.

Some, though not all, scientific explanations are reductive. Since all physical features are based on the parameters of structure and dynamics, they can, to certain extents, be related to one another and explained with reference to one another. Furthermore, it is sometimes assumed that there is a hierarchy of explanation. Some higher-level physical facts may, in principle, be explained by lower-level physical facts, since the higher-level structures and dynamics are constituted by or emerge from the interplay between lower-level structures and dynamics. Hence, sound might be explained in terms of pressure waves, which in turn can be explained in terms of the movements of air particles. Accordingly, some higher-level physical facts can be derived from the lower-level physical facts, such as the property of diffusion being predicted from the mass movement of molecules in a medium. This is because the lower-level physical facts entail the higher-level physical facts, and so the higher-level physical facts are logically supervenient on the lower-level physical facts. Of course, such reduction may not be feasible across some domains. This would indicate need for a pluralistic approach to science, whereby different domains require different methods and concepts. Nonetheless, while a pluralistic approach may ultimately be required, it is plausible that at least some higher-level physical facts can be reductively explained by lower-level physical facts.

In much of this chapter, I shall be concerned with the epistemic claim that a physical scientific account of consciousness is impossible. This will be followed up in chapter four with the related ontological claim that the reason why a physical account of consciousness is impossible is because consciousness is not physical. I argue that given a complete structural and dynamical physical account, the existence of consciousness will always be an extra fact to consider. It is an extra fact which cannot be entailed by the low-level physical facts, and so it is not logically supervenient on them. Rather, it is an entirely separate issue to consider, over and above the physical facts. As David Chalmers (1996) notes, structural and dynamical facts yield only further structural and dynamical facts. They cannot encapsulate the subjective quality of experience.

It follows that consciousness is fundamentally beyond science. If science is the study of the structure and dynamics of the world we experience, and consciousness is not conditioned by structure and dynamics, then science cannot account for consciousness. Indeed, the scope of scientific enquiry is the objective world which is experienced by us, but consciousness is not to be found within this objective world, for it is the subjective experiencer of this world. Since science is the study of the objective, then it follows that it cannot account for what is fundamentally subjective. After all, any study of the objective depends upon our ability to refer to the feature being studied in the third-person as other, as one does through experimentation and observation. However, I argue that this objectification is impossible with consciousness, because consciousness is irreducibly first-person. It cannot be accessed as an object of experience, because it is the subject of experience.

Beyond science

This section illustrates the intangibility of consciousness to science by showing how attempts to arrive at a scientific account of consciousness in various disciplines have been unsatisfactory. To be clear, my aim is not to show that these scientific approaches have not been valuable. I think it is clear that these scientific approaches have provided a great deal of valuable insight into various psychological aspects of the mind. Rather, my aim is to show that despite their successes with explaining these various psychological aspects of the mind, these approaches ultimately fail to explain the existence of consciousness.

Neurophysiology: Given that our experiences seem to be correlated with the activities of various parts of our brains, it is understandable that many have supposed that brain science can tell us about the nature of conscious experience. In recent decades, several neurophysiological models of "consciousness" have been attempted, such as those by Gerald Edelman (1989), Francis Crick and Christof Koch (1990), and Antonio Damasio (1999). I have put "consciousness" in quotation marks, because these models do not actually explain consciousness, but instead attempt to explain various psychological capacities that are erroneously conflated with consciousness. For example, Edelman's model suggests that "consciousness" can be explained by the way different brain structures interact to allow the conceptual categorisation of perceptual signals before they contribute to memory. Furthermore, he links this to the generation of language by postulating links with regions of the brain involved in speech production. It seems that what Edelman calls "consciousness" is not consciousness at all, but a form of higher-level perception. While his model explains this psychological capacity, it fails to account for the the subjective quality of experience. Similarly, the model suggested by Crick and Koch explains how different modalities of perceptual information are bound and unified by 40-hertz oscillations in the visual cortex. Again, this model appears to be a structural and dynamical account of higher-level perception, but it fails to account for the subjective quality of experience. According to Damasio (1999), what is central to "consciousness" is the notion of homeostasis. This refers to the coordinated and regulatory activities that enable a biological system to maintain an organised steady state. From a thermodynamic perspective, homeostasis could be analysed as a process that locally minimises entropy. However, while a model of homeostatic entropy minimisation may explain how a biological system resists degradation, it fails to explain why such a biological system is accompanied by subjective experience. Given the physiological and mathematical facts about homeostasis and entropy minimisation, the existence of consciousness remains a further fact. Accordingly, I argue that the claim that consciousness can be explained by neurophysiology is false. Any neurophysiological account of the structure and dynamics of a cognitive system can only yield further structural and dynamical information, but cannot yield information about the subjective quality of experience. Even if one discovers the neural mechanism that is correlated with a certain subjective experience, then one still would not be explaining consciousness.

Rather, one would be presupposing subjective experience and showing that it tends to accompany a certain neural mechanism, but this correlation and the very existence of subjective experience itself would remain unexplained. And so, the discovery of a neural mechanism would fail to account for consciousness, because the presence of consciousness would remain a further fact over and above the neural mechanism.

Psychophysics: This experimental discipline is concerned with the capacities and limits of an organism's sensory and perceptual systems. It studies what an organism can and cannot perceive, as well as the relations between changes in the stimuli and the organism's responses. A notable exponent of psychophysics was Hermann von Helmholtz, who, in his *Handbuch der Physiologischen Optik* (1867), investigated the thresholds of the visual system and formulated theories about colour vision and perception based on the activities of different receptors. Psychophysics is sometimes suggested to be the study of the limits of "experience", but I argue that this is mistaken. Psychophysics is a study of the dynamics involved in the psychological processes of awareness and perception. Indeed, subjective experience may be correlated with awareness and perception, but it is nonetheless a further fact over and above these psychological processes. Therefore, while psychophysics may examine the informational contents of awareness and perception, it does not explain the existence of consciousness.

Introspectionism: This approach to psychology, associated with William James (1890), suggests that one can learn about one's cognitive processes from the facts that one reveals in one's introspective reports. This is often considered to be a study of "consciousness", but I argue that this is mistaken. Insofar as it is based on verbal reports, introspectionism presupposes introspection and reportability, rather than consciousness specifically. That is to say, it works with the informational content of one's awareness. Of course, the informational content of awareness may be correlated with a subjective quality of experience, but this correlation is only assumed and not explained. Consciousness remains an extra fact to consider over and above the informational content of awareness.

Psychoanalysis: Psychoanalytic theory does not specifically attempt to explain consciousness, but it does posit the partitioning of the mind into "unconscious" and "conscious" elements. This terminology may be somewhat misleading, for it could result in the conflation of consciousness with psychological aspects of the mind. In the terminology I am taking to be true for the purpose of this

book, conscious means that the system is associated with consciousness, whereas nonconscious means that the system is not associated with consciousness. In psychoanalytic theory, however, the "conscious" is the aspect of the mind whose contents one can access through awareness and can process deliberatively, whereas "unconscious" mental contents are not immediately accessible to awareness and their processing occurs independently of any deliberative effort on the part of the individual. From this, it is clear that the psychoanalytic concepts of the "conscious" and the "unconscious" describe psychological aspects of the mind, which are involved in the generation of behaviour. They do not pertain to consciousness itself. In fact, the distinction between the "conscious" and the "unconscious" is not consciousness, but self-awareness and related psychological properties, such as introspection and reportability. While these are present in the "conscious" aspect of the mind, they are lacking in the "unconscious". The "conscious" aspect of the mind is that of which one is aware and over which one has deliberative control, whereas one has no such awareness or deliberative control over the "unconscious". These features are independent of the subjective quality of experience. Nonetheless, it might be noted that conscious experiences are correlated with events in the "conscious" aspect of the mind, but this correlation is a further fact that is beyond the scope of psychoanalytic theory itself. Indeed, "conscious" mental contents are, indeed, associated with particular qualia, whereas "unconscious" mental contents tend to be inferred or experienced through their manifestations in behaviour. However, psychoanalysis cannot account for this correlation, nor does it try to account for it. Rather, it is concerned with the dynamics of the mind, and how these relate to the motivation of behaviour and the psychopathology of neurosis. This emphasis on the structural and dynamical aspect of the mind can be exemplified in Sigmund Freud's model of the psychic apparatus in *The Interpretation of Dreams* (1900). In his model, Freud places the "conscious" and the "unconscious" as compartments within a reflex arc, and explains wish fulfilment and dream formation in terms of the dynamics within this reflex arc. Consciousness itself is not explained. Moreover, even if there turns out to be a structural and dynamical difference between "conscious" and "unconscious" processes, such as a neural difference between ordinary visual perception and blindsight, then this would still fail to explain consciousness, because the presence of consciousness would still need to be acknowledged as a further fact that happens to be correlated with this difference.

Cognitive Psychology: This is the discipline which studies the dynamics that are involved in cognitive processes. Explanations in cognitive psychology often take the form of cognitive models which trace the causal dynamics of information flow between different modules. Its scope is broad, examining a variety of psychological capacities, such as perception, memory, attention, introspection, reportability, self-awareness, intentionality, language, and problem solving, amongst others. However, I argue that cognitive psychology is insufficient to account for consciousness. Again, it provides structural and dynamical explanations of psychological capacities, but it cannot account for the subjective quality of experience. Over and above these structure and dynamics facts, the existence of consciousness still remains an extra fact to consider. Nonetheless, cognitive models of "consciousness" have been attempted, such as those by Bernard Baars (1988), Daniel Dennett (1991), and Paul Churchland (1995), but these turn out to be sophisticated accounts of various psychological capacities which are incorrectly labelled "consciousness". While these psychological capacities may be explained, consciousness itself remains unexplained. For example, Dennett's model (1991) suggests that information processing involves multiple drafts that "compete" with one another for attention. This is effectively an account of the dynamics involved in attention and reportability, but it fails to account for consciousness. Given all the details about the causal dynamics of the proposed system, there is still no reason why such a system should be accompanied by consciousness. Furthermore, as noted by Chalmers (1996), even if such a system could, with the use of a hypothetical "experience meter", be shown to be accompanied by consciousness, then this would, at most, show that there is a correlation between a certain sort of system and subjective experience. This does not explain consciousness, but presupposes its existence. Hence, the suggestion that subjective experience could be explained by some psychological capacity, such as memory or introspection, is false, because the presence of consciousness is a further fact over and above the structure and dynamics of such a psychological capacity.

Computer Science: Given that cognitive models work at the relatively abstract level of information flow between modules, their realisations are, in principle, substrate neutral. Any system which successfully realises the causal dynamics proposed by, for example, a model of learning will be learning, regardless of what materials make up the system. Hence, it is theoretically possible for cognitive processes to be performed by computational machines. This is the

focus of the field of artificial intelligence, as pioneered by Alan Turing (1950). A computation is an abstract model that provides information about the organisation of a physical system in mathematical terms. Computations are said to be implemented on physical systems, such as machines based on silicon integrated circuits. However, given the substrate neutrality of these computations, other systems, such as those in the brain, might potentially also be described in computational terms. Indeed, the field of computational neuroscience has used mathematical algorithms and computer simulations to model neural networks and cognitive processes. However, I argue that consciousness cannot be modelled by a computation. A computation is effectively a mathematical abstraction describing certain structural and dynamical features of a system. While it can supply information about structure and dynamics, there is nothing in such a mathematical abstraction that encapsulates the subjective quality of experience. That is to say, over and above all the information provided by a mathematical abstraction, the existence of consciousness remains an extra fact to consider. Thus, the claim that consciousness can be mathematically explained or modelled is false. Given its irreducible subjectivity, consciousness is outside the scope of mathematical modelling. Nonetheless, computational accounts of "consciousness" have been attempted, although I argue that these do not actually account for consciousness, but provide explanations of the processes associated with various cognitive capacities that are erroneously conflated with consciousness. For example, Douglas Hofstadter (1979) suggests that "consciousness" involves a recursive "strange loop", wherein a system feeds information back to itself, but I argue that this is insufficient. His model offers a structural and dynamical account of how a system is capable of recursion, but the existence of consciousness itself remains unexplained. In fact, the feature which he calls "consciousness" is not consciousness at all, but a form of self-awareness. Of course, it may turn out that a certain implementation of a computation is accompanied by consciousness, but this would only suggest that there is a correlation between a certain sort of computational system and subjective experience. The existence of consciousness itself would still need to be presupposed over and above the information provided by the computation. Hence, while it is possible that a complex artificial system could become associated with consciousness, this would not indicate that consciousness itself has been computationally modelled, but instead would indicate that the feature of the system that has been

computationally modelled happens to be contingently correlated with consciousness. Given the structural and dynamical facts about the implementation of the computation, the existence of consciousness is still a further fact to consider.

Quantum Mechanics: This a subfield of physics, associated with Niels Bohr (1922), Werner Heisenberg (1930), and Erwin Schrödinger (1944), which studies matter and energy at a subatomic level. Quantum mechanics introduced new concepts to physics, including the quantisation of energy, the intrinsic uncertainty of measurements, a wave-particle duality, and the nonlocality of events in spacetime. Accordingly, it marked a paradigm shift that replaced the determinism of classical physics with uncertainty and nonlocality. It has been suggested that quantum mechanics might reveal a physical explanation of consciousness, but I argue that this is false. Although quantum mechanics is a radical theory, it ultimately remains a theory about the structure and dynamics of the physical world. The difference between quantum mechanics and classical physics is that the structures and dynamics posited by quantum mechanics are indeterminate and nonlocal. Nonetheless, given that it is still a structural and dynamical theory, quantum mechanics cannot account for consciousness. Nothing in the structural and dynamical facts entails the subjective quality of experience, and so the existence of consciousness is still a further fact to consider over and above the physical facts. Like many of the other scientific accounts I have discussed, most quantum mechanical accounts of "consciousness" do not explain consciousness itself, but various psychological capacities which are then misleadingly conflated with consciousness. For example, Roger Penrose (1989) and Stuart Hameroff (1994) suggest that quantum collapse in brain microtubules may be the mechanism behind certain cognitive processes. While this may potentially help us to understand such processes, it does not say anything about consciousness. Other accounts, such as those by Dana Zohar (1990) and David Hodgson (1991), posit certain quantum mechanical properties as the physical correlates of conscious experiences. However, these accounts can only assume a correlation between the physical and the phenomenal. Positing that a given quantum mechanical state is associated with consciousness does not explain consciousness, but presupposes consciousness and assumes that it is correlated with the quantum mechanical state. Over and above the physical facts about the structure and dynamics of the quantum mechanical state, the existence of consciousness remains a further fact.

Evolutionary Biology: This is a subfield of biology, pioneered by Charles Darwin (1859), which studies how biological systems have been shaped by evolution. A notable feature of this world is that consciousnesses tend to be associated with such biological systems. Accordingly, it may be tempting to hope that evolutionary biology could explain why organisms became conscious, but I argue that this is mistaken. While evolution certainly pertains to the physical features of biological systems, the suggestion that evolution pertains to the first-person subjectivity of consciousness is false. It is reasonable to expect evolutionary considerations to feature in explanations of traits whose discernible behavioural effects may have influenced survival and reproduction in previous generations. For example, explanations of why organisms have certain cognitive capacities, such as attention and introspection, may appeal to evolutionary considerations and to the ontogenetic effects of the social environment. Consciousness, however, is not structural and dynamical aspect of cognition, but is the phenomenon of first-person subjectivity. It is an extra fact over and above the physical facts about an organism, and so it does not have physical or behavioural effects that could influence survival or reproduction. Since evolution by natural selection only concerns traits whose physical and behavioural effects influence survival and reproduction, evolutionary biology cannot account for the existence of consciousness. Furthermore, even if consciousness turns out to be correlated with a certain cognitive capacity, an evolutionary account may be able to provide a partial explanation of the presence of that cognitive capacity in an organism, but the existence of consciousness itself would remain unexplained. Hence, this would not be an evolutionary explanation of consciousness itself, but an evolutionary explanation of a trait with which consciousness is contingently correlated. Again, the existence of consciousness remains a further fact over and above all the facts about a biological system's evolutionary history.

What I have shown here is that science can account for the structure and dynamics of the physical world, but any such account ultimately fails to account for consciousness. A physicist may tell us that red light is a transverse electromagnetic wave of a certain wavelength, or photons at a certain energy level. There is no mention here of the subjective quality of a red experience. A biochemist may then tell us that when these photons hit one's retina, a chemical reaction occurs in one's photoreceptors. Again, there is no mention of the subjective quality of a red experience. A physiologist may tell us that the stimulation of one's retina by red light causes the

propagation of electrical impulses down the optic nerve, resulting in the activation of certain neurones in the visual cortex. Again, the subjective quality of red is left out of the account. A psychologist may perhaps tell us about the processes involved in one's ability to discriminate red and to respond to it in appropriate ways. Again, the subjective quality of a red experience is not mentioned. Given that it is the study of the structure and dynamics of the physical world that is experienced, science cannot possibly account for the existence of the subjective experiencer that is consciousness. It must be taken as true that consciousness is beyond the scope of scientific enquiry.

Physicalism falsified

In chapter four, I shall provide more general arguments against physicalism and for dualism, but in this section, I shall critically examine specific physicalist approaches and show that they fail to account for consciousness. Some of these physicalist approaches appeal to the scientific disciplines discussed earlier in this chapter. I shall argue that physicalism is false because, given any physical account of the structure and dynamics of the world, the existence of consciousness will still be an extra fact to consider.

Eliminativism: This view, anticipated by Thomas Hobbes (1655), suggests that there are no such things as mental properties or, at least, that there is no need to invoke the term "mental", since all properties can be explained in purely physical terms. Eliminativism has been supported by Willard van Ormand Quine (1960) in the philosophy of science, and more recently by Paul Churchland (1985), Patricia Churchland (1988), and Daniel Dennett (1991) in the philosophy of mind. I argue that eliminativism is false with regard to consciousness. As noted in chapter one, it is impossible for me to deny the existence of my consciousness, for my consciousness is my very first-person existence. Accordingly, the fact that I am acquainted with the existence of my consciousness refutes eliminativism. The eliminativist might then suggest that consciousness is an illusion, but I argue that this is incoherent, because an illusion is itself a conscious experience, and so it necessitates the existence of consciousness. Therefore, this sort of illusionist eliminativism is necessarily false, because the very existence of consciousness is necessary for the discernment of what is real and what is illusory. Given this, the eliminativist might assume a more modest stance, which is to acknowledge the existence

of consciousness but to then claim that the concept can be reframed exclusively in physical terms, thus making any reference to "consciousness" unnecessary. However, I argue that this also fails. Any physical account based on structural and dynamical information can yield only further structural and dynamical information, but necessarily omits information about consciousness. The existence of consciousness remains an extra first-person fact over and above the third-person structural and dynamical facts. And so eliminativism is false, because third-person physical facts cannot capture the first-person subjectivity that is essential to consciousness Furthermore, the scientific data on which such an account is based is ultimately acquired through experience by consciousness, and so such an account fails to eliminate reference to consciousness. Thus, eliminativism is necessarily false with regard to consciousness.

Identity Theory: According to identity theory, mental states are identical to neural states. Type identity theory, which suggests that types of mental states are identical to types of neural states, was developed by U. T. Place (1956) and J. J. C. Smart (1959), while a form of token identity theory, which suggests that the identities are only between individual mental events and corresponding neural states, was developed by Donald Davidson (1970). Psychophysical identity is not supposed to be an analytic fact, but a contingent and synthetic fact, established through empirical discovery. For example, according to the identity theory, the mental state, pain, and the neural state, the firing of C-fibres, might have been different features, but it just happens that they were discovered to be identical. The presumed identity between pain and the firing of C-fibres is then assumed to be analogous to the identity between water and H_2O. Identity theorists suggest that the identity between water and H_2O is also a contingent fact established through empirical discovery. They also suggest that we can conceive of a possible world wherein water did not turn out to be H_2O, just as we can conceive of a possible world wherein pain did not turn out to be the firing of C-fibres. However, Saul Kripke, in *Naming and Necessity* (1980), argues that all identities are necessary, provided that the expressions on both sides of the identity statements are rigid designators. Water is H_2O and this is the case in every possible world. The appearance of contingency is just an illusion. For example, when we imagine a possible world wherein water is not H_2O, we are, in fact, not imagining water, but "watery stuff", a substance which is superficially indistinguishable from water. However, the essential property of water is its molecular structure H_2O and, since this "watery stuff" does not have the molecular

structure H_2O, "watery stuff" cannot be water. Therefore, water is H_2O in every possible world. The identity is necessary. Also, Kripke argues that the presumed analogy of water and H_2O with pain and the firing of C-fibres is false. If pain and C-fibres are identical, then this identity too must be necessary, but according to Kripke, there is no such identity here. Whereas the expressions "water" and "H_2O" rigidly designate the same feature, the expressions "pain" and "the firing of C-fibres" rigidly designate different features. The expressions "water" and "H_2O" both refer to the same physical compound. What is essential to this compound is its molecular structure. By contrast, "the firing of C-fibres" refers to a structural and dynamical process, whereas "pain" refers to a subjective experience. What is essential to the firing of C-fibres is its neural mechanism, whereas what is essential to pain is its phenomenal quality. Hence, we can genuinely conceive of possible worlds wherein pain is not accompanied by the firing of C-fibres, and, conversely, wherein the firing of C-fibres is not accompanied by pain. Given that the relation between the firing of C-fibres and pain is contingent, pain is not identical with the firing of C-fibres. Furthermore, one cannot argue that the pain that is imagined without the firing of C-fibres is not pain, but "painy stuff", an experience which feels like pain but is not pain, because all that is essential to pain is its phenomenal quality. All that an experience has to exhibit to be pain is for it to feel like pain. This suggests that "painy stuff" is, in fact, pain. So far, the focus has been specifically on pain and the firing of C-fibres, but the argument can be made more general. Consciousness is not identical with a brain state *X*, because the expressions "consciousness" and "brain state *X*" rigidly designate features that are essentially different. What is essential to consciousness is its first-person subjectivity, whereas what is essential to brain state *X* is its third-person structure and dynamics. Given that first-person subjectivity is fundamentally different from third-person structure and dynamics, it follows that identity theory is false with regard to consciousness.

Behaviourism: This is the view that mental states can be defined analytically in terms of behaviours, such that statements about mental states are synonymous with statements about behaviours. Such a view is associated with Gilbert Ryle (1949) and, arguably, with Ludwig Wittgenstein (1953). An obvious problem with behaviourism is that it fails to account for phenomenality. Behaviourism restricts itself to descriptions of behaviours, but nothing in these descriptions entails first-person subjective

experience. Given the complete behavioural facts about an individual, the existence of consciousness still remains a further fact to consider. Therefore, behaviourism is false with respect to consciousness. With respect to intentionality, behaviourism may be more relevant, insofar as intentional states have causal roles in the production of behaviours. Hence, it could at least make theoretical sense to analyse intentional states in terms of behavioural dispositions. Nonetheless, such a behaviourist project is not entirely successful. As Roderick Chisholm (1957) argues, this is partly because behaviourist analyses are circular. They attempt to define beliefs and desires in terms of behavioural dispositions, but these behavioural dispositions are only relevant if one presupposes further beliefs or desires. For example, assume that I believe it is sunny outside. The behaviourist analyses this in terms of my behaviour. Perhaps I take a pair of sunglasses with me when I go outside. However, the fact that I take a pair of sunglasses with me when I go outside can only be taken to indicate my belief that it is sunny if the behaviourist presupposes two facts about my mentality. These are, first, that I dislike sunlight in my eyes and, second, that I believe my sunglasses will prevent me from getting sunlight in my eyes. This example shows that mental states are only analysable in terms of behaviours if we presuppose further mental states. Therefore, behaviourism does not remove the need to invoke intentionality.

Functionalism: According to functionalism, mental states are defined by the causal roles they have in the overall working of an organism. Each type of mental state is characterised in terms of a disposition to act in certain ways or have other mental states, given certain sensory inputs or preceding mental states. For example, assume that I feel pain after biting into a chilli pepper, which then makes me feel anxious, and causes me to pour a glass of water. The functionalist would say that "being in pain" is synonymous with "being in the state *P*, which is caused by biting a chilli pepper, and which in turn causes both the state *A* and the pouring of water". Therefore, like behaviourism, functionalism defines mental states in terms of dispositions. However, unlike behaviourism, functionalism posits internal mental states and is not restricted to descriptions of behavioural outputs. This is a method which has been advocated by David Armstrong (1968), Jerry Fodor (1968), and David Lewis (1990). However, I argue that functionalism is false with respect to consciousness. Functionalism is effectively an account of the structural organisation and causal dynamics of intentional processes in a cognitive system. Nothing in such a structural and dynamical

account entails the subjective quality of experience. Given a complete functionalist analysis of the structural organisation and causal dynamics of a cognitive system, the existence of consciousness remains a further fact to consider.

Embodied Cognition: In their book *The Embodied Mind* (1992), Francisco Valera, Evan Thompson, and Eleanor Rosch propose that cognition is shaped and sustained by the kind of body an organism has and the way this body is embedded in the environment with which the organism interacts. Under this view, perception does not involve an abstract representation, but is a dynamic and interactive process. While the enactive approach of embodied cognition offers a promising approach to cognition, I argue that it fails to account for consciousness. Embodied cognition provides a structural and dynamical account of how perception and action are shaped by the organism's embeddedness within the environment, but nothing in this structural and dynamical account entails the presence of first-person subjectivity. And so, while it is plausible with respect to perception and action, the enactive approach of embodied cognition is false with respect to consciousness, because the existence of first-person subjective experience remains a further fact over and above the structural and dynamical facts about embodied interactions.

Emergentism: Given the failure of reductive explanation of subjective experience, some appeal has recently been made to emergent properties in the physical sciences with the hope that they might inform a nonreductive explanation. It has been suggested that the relation between emergent properties and lower-level physical properties could be analogous to the relation between phenomenal qualities and brain states. Emergent properties are higher-level physical properties which are unpredictable from lower-level physical properties, but are nonetheless physical. For example, a colony of ants appears to follow a highly ordered pattern, but this cannot be predicted from the behaviour of an individual ant. Rather, the patterns emerge from the complex interactions between the ants in the colony. It has been suggested by John Searle (1992) that experience may be an emergent property of the brain, just as liquidity is an emergent property of H_2O. However, I argue that emergentism is false with respect to consciousness, because it rests on a false analogy. While emergent properties, such as liquidity, may not be immediately obvious or predictable from the lower-level physical properties, they are still logically supervenient on the lower-level physical properties. As Chalmers (1996) notes, this is because emergent properties are higher-level structural and dynamical

features which are constituted by the lower-level structural and dynamical features. Hence, when all the lower-level physical facts are given, these higher-level emergent properties will still be entailed, even though they may not be immediately obvious or predictable to us. However, consciousness is not a structural and dynamical feature, and so it is not logically supervenient on the lower-level physical facts. Accordingly, consciousness cannot be said to emerge from the lower-level physical facts. That is to say, when all the lower-level physical facts are given, the higher-level physical facts may be entailed, but the existence of consciousness is still an extra fact to consider. Therefore, emergentism is false with regard to consciousness. Given that consciousness is not an emergent property, it must be accepted as being ontologically fundamental.

Necessitarianism: As noted by Chalmers (1996), a potential strategy that could be used to attempt to defend physicalism is the assumption of a necessary connection between physicality and phenomenality. This would involve suggesting that the phenomenal facts are necessitated by the physical facts, even though the phenomenal facts are distinct from the physical facts. However, as Chalmers concedes, this suggestion is deeply problematic and unwarranted in light of how we understand modality. First, the necessitarian claim that something may be logically possible yet could not happen is false, insofar as whether something is possible tends to be defined by whether it can happen. Second, even if it is assumed that there is such a strong relation between physicality and phenomenality, then this would still be insufficient to defend physicalism, because it still acknowledges that physicality and phenomenality are distinct domains. Rather, it would amount to a form of dualism, whereby physicality and phenomenality are essentially separate domains that are strongly associated. Even where there is transcendental dependence, the domains are still ontologically distinct. For example, a phenomenal quality is transcendentally dependent on the existence of consciousness, because consciousness is the existence wherein such a quality manifests. Yet, it is true that consciousness is a separate feature from the phenomenal quality. Given all the facts about the phenomenal quality, the first-person individuation of consciousness remains a further fact. Notably, there remains the fact that this quality is experienced by *me* and not by *you*. Moreover, consciousness could exist independently without the phenomenal quality, which indicates that consciousness is a more fundamental feature than the phenomenal quality. Therefore, necessitarianism fails to undermine

dualism, because it is true that a necessitated entity could exist as an ontologically separate entity. Third, necessitarianism is undermined by the conceivability of modal variation between distinct domains. This recalls David Hume's (1748) proposal that there is no necessary connection between distinct events, because it is conceivable that any such association that happens to obtain between such events might not have obtained. Modal variation is always conceivable between domains that are not linked by identity, logical entailment, or transcendental dependence. For example, Chalmers (1996) notes that a dropped stone may move toward the ground in this world, but it is conceivable that there is a possible world wherein the dropped stone moves away from the ground. Such counterfactual reasoning requires the relevant relation to be contingent. Similarly, modal variation is always conceivable between physical and phenomenal domains. As noted above, a phenomenal quality is transcendentally dependent on the existence of consciousness, because such a quality is subjective, and so only manifests in the subjective existence of consciousness. However, there is no such dependence in the relation between physicality and phenomenality. Whereas physicality is essentially third-person, phenomenality is essentially first-person, and so the two domains have fundamentally different essential natures that do not entail any connection with each other. There is nothing in the third-person facts about the physical world that entails a connection with first-person subjectivity, just as there is nothing in the first-person subjective character of experience that entails a connection with the third-person physical world. Therefore, necessitarianism is false regarding the relation between physicality and phenomenality. That is to say, the claim that there is a necessary connection between physicality and phenomenality is false.

Panpsychism: This is the view that all matter has physical and mental properties. Such a view is exemplified by the substance monism of Baruch Spinoza (1677), the monadological monism of Gottfried Wilhelm von Leibniz (1714), and the process theory of Alfred North Whitehead (1933). Under this view, mentality is an intrinsic property of matter. Moreover, panpsychism sometimes has an emergentist component, insofar as it suggests that more complex configurations of physical matter are associated with more complex mental qualities. However, I argue that monist panpsychism is false for two reasons. First, the claim that mentality is an intrinsic property of matter suggests that there is a necessary connection between mentality and physicality, but I have already shown that this is deeply problematic. As noted earlier, there is no necessary

connection between distinct domains that are not linked by identity, logical entailment, or transcendental dependence, because modal variation is always conceivable between such domains. Given that matter is essentially third-person and consciousness is essentially first-person, the two features have fundamentally different essential natures that do not entail any connection with each other. There is nothing in the objective third-person facts about matter that entails a connection with first-person subjectivity, just as there is nothing in the first-person subjective character of experience that entails a connection with a third-person configuration of matter. Thus, the claim that consciousness is an intrinsic property of matter is necessarily false, because the first-person ontology of consciousness and the third-person ontology of matter are essentially distinct domains that do not entail each other. There is no necessary connection between matter and consciousness, and so necessitarianism is false regarding the relation between physicality and phenomenality. This is shown by the fact that it is logically conceivable for the relations between phenomenal qualities and physical properties to vary across possible worlds. For example, a given structural and dynamical property may be associated with a given phenomenal property in this world, but it is conceivable that this structural and dynamical property could be associated with a different phenomenal property in another possible world. Hence, given all the physical facts about a certain configuration of matter, the presence of consciousness remains a further fact to consider. This suggests that a putative feature that possesses a physical property and a mental property would not be a single mereologically simple unit, but at would just describe a conjunction of a physical property and a mental property which are ontologically separate from each other. Thus, monist panpsychism is false, because it fails to acknowledge the ontological gap that is entailed by the contingent relation between first-person phenomenality and third-person physicality. Second, the emergentist version of panpsychism is problematic insofar as it suggests that mental properties can be combined just as physical properties can. As noted in chapter two, the identity of any given consciousness is determined by its unique first-person individuation, which is essentially different from the first-person individuation of any other consciousness. And so, consciousnesses exist as ontologically discrete units that are essentially separate from one another in virtue of their unique ipseities. This indicates that the claim that consciousnesses could undergo fusion is false. Fission and fusion are impossible with regard to consciousness, because the first-

person individuation of consciousness is discrete. Therefore, emergentist panspsychism is false.

Neutral Monism: This departs somewhat from standard physicalism, but I include it here as a form of physicalism because, like many other physicalist positions, it attempts to give a reductive account of experience. Neutral monism is a position, associated with Ernst Mach (1886) and Bertrand Russell (1927), which claims that physical and mental properties can be reduced to a single underlying neutral property. This putative property is neither physical nor mental, but physical and mental properties are purported to emerge from it. I argue that neutral monism is false for similar reasons to why reductive physicalism and monist panpsychism are false. First, insofar as the putative neutral property is supposed to be nonexperiential, there is an ontological gap between this nonexperiential neutral property and an experiential mental property. Given all the nonexperiential neutral facts about the neutral property, the experiential nature of the mental property remains a further fact. Therefore, neutral monism is false, because nonexperiential neutral facts fail to account for the experiential existence of consciousness. Indeed, given this nonentailment from the nonexperiential to the experiential, any form of monism is necessarily false with regard to consciousness, because it fails to account for this ontological gap between third-person objectivity and first-person subjectivity. The first-person existence of consciousness is inevitably a further fact over and above the third-person facts. Second, given that physicality is essentially third-person and phenomenality is essentially first-person, the two domains have distinct essential natures that do not entail any neutral connection with each other. There is nothing in the third-person facts about the objective world that entails a neutral connection with first-person subjectivity, just as there is nothing in the first-person subjective character of experience that entails a neutral connection with the third-person objective world. Hence, neutral monism is false, because the first-person subjective character of experience does not entail any connection with a neutral property. Again, a putative neutral property that is associated with a physical aspect and a mental aspect would not be a single mereologically simple unit, but would just describe a conjunction of a physical property and a mental property which are ontologically separate from each other. Third, the claim that consciousness could be reduced to a more fundamental property is false, because consciousness is the first-person existence through which any property is realised. Hence, it must be taken as true that the existence of consciousness is

fundamental, as the realisation of any property presupposes the prior existence of consciousness. It is false to suppose that something could be more fundamental than consciousness, because something is only realised through consciousness. Otherwise, it would just be a potential with no reality. Insofar as it is nonexperiential, this potential would fail to account for the experiential nature of consciousness. Consciousness would remain a further ontological fact. Thus, given its fundamental nature, it must be taken as true that consciousness exists independently as an ungrounded entity.

Promissory Materialism: Although no satisfactory physical account of conscious experience has yet been given, one might still hope that a satisfactory physical account could be given in the future. However, I argue that promissory materialism is a false hope. This is because physical accounts are structural and dynamical accounts, and so they can only capture structural and dynamical facts. They cannot capture the subjectivity of consciousness. Hence, the claim that consciousness could be physically explained is false. In response, one might suggest that new physical properties could be discovered which are not structural and dynamical. My reply to this is twofold. First, the suggestion is analytically false. A property is defined as being physical in virtue of its being structural and dynamical, and so any new property which is discovered not to be structural or dynamical cannot be said to be physical. Second, even if such a property is discovered, I argue that it would fail to explain consciousness. Insofar as such a property is nonexperiential and consciousness is experiential, there would remain an ontological gap between the nonexperiential and the experiential. Over and above the facts about that property, the existence of consciousness would remain a further fact. Therefore, promissory materialism is false.

I have shown in this section that physicalism fails to account for consciousness. Beyond all the physical facts about the world, the existence of consciousness remains an extra fact. Therefore, physicalism is necessarily false with regard to consciousness. In the following section, I shall say more about how the issues discussed above relate to the conflation of consciousness with intentionality.

Intentionality

As noted in chapter one, consciousness often falsely conflated with various psychological capacities, such as awareness and perception. Usually, when a scientific theory claims to have explained

"consciousness", it has actually explained a psychological capacity that has been erroneously labelled "consciousness". On closer analysis, it often turns out that such a psychological capacity does not involve consciousness, but involves intentionality.

Intentionality is a concept associated with Franz Brentano (1874) and pertains to the aboutness or directedness of states of mind. Mental states are often representations of things. For example, when I think of an apple, my thought is about an apple.

It is not difficult to see how intentionality is related to psychological features such as awareness and perception. To be aware implies awareness of something. To perceive implies the perception of something. Other psychological features have more specific intentional objects. For example, introspection is the awareness of one's own internal state. Likewise, self-awareness is the awareness of oneself as an individual, which may involve some kind of representation of oneself. It could be contended that intentionality is what mental states have in common, and so for a state to be mental is for it to be intentional.

According to John Searle (1992), experience is essential to intentionality, but I argue that this is mistaken. Indeed, this claim seems to involve a false conflation between consciousness and intentionality. Herein, I argue that Searle is wrong about conscious experience. Subjective qualities are not essential to intentional states. Accordingly, I propose that there is no necessary connection between consciousness and intentionality, although the two are sometimes contingently related. This also has implications for the coherence of the concept of an unconscious mental state.

As Searle suggests, an intentional state has an "aspectual shape", which is supplied by its phenomenal quality. Given that what we call "unconscious mental states" have no phenomenal qualities associated with them, Searle argues that they cannot be intentional states. Accordingly, he claims that they are not mental states at all, but rather are just neurophysiological states. Some objections can be made to Searle's position. For example, Norton Nelkin (1993) argues that a phenomenal quality may accompany the introspection of an intentional state, but that such a phenomenal quality is not part of the state itself. Therefore, according to Nelkin's analysis, a phenomenal quality may be associated with the introspection of an intentional state, but is can be conceptually separated from the state itself. It follows that a phenomenal quality is not essential to intentionality.

What is this "aspectual shape" if it is not a phenomenal quality? A possible suggestion is that it is the linguistic content of the

intentional state, which will depend on the social norms and conventions of the interpersonal setting wherein one is situated. This linguistic content determines what the state is. For example, "being excited about a journey" is different from "being excited about a party" and "being excited about a party" is different from "being apprehensive about a party". If we apply a functionalist analysis, we could say that states with different linguistic content have different causal roles. For example, being excited about the party causes me to act in a different way towards the people at the party from how I would behave towards them if I was apprehensive about the party.

Viewed in this light, different intentional states involve dispositions to behave in different ways. There is no need to invoke subjective qualities here. Think about how we ascribe intentionality to others. Although we do not doubt that others are conscious, we have no direct access to their subjective experiences, and yet we are able to ascribe intentional states to them. For example, we can say that someone is frightened without having any knowledge about the phenomenal quality of that person's subjective experience. Accordingly, the presence of consciousness is an extra fact over and above the facts about intentionality.

What follows from this is that the concept of an unconscious mental state is coherent. Phenomenal qualities are not essential to intentional states, and so there is no logical contradiction in the notion of a mental state that is not associated with phenomenal quality. Does this mean that there actually are, in the world, such things as unconscious mental states? Do what we call "unconscious mental states" actually possess intentionality, or are they just neurophysiological states to which we mistakenly attribute meaning?

I suggest that there are such things as unconscious mental states, but that they are higher-level theoretical concepts that we infer and ascribe to one another to understand and make meaningful sense of one another's actions. Indeed, intentionality is itself an explanatory concept we use to help us understand and account for behaviour. For example, consider that someone says to me, "I am anxious". I could attempt to characterise this speech act in neurophysiological terms, by describing the activity in the person's limbic system, which lead to the coordinated firing of certain neurones in various areas of the cerebral cortex, and which finally result in the movement of the lips and tongue and the production of sounds from the person's vocal cords. However, as well as being convoluted and requiring a detailed knowledge of the person's neurophysiology, this characterisation would omit important contextual information about the interpersonal

setting that could be relevant to the implications of the speech act. Instead, I account for the person's behaviour meaningfully by attributing, to the person, an intentional state with semantic content, which in this case is "being anxious".

The use of a higher-level concept such as intentionality in an explanation makes the explanation simpler and more comprehensive. It is simpler, because it does not require convoluted details about lower-level processes. I can understand the person's behaviour by appealing to the intentional state "being anxious", rather than by bringing up details about neurophysiological processes. It is more comprehensive, because it considers contextual factors, such as how the meaning of the attributed intentional state must be interpreted relative to the linguistic and social norms and conventions of the interpersonal setting wherein the person is situated.

Given that it is a higher-level theoretical concept, I argue that there is no problem with ascribing intentionality to "unconscious mental states". As I stated earlier in this chapter, an "unconscious" mental state differs from a "conscious" mental state, first, with respect to the fact that we can introspect and report "conscious" mental states but not "unconscious" mental states and, second, with respect to the fact that "conscious" mental states have specific qualia associated with them whereas "unconscious" states do not. We experience unconscious mental states only through their behavioural manifestations. However, since qualia are not essential to the semantic content of intentional states, I argue that if we are able to ascribe intentionality to "conscious" mental states, then there is no reason why we cannot ascribe it to states that are unconscious.

A case where it is useful to ascribe intentionality to unconscious states is in the study of the associations between ideas in a chain of thoughts. For example, consider that I have a thought of an espresso, which is followed by the thought of a continental café, which is then followed by the thought of my birthplace in Burma. The psychological theory of associationism, advocated by William Hamilton (1865), suggests that each idea in a chain of thoughts is thematically associated with the ideas immediately preceding and following it. In the chain of thoughts presented above, the first two thoughts are obviously associated with each other, since espresso is served in a continental café. However, there appears to be no obvious association between the thought of a continental café and the following thought of Burma. What links these two thoughts?

It is possible to assert that all that occurs between the thought of a continental café and the following thought of Burma are neural

processes in the brain, or what William Carpenter (1874) called "unconscious cerebration". Under this view, there is no intermediate thought with intentional content between the two thoughts. However, this would contradict associationism, insofar as it denies that each thought is thematically associated with thoughts ones preceding and following it. To resolve this, Hamilton (1865) proposed that there are intermediate thoughts with intentional content which link seemingly unrelated thoughts, but that these intermediate thoughts are not straightforwardly accessible to introspection. That is to say, they are unconscious thoughts. And so, regarding the chain of thoughts I presented above, it can be conjectured that the unconscious thought which connects the thought of the continental café and the thought of Burma is the thought of familial love, as this specific continental café is run by a family whose love and care for one another reminded me of my family in Burma. This purported intermediate thought may be unconscious, but it is thematically associated with the other thoughts in a manner that is compatible with associationism.

By considering intentionality as a higher-level theoretical concept, it also follows that we are able to ascribe it to other processes, such as the workings of certain artificial systems. Furthermore, given that intentionality is independent of consciousness, the ascription of intentional states to these artificial systems makes no assertions about whether they have subjective qualities or not. These artificial systems may, indeed, turn out to be conscious, but this is orthogonal to the question of whether they can be said to have intentional states. Thus, we may be able to describe certain artificial systems as being "intelligent" without having to worry about whether their workings are accompanied by qualia.

And so, it is true that consciousness and intentionality are ontologically separate features. Consciousness refers to first-person subjective existence, whereas intentionality refers to a psychological property that is ascribed to an individual to explain the individual's behaviour. Indeed, many of the things to which we ascribe intentionality in this world do, in fact, turn out to have consciousnesses associated with them. However, this association between the presence of intentionality and the presence of consciousness is contingent. While they are often associated with each other in this world, intentionality and consciousness can come apart. Accordingly, it is a mistake to conflate consciousness with psychological properties such as awareness and perception, because these properties pertain to intentionality and not to consciousness.

IV
Verifying Dualism

In chapter three, I argued that a physical account of consciousness is impossible and that consciousness is fundamentally beyond science. In this chapter, I argue that this is because consciousness is not physical, but exists as a separate entity. Given all the physical facts about the world, the existence of consciousness is a further fact to consider. Therefore, physicalism is false. I begin by presenting some arguments against physicalism, which demonstrate that phenomenality does not supervene on physicality. In virtue of this nonentailment from physicality to phenomenality, dualism is true.

The arguments

The knowledge argument

The central thesis of the knowledge argument for dualism is that full knowledge about the subjective quality of experience can be acquired only by having the experience oneself. No amount of physical knowledge about the structure and dynamics of the stimulus that causes the experience, or of what happens in the brain when one has such an experience, entails what the experience is *like* qualitatively. It follows from this that the physical facts do not exhaust the phenomenal facts, and so physicalism is false.

This argument is illustrated by Frank Jackson (1982) through the thought experiment of Mary, the colour scientist who has spent her entire life in a monochrome environment exclusively coloured in black, white, and shades of grey. Consequently, Mary has never subjectively experienced the colour red. Nevertheless, via black-and-white television, she has become the world's leading expert on colour perception, and knows everything physical that occurs inside a perceiver's brain when the perceiver sees red. Furthermore, Mary lives in a time in which we have a completed physics, and so also has learned everything physical about the structure and dynamics of red light. Mary, therefore, can be said to have the complete knowledge of the physical facts about the colour red and its perception.

However, as Jackson argues, Mary does not have the complete facts about red experience, since she cannot derive, from the physical

facts, the subjective quality of a red experience. She knows all about the structure and dynamics of red light and red perception, but having never experienced red herself, she does not know what red is like qualitatively. Thus, Jackson proposes that if Mary is released from her monochrome compound and sees the colour red for the first time, she gains knowledge that she did not possess before. This knowledge is of the subjective quality of red experience, and, since Mary already had possessed the complete physical knowledge of the colour red and its perception, this new knowledge she gains must be nonphysical. The conclusion, from this, is that there exist phenomenal facts over and above the complete physical facts, and that these phenomenal facts cannot be explained by, or identified with, the physical facts. Thus, physicalism is false.

The argument can be taken even further by considering systems different from ourselves, such as machines. It can be conjectured not only that no amount of physical information about a system can tell us anything about what its experience its like, but also that no amount of physical information can tell us whether it is accompanied by experience at all. Nothing in the physical facts logically entails the actual existence of consciousness itself. As David Chalmers (1996) notes, when we have the complete physical facts about a system, the issue of whether or not the system is accompanied by consciousness remains a further fact. Although the system may, in fact, be conscious, it is just as logically compatible with the physical facts that the system is not accompanied by consciousness.

As Chalmers suggests, we could design a computer with simple cognitive and perceptual abilities, such as that of recognising colours. The computer may even be modelled on the human visual system and categorise colours in a similar manner to us. However, even if we know everything physical there is to know about the computer's circuits, there would still remain the question of whether the computer's processing is accompanied by consciousness. Even with the complete physical facts about the computer, we cannot derive the answer to this question. Of course, it may actually be the case that the computer is associated with consciousness, but the point is that this fact is not entailed by the complete physical facts. And so, despite our physical knowledge of how the computer works, the existence of consciousness is still a further fact to consider.

A similar approach is taken by Thomas Nagel in "What Is It Like to Be a Bat?" (1974), in which he argues that no amount of physical knowledge about a bat's nervous system can tell us what it is like to be a bat. Nagel's argument is based on the fact that this phenomenal

knowledge is only available from a bat's perspective. That is to say, this phenomenal knowledge is first-person knowledge that is available only to a subject via the subject's own experience. In contrast, physical knowledge is third-person knowledge of objective facts, and is available to all. This objective physical knowledge cannot provide us with any information about the subjective, for it only provides third-person information about the structure and dynamics of the object that is experienced, but not first-person information from the viewpoint of the experiencer. In fact, Nagel argues that we cannot even imagine what it is like to be a bat. Rather, when one tries to imagine what it is like to be a bat, one is, in fact, imagining what it is like for one to behave like a bat. The imagining is being done from the first-person point of view of the imaginer, and so provides no insight into the first-person point of view of the bat.

The argument presented by Nagel appeals to the fact that bats are so different from humans, and so a bat's perspective seems so inaccessible to us. However, I suggest that this dissimilarity between bats and humans is not actually necessary for Nagel's argument to work. I propose that one cannot even know for certain what it is like to be another human being. Given the first-person subjectivity of consciousness, one has privileged access to one's own experience, but not to anyone else's. Therefore, one cannot know what it is like to be someone else. The most one can do is to assume that the other person's experience has a similar quality to one's own experience based on the fact that the two people are physiologically similar.

For example, Mary and Toby both look at a red apple, and agree that the apple is red. This is objective knowledge, available to both Mary and Toby, about a third-person object, namely the apple. However, Mary has no access to Toby's experience of red, and Toby has no access to Mary's experience of red. Even though they both consciously experience the apple and agree that it is red, they have no idea about the subjective quality of the other's experience of red. Indeed, Mary's experience of red may be of a completely different quality to Toby's experience of red, but since they have no access to each others' experiences, they would never know this for sure. All we can say is that Toby and Mary have good reason to assume that their experiences of red are the same, based on the fact that they are embodied in similar ways.

I have presented Jackson's argument alongside Nagel's, because both demonstrate the same principle, specifically that no amount of physical knowledge can provide us with the complete information about the subjective quality of experience. In the case of Jackson's

argument, Mary, despite having the complete physical knowledge about the colour red and its perception, lacked phenomenal knowledge about red experience before her release. Similarly, Nagel argues that even if we were to possess the complete physical knowledge about a bat's neurophysiology, we would not know what it is like to be a bat.

And so, the key conclusion of the knowledge argument is that facts about consciousness are not entailed by physical facts, and so are separate facts over and beyond the physical facts. As Chalmers (1996) notes, consciousness is not logically supervenient on the physical. This is seen as a powerful refutation of physicalism, and there have, indeed, been several physicalist replies to it, but I argue that these replies necessarily fail. I now consider some of these replies and defend the knowledge argument from them.

Although Jackson agrees that Mary gains knowledge about her own experience upon her release, he argues that it is the knowledge about the experiences of others that she gains which is of particular importance for the knowledge argument against physicalism. Before her release, she had the complete physical knowledge about the colour red and what happens in a person's brain when that person perceives red. However, she did not have the complete knowledge of that person's subjective experience of red. Jackson suggests that upon her release, Mary sees a red apple, and finds out what a red experience is like. From this new experience of hers, she gains knowledge about the subjective experience that a person has when that person sees red.

However, this argument is based on the assumption that the experience that Mary has when she sees red is the same experience that someone else has in the same situation. As I argued earlier with reference to Nagel, experiences are subjective, and so we cannot know about the subjective qualities of the experiences of others. We can only know about the subjective qualities of our own experiences. Therefore, given this irreducible subjectivity of consciousness, Mary cannot know for sure whether her experience of red is the same as someone else's experience of red. It follows that upon her first experience of red, she only gains knowledge about her own experience, not about the experiences of others.

Why, in his argument, does Jackson feel that he needs to show that Mary gains knowledge about the experiences of others, instead of simply leaving us with the conclusion that she gains knowledge about her own experience? Perhaps he thinks that while both the physicalist and the dualist could agree that Mary gains new

knowledge about her own experience upon her release, the physicalist will just claim that this knowledge is just about a new brain state in which Mary had never previously been. If this is so, then the fact that Mary learns something new about her own experience upon her release cannot be used as an argument against physicalism, for the physicalist can just claim that this new knowledge is just further physical knowledge. Instead, it is only the knowledge Mary gains about the subjective experiences of others that shows that physicalism is false, since it shows that despite her complete physical knowledge about the brain states of others when they perceive red, she did not know the subjective quality of their red experiences.

Contrary to the above, I argue that Mary does gain knowledge about her own experience and that this is enough to refute physicalism. Note that before her release, Mary had the complete knowledge of all the physical facts about the colour red and its perception. From these physical facts, she should, in principle, have been fully capable of predicting what brain state she would be in if she were to see red, yet she was still unable to derive what a red experience is like qualitatively. Therefore, the physicalist's claim that the new knowledge Mary gains from her first experience of red is just new knowledge about a brain state is false. Indeed, Mary could not gain any new physical knowledge about a brain state after her release, because she already knew everything physical there is to know about that brain state before her release. Nevertheless, despite her possession of this knowledge, she did not know what the subjective quality of a red experience is like before her release and only gains this knowledge after her release.

A possible reply to this is that if Mary had the complete physical knowledge about her brain states before her release, she could have easily worked out what a red experience is like by using this physical knowledge to manipulate her brain in such a way that causes her to have a red experience. This, however, is beside the point. Indeed, Mary could manipulate her brain in such a way that causes her to have a red experience and she could learn about the subjective quality of experience via this method. However, the fact that she has to make herself experience red in order to learn about its subjective quality shows that her knowledge of the physical facts, on its own, could not provide this information. After all, she already had possessed the complete physical facts about the colour red and its perception before she manipulated her brain, but, solely from these facts, she could not work out the subjective quality of a red

experience. Only through having the experience herself does she gain this extra fact.

And so, we do not need to appeal to knowledge about the experiences of others for the knowledge argument to be sound. The fact that Mary gains new knowledge about her own experience after her release is enough to show that subjective experience is not entailed by the physical facts. Given that subjective experience is a further fact beyond the physical facts, dualism is true.

Another objection to the knowledge argument, raised by Lawrence Nemirow (1990), questions the sort of knowledge that Mary gains from her experience of red. Specifically, Nemirow deploys the distinction between "knowing that" and "knowing how". This is the distinction between the knowledge associated with new facts and the knowledge associated with new abilities. He then proposes that Mary, upon her first exposure to red, gains the latter sort of knowledge, namely the practical ability to recognise the subjective quality of red. The acquisition of this ability does not necessarily involve the learning of new facts. Therefore, if new abilities are all that Mary acquires from her first experience of red, then the knowledge argument loses its force. Since no new facts are learned from Mary's experience of red, it follows that there is not any information left out of her previous physical account of red.

The reply to this is straightforward. Mary does gain a new ability from her first experience of red, but she also gains new facts. From her first experience of red, Mary gains a fact about what red is like phenomenally. As Chalmers (1996) notes, before her release, Mary did not know what it is like to experience red, since the only knowledge she possessed were structural and dynamical facts about red light and perception. As far as she was concerned, the experience of red could be like anything, or it might not be like anything at all. However, after her release, she learns that the experience of red is, in fact, like what it is like when she first experiences it. Therefore, Nemirow's ability reply does not hold. As well as gaining a new ability, Mary gains new facts, and these facts are over and above the complete physical facts that she previously knew.

Going even further, Daniel Dennett (1991) claims that Mary learns nothing at all from her first experience of colour following her release. That is to say, she does not even gain new abilities. To make this claim, Dennett focuses on the assumption in the argument that Mary has all the physical facts about colour perception. From this, he argues that if Mary has all the physical facts, which include facts about one's behavioural and cognitive reactions to perceiving certain

colours, then upon her first experience of seeing a particular colour, she should be able to use her neurophysiological and psychological knowledge to identify what colour she is seeing, perhaps by observing her own behavioural reactions and thoughts to perceiving that colour.

To illustrate this, Dennett suggests a thought experiment involving Mary's release. Upon her release, Mary's captors decide to trick her and present her with a blue banana, while claiming to her that the banana is, in fact, yellow. However, with her complete physical knowledge of colour perception, Mary notices that her behaviour and thoughts upon seeing the banana correspond closely to what she knows are the natural reactions one has when one sees the colour blue. Therefore, Mary is able to recognise that the banana that has been presented to her is not yellow, but is blue.

According to Dennett, therefore, Mary does not even gain any new abilities from her first experience of colour, for she had already possessed the ability to recognise colours before her release, solely from her physical knowledge. This may be so, but Dennett's argument misses the point. The question is not whether Mary gains new abilities or not, but whether she gains new facts. From my previous discussion of the ability reply, we can conclude that Mary does indeed gain a new fact about what a red experience is like qualitatively, which Dennett's argument does not consider.

It seems that Dennett's argument is focusing not on consciousness, but on reportability. This refers to one's ability to recognise and report the contents of one's awareness and perception. It is a dynamical aspect of an individual's operation, and so there is no reason why it cannot be explained in physical terms of structure and dynamics, perhaps with the aid of a cognitive model that charts the flow of causal dynamics in one's mental apparatus that consequently result in the generation of behaviour. This is entirely different from consciousness. Whereas reportability is referring to a psychological capacity, consciousness is referring to the subjective quality of experience. Nothing in accounts of reportability suggests that it should be accompanied by conscious experience. Dennett's description of Mary recognising a blue banana is merely a description of her capacity to report. Her cognitive abilities allow her to be able to discriminate blue from yellow, and she is able to communicate this to her captors. However, nothing is mentioned about the subjective quality of her experience of blue. Therefore, Dennett's argument is unsound, because he falsely conflates consciousness with reportability.

From a different angle, Paul Churchland (1985) argues that the knowledge argument proves too much. Recall that the knowledge argument suggests that Mary learns everything physical about the colour red and its perception while she is in her monochrome isolation, perhaps, via a series of black-and-white television programmes, books, and journals. From these, she gains complete physical knowledge about the colour red.

Building on this picture, Churchland suggests that in addition to being educated about the physical facts about the colour red, Mary also receives a series of lectures, over her black-and-white television, from a dualist, who teaches her everything phenomenal about a red experience. According to Churchland, if this was the case, then Mary would know everything physical and phenomenal about the colour red before her release. However, if she sees a red apple for the first time after her release, one would still have the intuition that she learns something new from the experience and that this knowledge could not be about a physical or a phenomenal fact, since she already knew all the physical and phenomenal facts before her release.

And so, Churchland is using the same argument that Jackson uses against physicalism to argue against dualism. If the new facts that Mary learns upon her experience of red are not physical or phenomenal, then it follows that dualism is no better off than physicalism in explaining all the facts. Rather, a new position that accounts for physical, phenomenal, and nonphysical-and-nonphenomenal facts is needed. This would seem to lead to absurdity, as the same argument can, again, be used on this new position. If Mary was educated about the physical, phenomenal, and nonphysical-and-nonphenomenal facts before her release, and still learned new facts upon her first experience of red, then it appears that this new position cannot account for all the facts either. Yet another position is needed, but once again, this will be unsatisfactory, since the same argument can again be used against it. Therefore, Churchland claims that the knowledge argument proves too much, for it can be adapted and used to criticise its own conclusions. If we use the knowledge argument, we will never find a satisfactory position.

In reply, Jackson's (1986) contends that lectures over black-and-white television can teach Mary everything physical about the colour red, but they cannot teach Mary everything phenomenal about a red experience. As he notes, "you do not need colour television to learn physics or functionalist psychology". While Mary can learn the physical facts about red from black-and-white television,

Churchland's suggestion that Mary learns everything phenomenal about red from black-and-white television is impossible.

Here, Jackson is appealing to the notion of epistemic asymmetry. Physical facts are epistemically third-person. They are objective facts concerning the structure and dynamics of the external world, and so they are, in principle, accessible to all and can be communicated. In contrast, phenomenal facts are epistemically first-person. They are not facts about the external world, but about one's individual subjective experience. Therefore, there is a degree of first-person privacy about phenomenal facts. One has privileged access to one's own qualia, but not to the qualia of others. Our knowledge of qualia comes from our own experiences of them.

It follows that Mary can learn all the physical facts about the colour red and its perception from lectures via black-and-white television, since these facts are epistemically third-person. They are structural and dynamical facts about external features, and so they can be fully communicated to Mary via black-and-white television. However, phenomenal facts about red qualia cannot be communicated to Mary in the same way. Phenomenal facts are epistemically first-person, and so Mary's knowledge of them can only come from her own subjective experience. Subjective qualities cannot be communicated objectively in the third-person, because they are fundamentally experiential.

Hence, Churchland's objection fails to undermine the knowledge argument. His argument is based on the premise that Mary learns all the phenomenal facts about red qualia before her release. However, epistemic asymmetry indicates that this is not possible. Phenomenal knowledge is only acquired through direct experience. Given this, the knowledge argument cannot be used against its own conclusion.

The conceivability argument

This argument also defends the idea that consciousness is not logically supervenient on the physical. As has been shown by the knowledge argument, even when the complete physical facts about a system are established, the question of whether consciousness accompanies the system is left open. The physical facts do not entail the existence of consciousness, and so it follows that consciousness is an extra fact over and above the physical facts.

From this, the conceivability argument proposes that given any physical system, it is logically conceivable that such a system could lack conscious experience, even if it is a fact that the system is

actually accompanied by consciousness in this world. The reason for this is that one can describe and explain the structure and dynamics of a physical system without having to refer to subjective experience at all. Therefore, when given all the third-person physical facts about structure and dynamics, the presence of first-person subjective experience remains a further fact to consider. The third-person physical facts about a system fail to account for why there is an individuated first-person subject accompanying this system.

This argument is developed by David Chalmers in *The Conscious Mind* (1996), wherein he argues for the logical conceivability of zombies, or biological systems that physically indistinguishable from humans but which are not associated with consciousnesses. As Chalmers notes, these zombies are very different from the "zombies" seen in horror films. The "zombies" seen in horror films have notable differences from humans with regards to their psychological capacities, but it is reasonable to assume that these "zombies" are conscious. By contrast, the zombies proposed by Chalmers are nonconscious, but are physiologically and behaviourally indistinguishable from humans. It is this nonconscious kind of zombie to which I shall be referring throughout this chapter. Consider that I have a zombie twin, which is physically indistinguishable from me in every respect. However, a difference is that I, myself, am conscious, whereas my zombie twin is not. While there is consciousness associated with me, there is no consciousness associated with my zombie twin. However, since my zombie twin is physically indistinguishable from me, we both behave in indistinguishable ways. The same stimuli bring about the same neural processes in our brains, and so we both have the same reactions and perform the same actions. The difference is that my behaviour is accompanied by consciousness, whereas my zombie twin's is not. The neural processes in my brain are accompanied by subjective experience, whereas there is no experience at all in the case of my zombie twin.

According to Chalmers, there is no logical contradiction in this seemingly peculiar situation. The behaviour of an organism depends on its physical structure and dynamics, but nothing in these structural and dynamical properties entails the presence of subjective experience. It is perfectly conceivable to characterise the activity of a system in terms of structure and dynamics without having to bring up consciousness at all. Therefore, given the complete physical facts about a system, the existence of consciousness is an extra fact over and above the physical facts. It also follows that the physical facts

cannot help us to distinguish me from my zombie twin, since we are physically indistinguishable.

Another way to approach this argument is to consider what Chalmers calls "nonstandard realisations" of the causal structure of a system. That is to say, the causal dynamics within the human nervous system which are responsible for the generation of behaviour can, in principle, be realised in different ways. The example used by Ned Block (1978) is the construction of an isomorphic realisation of the brain using the entire population of a country. In this realisation, neurones are replaced by individual members of the population, organised in such a way that its overall activity is analogous to that of a human brain, but on a much larger scale. The significance of this hypothetical example is not to show whether or not such a realisation would be accompanied by consciousness. It would be reasonable to suggest, in this world, that it would not. Rather, this example shows that even though consciousness may not accompany such a realisation in actuality, it is equally coherent logically whether it does or does not. Given the complete physical information, the presence or absence of consciousness is still an open question. It is not logically entailed by the physical facts about a system's organisation. Therefore, consciousness does not logically supervene on the physical.

One objection to the conceivability argument is that zombies are naturally impossible. That is to say, the existence of zombies is incompatible with the regularity of the laws of nature. Given the fact that I am conscious, it is perfectly reasonable to assume that any physically indistinguishable replica of me would also be conscious. In fact, given the fact that my replica is physically indistinguishable from me, it would appear quite arbitrary for me to be conscious and for it not to be. Thus, zombies may be naturally impossible in the actual world. However, this objection misses the point of the conceivability argument. As noted by Chalmers (1996), the question is not whether zombies are naturally possible, but whether the idea of a zombie is conceptually coherent. I argue that there is no contradiction in the idea. Given that the behaviour of a person can be explained entirely in physical terms of structure and dynamics, there is no need to bring subjective qualities into the picture. Thus, it is entirely coherent to think of a zombie that is structurally and dynamically indistinguishable from that person, but lacking consciousness. In agreement with Chalmers, I suggest that the mere logical coherence of the idea of a nonconscious human replica is enough to establish a conclusion. It reveals that having considered all

the physical properties of a system, whether this system is accompanied by consciousness is still an extra fact to consider. Indeed, Chalmers accepts that nonconscious human replicas may not be naturally possible in this world, but nothing in the physical facts entails that this has to be the case.

Some critics, including Anthony Marcel (1988) and Robert van Gulick (1989), have argued against the logical conceivability of zombies by proposing that phenomenal states have causal roles. For example, Marcel suggests that phenomenal states have roles in the ability to initiate actions with respect to parts of the world that are being experienced, the formation of an integrated concept of self from the experience of autobiographical memory, the ability to learn new nonhabitual tasks, and the ability to form plans of action. From this, van Gulick argues that zombies are not logically possible, since any nonconscious human replica would lack these causal roles proposed by Marcel, and, therefore, would not be structurally and dynamically indistinguishable from a conscious person.

I argue that Marcel and van Gulick fail to undermine the conceivability argument, because they are mistakenly conflating the phenomenal with the psychological. The abilities proposed by Marcel and van Gulick are structural and dynamical properties involved in cognition and behaviour. Insofar as they are structural and dynamical properties, they can be explained in structural and dynamical terms. Consciousness does not need to feature in the explanation. Not only are phenomenal qualities not required to explain cognitive abilities, but cognitive abilities are unable to explain phenomenal qualities. Subjective experience remains a further fact over and above the cognitive abilities.

The explanatory gap argument

Whereas the conceivability argument focuses on the actual presence of consciousness, Joseph Levine's (1983) explanatory gap argument focuses on the subjective qualities of particular conscious experiences. For example, consider the qualitative nature of a subjective experience, such as that of seeing the colour red. Here I am not concerned with any physical properties of the colour red, such as its wavelength or the neural processes that occur upon its perception, but with the actual subjective quality of a red experience. Upon reflection, it is clear that such an experience is arbitrary. As noted by Robert van Gulick (1993), a phenomenal hue has no structure, for it is a basic simple. It is a unique quality of its own

kind, and so any connection between it and anything else cannot be anything but arbitrary. It follows that there is no explanatory connection between physical events in the brain and the qualitative nature of subjective experiences. Physical information cannot explain subjective experience, and so physicalism is false.

Although the explanatory gap argument was developed by Levine (1983), the idea that qualia are arbitrary was anticipated by John Locke. In *An Essay Concerning Human Understanding* (1689), Locke notes that the ideas which are evoked by secondary qualities, such as the experiences of colour and taste, bear no relation to the events in the objective world that evoke such qualities in our minds. Rather, these ideas and their corresponding physical events are merely correlated in an arbitrary manner.

To further illustrate this explanatory gap between the physical and the phenomenal, Levine suggests the logical possibility of the spectral inversion of visual qualia, while one's neurophysiological processing remains the same. Perhaps the most vivid way to illustrate this is to consider a thought experiment described by Chalmers (1996), wherein we are presented with a hypothetical conscious being, who is physically indistinguishable from me, but whose visual qualia are inverted. That is to say, in situations in which I would have a red experience, this being will have a blue experience. Nevertheless, its activity remains indistinguishable to mine. It possesses the same neural mechanisms of colour processing as me, and, consequently, displays the same behaviour in response to seeing a red object. The only difference is that the phenomenal qualities that accompany this activity are different.

According to Levine and Chalmers, there is no logical contradiction here. Nothing in the physical activity of the brain states that one type of processing should be accompanied by one particular type of experience rather than another. Red processing with blue phenomenology is just as coherent as red processing with red phenomenology. The accompanying qualia are extra facts over and above the structure and dynamics of neural activity.

An objection to the explanatory gap argument is raised by Larry Hardin (1988), who argues that visual qualia are not as arbitrary as Locke and Levine suggest, but rather lie within a highly organised and asymmetrical colour space. Accordingly, Hardin argues that any changes made to this colour space, such as inversion, would disrupt its structural organisation and lead to unforeseen consequences. The arguments he gives for the asymmetrical organisation of the human colour space are as follows.

First, Hardin notices that certain colour qualia have other nonvisual qualia associated with them. For example, the experience of red is commonly associated with a "warm" feel, and the experience of blue with a "cool" feel. Furthermore, these properties are correlated with certain physiological reactions which occur upon the perception of certain colours. For example, the perception of a red object evokes "warm" reactions in one's physiology, and the perception of a blue object evokes "cool" reactions. If the colour spectrum of one's visual qualia is inverted, Hardin argues that a peculiar situation would arise wherein the "warm" phenomenology of a red experience would be dissociated from its "warm" physiological reaction. Rather, when confronted with a "warm" red object, one would have a "cool" blue colour experience, but still display a "warm" reaction. According to Hardin, this an odd idea.

Second, Hardin notes that some colours seem to be experienced as unary, such as red and blue, others seem to be experienced as binary, such as purple being a combination of red and blue, and some seem unimaginable, such as a colour that is both red and green. He explains this by referring to the neurophysiology of colour vision. Our colour spaces are not symmetrical, because we possess two underlying opponent colour channels. One channel discriminates between red and green, while the other channel discriminates between blue and yellow. Since red and green both are opponents on the same channel, a colour cannot be both red and green. Binary colours are combinations of the outputs of the two different channels, and so purple is a possible colour because it is a combination of a red output from one channel and a blue output from the other channel. Moreover, Hardin notices how some colours seem to be more dominant than others. For example, green is more dominant than yellow, in the sense that the intermediate colours between yellow and green are perceived as shades of green, rather than shades of yellow.

From this, Hardin concludes that visual qualia are not basic simples, but have a complex organisational structure that cannot be disrupted. Therefore, he suggests that the links between visual qualia and neural processes are not arbitrary. However, I argue that Hardin's objection fails to undermine the explanatory gap argument.

Concerning Hardin's first argument, a possible reply would be to bite the bullet and simply accept that an individual with inverted colour qualia would indeed have a "cool" experience when displaying a "warm" reaction. This is the approach suggested by Chalmers (1996). As noted by the explanatory gap argument, there is nothing logically incoherent about the notion of a "cool" experience

accompanying a "warm" reaction. Once we have the complete physical information about the causal dynamics in one's nervous system when one is having a "warm" reaction, the phenomenal quality of the accompanying subjective experience is a further fact.

While Chalmers' argument is sufficient to refute Hardin's objection, I argue that it is unnecessary, because we do not need to concede Hardin's claim that spectral inversion would result in colours being dissociated from their associated properties. According to Hardin, the "warm" and "cool" feels that are associated with our experiences of red and blue are properties of the red and blue experiences, but I disagree. Instead, I propose that the properties of "warmth" and "coolness" do not belong to the red and blue experiences themselves, but to the "warm" and "cool" judgements that occur upon perceiving red and blue objects respectively. These properties are only thought to belong to the red and blue experiences, because, upon perceiving red and blue objects, these red and blue experiences tend to occur respectively with the "warm" and "cool" judgements that are accompanied by these "warm" and "cool" phenomenal feels. Such judgements are partly culturally conditioned, as red and blue are conventionally used in some cultures to signify warmth and coolness. Hence, one associates a red experience with "warmth" because one's perception of an object associated with this experience, namely a red object, evokes a culturally conditioned "warm" judgement and an accompanying "warm" feel. Conversely, a blue experience is associated with "coolness" because the perception of a blue object evokes a culturally conditioned "cool" judgement and an accompanying "cool" feel. However, for a subject with inverted colour qualia, the "warm" judgement and "feel" associated with perceiving a red object would not be accompanied by a red experience, but by a blue experience, and so it would be this blue experience that would become associated with "warmth". Conversely, the red experience that accompanies the "cool" judgement that this subject has upon perceiving a blue object would become associated with "coolness". It follows that spectral inversion would not necessarily result in the properties of colour experiences being dissociated from their corresponding judgements.

The above suggests that the properties we ascribe to colour experiences are not actually properties of the experiences themselves, but are properties associated with the corresponding psychological judgements. For example, the "warmth" we associate with the colour red is not actually a property of the experience of red, but a judgement that we make upon perceiving red. This can be

thought of as analogous to the colour of an object, as opposed to one's experience of that colour, being a property of the object, rather than a property of the experience. To illustrate this, consider Mary and Toby experiencing a red apple. Now consider that Toby has inverted colour qualia. Despite the fact that they are having opposite colour experiences when they look at the apple, both Mary and Toby agree about the fact that the apple is red. This is because objectively, the apple possesses the property of redness, but subjectively, this redness is experienced differently by each of the individual experiencers' consciousnesses. Similarly, the "warmth" and "coolness" we associate with red and blue are not directly related to the experiences of red and blue, but are judgments we make during red and blue perception. As I have already suggested, a subject with inverted qualia could still make a "warm" judgement upon perceiving red, even if this reaction is accompanied by a blue experience. Subjective qualities can come apart from the properties ascribed to objects and the judgements about them, and so Hardin's objection fails to undermine the logical possibility of inverted qualia.

Concerning Hardin's second argument, which suggests that there are asymmetrical relationships between unary and binary colours, even if we concede that the human colour space may not be symmetrical, Sydney Shoemaker (1982) argues that there could be beings whose colour spaces are symmetrical. When this is taken into account, Hardin's objection to inverted qualia no longer presents any difficulty, since such inversion would not affect the organisational structure of a symmetrical colour space. For example, a conscious being may have a colour space with the colours *A* and *B* as the two extremes, which are associated with the perception of light with short and long wavelengths, respectively. This spectrum is entirely symmetrical, with the intermediate colours simply being combinations of *A* and *B* in varying proportions. However, there is nothing about the physical characteristics of this being's nervous system that indicates that the perception of short wavelengths has to be accompanied by the experience of *A* rather than *B*, or that the perception of long wavelengths has to be accompanied by the experience of *B* rather than *A*. There is no logical contradiction in conceiving of a second being who is physically indistinguishable from the first being, but has inverted visual qualia, so that *B* is experienced upon the perception of short wavelengths and *A* is experienced upon the perception of long wavelengths. Therefore, given a symmetrical colour space, the physical facts do not entail facts about the polarity of this colour space.

A possible objection to this is that we have good reason to believe that the human colour space is asymmetrical, and so, by proposing the possibility of beings whose colour spaces are symmetrical, Shoemaker is just avoiding the issue. Even if the spectral inversion of a symmetrical colour space, such as that of Shoemaker's hypothetical being, is possible, this still does not show that the spectral inversion of an asymmetrical colour space, such as that of a human, is possible.

In reply, I argue that the mere logical conceivability of the spectral inversion of a symmetrical colour space is enough to rebut Hardin's objection. Indeed, the asymmetry of the human colour space may render its inversion implausible, but there is no reason why we must limit ourselves to talking about the human colour space. After all, the point of the explanatory gap argument is to show that there is no entailment from the physical facts about brain states to the subjective quality of experiences, not to show that the human colour space can be inverted. The very fact that we can logically conceive of a situation in which colour qualia can be inverted shows that these qualia cannot be logically entailed by the physical facts, for, if they were logically entailed by the physical facts, the idea of such an inversion would be inconceivable to us.

Another reply is suggested by Shoemaker (1981) that could apply to beings with asymmetrical colour spaces, such as humans. His argument is that although we may be able to map out the organisational structures of our colour spaces, there will always be something qualitative fundamentally left unexplained, namely the subjective qualities of individual experiences. That is to say, no amount of explanation can capture the phenomenal redness of a red experience. To illustrate this, Shoemaker presents a thought experiment involving conscious beings who are physically indistinguishable from humans, but whose colour qualia are completely different from ours. The phenomenal colour spaces of such beings have the same organisational structures as human colour spaces. And so, although their qualia are unlike our own, their interrelations are perfectly analogous to those between our own. Instead of red, they experience red*, which bears the same relations to their blue* and green* that our own red respectively bears to our blue and green. As Shoemaker notes, we may be able to explain, with appeal to the organisational structure of one's colour space, why a given brain process is accompanied by phenomenal red rather than phenomenal purple, but we cannot explain why it is accompanied by phenomenal red rather than phenomenal red*. The subjective quality

of any given experience eludes explanation, and so remains a further fact over and above the organisational structure of the colour space.

The modal argument

Another physicalist approach is to deny the significance of the explanatory gap altogether by assuming an identity between an experience and its corresponding brain state. This is the identity theory developed by U. T. Place (1956) and J. J. C. Smart (1959). Consider, for example, the identity between Mark Twain and Samuel Clemens. It is reasonable to ask questions such as why this writer decided to go under two names and why it took so long for one to discover that these two names refer to the same person. However, it makes no sense to ask why Mark Twain and Samuel Clemens are one and the same person. They just are. When we realise that one person is assuming two names, there is no point in asking why there is only one person. Likewise, if an identity is assumed between an experience, such as pain, and a neural mechanism, such the firing of C-fibres, then it makes no sense to ask why pain and the firing of C-fibres are the same feature, just as how it makes no sense to ask why Mark Twain is Samuel Clemens, or to ask why water is H_2O. According to the identity theory, pain simply is the firing of C-fibres and there is no need for further discussion.

The identity theory seems to suggest that the above relations are contingent. The statements "pain is the firing of C-fibres" and "water is H_2O" are not known *a priori*, but are known *a posteriori* through empirical discovery. The fact that water is H_2O is not an analytic fact, but a synthetic fact that had to be discovered by. Likewise, it was also a discovery that the firing of C-fibres results in pain.

An argument against the identity theory is presented by Saul Kripke in *Naming and Necessity* (1980). As Kripke notes, it is true that all identities are necessary, provided that the terms used to pick out the objects or individuals designate rigidly rather than flexibly. A flexible designator is a term which is used to refer not to the referent itself, but to the conditions which are satisfied by the referent in this world. An example of a flexible designator would be, "the highest mountain in the world", which in this world is Mount Everest. However, purported identities which involve flexible designators are contingent. For example, let us consider the statement "Mount Everest is the highest mountain in the world". This fact is contingent, for the term, "the highest mountain in the world", is a flexible designator. Indeed, in this world, Mount Everest is the highest

mountain in the world, but we can conceive of a counterfactual world wherein Langtang Lirung is higher than Mount Everest. In that counterfactual world, Langtang Lirung would be the highest mountain in the world. Mount Everest is the highest mountain in the actual world, because, in the actual world, Mount Everest happens to satisfy the conditions set by the flexible designator "the highest mountain in the world". However, in the aforementioned counterfactual world, it would not be Mount Everest that satisfies these conditions, but Langtang Lirung.

Flexible designators are associated with the descriptivist theory of reference, which was developed by Gottlob Frege (1892). This proposes that when we refer to an object, we are referring to the cluster of descriptions that characterise the object. Consider the example of Mount Everest. The cluster of descriptions in this case is the flexible designator, "the highest mountain in the world".

However, Kripke argues that although the descriptivist theory may be appropriate in cases in which we use flexible designators, it fails with respect to cases which use rigid designators. Instead, Kripke proposes a causal theory of reference. A rigid designator is a term which refers not to the cluster of descriptions which characterise an object, but to the object or individual itself. A proper name is an example of a rigid designator that fixes a referent, or a subject. Accordingly, a proper name is not a mere synonym for a cluster of descriptions. While this cluster of descriptions is contingent, the identity of the referent is necessary.

For example, when I use the proper name "Buddy Bolden", I think of the pioneer of jazz, but the name "Buddy Bolden" is not a mere synonym of the description "pioneer of jazz". As I have said, a description is a flexible designator, and so is contingent. I could think of a possible world wherein the referent "Buddy Bolden" does not fit this description. Perhaps that possible world is a world wherein Buddy Bolden did not become a musician. Rather, the proper name "Buddy Bolden" is a rigid designator that fixes Buddy Bolden as the referent. It identifies a specific individual and this identity is necessary. Indeed, in the actual world, wherein Buddy Bolden did become a musician, this specific individual does fit the description "pioneer of jazz", but there may be a counterfactual world wherein this same individual does not fit that description.

Would "Buddy Bolden" also denote Buddy Bolden in a counterfactual world wherein this particular individual was not given the name "Buddy Bolden", but instead was given a different name? I argue that it would, because the analysis is relative to how the

referent is fixed in the actual world. Regardless of what name the individual is given in other possible worlds, the rigid designator "Buddy Bolden", which fixes Buddy Bolden as the referent in the actual world, denotes that same individual across every world. This is entailed, first, from the thesis of transworld identity which states that the identity of an individual is maintained across worlds and, second, from the fact that a rigid designator exclusively designates a specific individual. If transworld identity is accepted, then the rigid designator "Buddy Bolden", which is fixed in the actual world, would still denote Buddy Bolden in a counterfactual world wherein Buddy Bolden had been given a different name.

Transworld identity captures how we talk about how we might otherwise have possibly been under different circumstances. For example, Kripke considers the statement "Humphrey might have won the election". Here, Humphrey cares about the possibility that he, Humphrey, could have won and not about the possibility that a counterpart could have won. This suggests that David Lewis' (1986) counterpart theory is false with regards to modal claims about how we might otherwise have possibly been. The thesis of transworld identity is true, insofar as such modal claims about how we might otherwise have possibly been are supposed to be about ourselves. There might, in a possible world, be another person, Humphrey*, who is a different person from Humphrey, yet resembles Humphrey. Nonetheless, "Humphrey might have won the election" is not about a possible world where Humphrey* had won, but is about a possible world where Humphrey had won.

Could there be a possible world wherein I have very different properties, such as a different genotype or a different birthdate? I argue that there could. Notably, Kripke claims otherwise, because he assumes the necessity of origin, which suggests that one is identified through the specific pair of gametes from which one originated. However, I propose that the necessity of origin is false with respect to modal claims about how we might have otherwise possibly been, because a given subject can, from a first-person point of view, conceive of the possibility of being associated with a body with a different origin. Given that my individuated first-person subjectivity is what essentially determines the identity of my self, there could be a counterfactual world wherein I have a different genotype or a different birthdate. My transworld identity would be maintained in this scenario, insofar as the same subjective self, namely my consciousness, is associated with my body in the actual world and with the body with the different origin in the counterfactual world.

Thus, a rigid designator fixes a referent and denotes that same referent throughout every possible world, regardless of the cluster of descriptions that the referent satisfies in each possible world. Accordingly, it must be accepted that the necessity of identity is true in the case where the terms that are used to pick out the referent are rigid designators. This can also be derived as follows. Given, first, that it is necessarily true that any given feature is identical with itself and, second, that identity is transitive, it follows that rigid designators that refer to that given feature denote an identical referent across every possible world. Indeed, the suggestion that identity can be contingent is false, because terms that may be coextensive in the actual world but that turn out to have different referents in a counterfactual world are not denoting the same property, which would indicate that the terms are not rigid designators and that the relation between the referents is not identity.

An example of an *a posteriori* necessary identity is the identity between water and H_2O. Both of the terms "water" and "H_2O" are rigid designators, and both refer to the same object. Since they both fix the same referent, water is identical to H_2O. Furthermore, this identity is necessary. Water is necessarily H_2O, for the identity holds in every possible world. According to Kripke, any appearance of contingency is just an illusion. Indeed, the fact that water is H_2O was something that needed to be discovered, and so it seems natural to suggest that water might not have turned out to be H_2O. That is to say, although water turned out to be H_2O in this world, one may speculate that there may be a possible world in which water did not turn out to be H_2O, but instead turned out to be *XYZ*. However, Kripke argues that this is contradictory, because water is H_2O in every possible world. In fact, the substance imagined in this other world is in water at all, but is a substance made from *XYZ* that resembles water. One may call this substance "watery stuff". However, for something to be water, it must be made out of H_2O, and since this "watery stuff" is not made out of H_2O, it cannot be water. Therefore, Kripke suggests that water is necessarily H_2O.

However, with the experience of pain and the firing of C-fibres, Kripke argues, there is no necessary identity, for whereas water is H_2O in every possible world, we can conceive of possible worlds in which the experience of pain is not accompanied by the firing of C-fibres, and, conversely, worlds in which the firing of C-fibres is not accompanied by the experience of pain. Furthermore, since all identities which involve rigid designators must be necessary, it follows that pain and the firing of C-fibres are nonidentical.

This is because the experience of pain and the firing of C-fibres essentially refer to different features. Whereas the firing of C-fibres refers to the neural mechanism that occur when one is stimulated with noxious stimuli, the experience of pain refers to the subjective quality of pain itself. Therefore, it is logically conceivable to dissociate the two from each other. We can conceive of the firing of C-fibres without the subjective quality of a painful experience.

This cannot be done in the case of water and H_2O, for water's molecular structure of H_2O is essential to it. It follows that whereas a substance which behaves like water, but is not made from H_2O, is not water, but "watery stuff", we cannot say that something that feels like pain, but is not accompanied by the firing of C-fibres, is not pain, but "painy stuff". According to Kripke, all it is for something to be pain is for it to feel like pain. Whereas the molecular structure of H_2O is essential to water, what is essential to pain is its subjective quality. Therefore, anything that feels like pain, such as this supposed "painy stuff", is in fact pain, whereas anything which behaves like water but is not composed of H_2O cannot be water.

Having considered this, we can see that the comparison of the situation with Mark Twain and Samuel Clemens with the situation with pain and the firing of C-fibres is a false analogy. Whereas the names "Mark Twain" and "Samuel Clemens" both rigidly designate the same person, the terms "pain" and "the firing of C-fibres" refer to different features. The firing of C-fibres is the neural state that occurs upon stimulation by a noxious stimulus and pain is the subjective experience that accompanies this neural state.

Indeed, we may conceive of a counterfactual world wherein Samuel Clemens had decided not to use the pseudonym "Mark Twain". However, relative to the actual world wherein he did use the pseudonym "Mark Twain", the expressions "Samuel Clemens" and "Mark Twain" rigidly designate the same person. Furthermore, we may conceive of a possible world wherein there was a different person named Mark Twain, who was an entirely separate individual from Samuel Clemens. In this scenario, the rigid designator "Mark Twain*" that denotes this other person is different from the rigid designator "Mark Twain" that denotes the person with whom Samuel Clemens is identical. That is to say, "Mark Twain" and "Mark Twain*" denote different referents. Only the former rigid designator pertains to the identity between Samuel Clemens and Mark Twain, because this is the rigid designator that denotes the person with whom Samuel Clemens is identical. This person with whom Samuel Clemens is identical and who is denoted by the rigid designator

"Mark Twain" is nonidentical with the other person named Mark Twain who is denoted by the rigid designator "Mark Twain*".

Hence, although it was an *a posteriori* discovery that "Mark Twain" and "Samuel Clemens" refer to the same person, this identity is necessary, and, according to Kripke, any appearance of contingency is illusory. However, there is no such identity between pain and the firing of C-fibres, because "pain" and "the firing of C-fibres" refer to different features. Whereas the expression "pain" rigidly designates the phenomenal quality that is essential to pain, the expression "the firing of C-fibres" rigidly designates the neural mechanism that comprises the firing of C-fibres. It is logically conceivable that these two features can come apart. Therefore, the identity theory is false. A subjective experience, such as pain, and a neural mechanism, such as the firing of C-fibres, are nonidentical.

Moreover, I argue that the above failure of identity can be shown to hold even without having to appeal to the necessity of identity. As noted above, Kripke's analysis indicates that the notion of contingent identity is impossible and incoherent, because terms that may be coextensive in the actual world but that turn out to have different referents in a counterfactual world are thereby not denoting the same property. Nonetheless, even if one countenances the notion of contingent identity, it would still be the case that subjective experience is nonidentical with a physical state. This is because the term that denotes the subjective experience and the term that denotes the physical state are not even coextensive in the actual world. For example, I noted above that the essential feature of pain that determines the referent of "pain" is a phenomenal quality, whereas the essential feature of the firing of C-fibres that determines the referent of "the firing of C-fibres" is a neural mechanism. Given that "pain" and "the firing of C-fibres" denote different referents in the actual world, it follows that pain is nonidentical with the firing of C-fibres, regardless of whether identity is necessary.

An objection to the modal argument is raised by David Papineau (2002), who suggests that one's intuition that a phenomenal quality and its corresponding brain state are essentially distinct is merely the result of the particular way in which one thinks about consciousness. According to Papineau, the phenomenal quality and the brain state are just the same item under two different modes of presentation. One can either think of this item physically as a brain state, or experientially as a phenomenal quality. Hence, dualism seems intuitive because one can know an item experientially without this knowledge revealing anything physical about the brain state.

In reply, I argue that Papineau's objection fails for two reasons. First, some of the arguments for dualism do not depend on the intuition that one can know an item experientially without knowing it physically. Instead, the knowledge argument and the explanatory gap argument show that one can have complete physical knowledge without having complete phenomenal knowledge. Hence, these arguments for dualism can still be sound even if the aforementioned dualist intuition can be explained. Second, Papineau's objection suggests that consciousness is epistemically unique, insofar as one can know it under an experiential mode of presentation. However, I argue that the fact that one can know consciousness under an experiential mode of presentation indicates that there is something ontologically unique about consciousness, insofar as its being known experientially entails a first-person ontology. Hence, positing an experiential mode of presentation does not explain consciousness away, but presupposes its unique existence. Importantly, this shows that Papineau's objection does not debunk the aforementioned dualist intuition, but rather it justifies the intuition and supports dualism, because it suggests that phenomenal knowledge requires the existence of a uniquely first-person mode of subjective experience over and above the third-person physical mode of presentation. Therefore, Papineau's claim that a phenomenal quality and a brain state are the same item under different modes of presentation is false.

Given that phenomenal and physical concepts denote different referents, we can conclude that it is true that consciousness is nonidentical with a physical state. In fact, in virtue of the fact that the referent of the rigid designator "consciousness" is determined by the unique first-person ontology that is essential to consciousness, it must be taken as true that consciousness is only identical with itself. Therefore, there remains a meaningful explanatory gap between the occurrence of physicality and the occurrence of phenomenality, for which physical facts cannot account. From this, it follows that physicalism is false. In order to account for consciousness, it must be acknowledged that dualism is true.

The subjectivity argument

While the above arguments focus on the qualitative character of experience, the subjectivity argument focuses on the first-person subjective ontology of experience. As noted by Thomas Nagel (1986), conscious experience does not occur in some neutral third-person "view from nowhere", but is necessarily individuated to a

given first-person subject. That is to say, *my* experience is fundamentally different from *your* experience, because *my* experience has a first-person individuation unique to *me* and *your* experience has a first-person individuation unique to *you*. Given the first-person ontology of consciousness, such individuation is discrete, such that *you* and *I* exist as discretely distinct first-person experiencers with unique ipseities that are essentially different. Accordingly, the identity or haecceity of a given consciousness is essentially determined by its unique first-person individuation.

By contrast, physical facts are third-person facts about the objective world. They occur in a neutral third-person space. Accordingly, physical events and processes can be described and explained in third-person structural and dynamical terms.

Insofar as they are exclusively third-person facts, physical facts fail to account for the first-person individuation of consciousness. Given all the third-person physical facts about the structure and dynamics of the objective world, the fact about the first-person individuation of consciousness remains a further fact to consider. That is to say, the third-person physical facts do not account for why conscious experiences are individuated to different first-person subjective viewpoints. For example, I, as a first-person subject, happen to have an experiential viewpoint associated with *this* body rather than with *that* body, but is just as consistent with the physical facts about the bodies that my experiential viewpoint could have been associated with *that* body rather than with *this* body. Indeed, the third-person physical facts fail to account for why this body is associated with experience that is individuated to a first-person subject at all.

Therefore, the first-person individuation of consciousness is a further fact that is separate from the third-person physical facts. The complete third-person physical facts about our bodies may yield information about their spatial locations and their causal organisations, but such third-person physical facts fail to explain why *my* consciousness accompanies the body which is at *this* spatial location rather than the body which is at *that* spatial location and why *your* consciousness accompanies the body which is at *that* spatial location rather than the body which is at *this* spatial location. That is to say, when all of the third-person physical facts about *this* body and *that* body are given, the fact the experience associated with *this* body is individuated to *me* rather than to you and that fact that the experience associated with *that* body is individuated to *you* rather than to me remain further facts to consider.

The above also highlights the problem with David Hume's (1740) suggestion that experience involves a bundle of perceptions without any perceiver of that bundle of perceptions. Such a view is problematic because it seems to portray experiential qualities as impersonal events in a neutral space. However, as noted above, experiential qualities are not impersonal events in a neutral space, but are individuated to specific first-person subjective viewpoints. And so, Hume's bundle theory is false, because it fails to account for the first-person individuation of consciousness. That is to say, it fails to account for why *this* cluster of perceptions is experienced by *me* and why *that* cluster of perceptions is experienced by *you.*

It might be objected that the above could be resolved by appealing to indexicality, but I argue that this objection fails, because first-person individuation is different from mere indexicality. Indexicality is a contextual fact about how a speaker is centred in the world, whereas first-person individuation is a substantive ontological fact about the existence of a distinct and discrete experiential subject. While it may be true that any fact about subjectivity is an essentially indexical fact, indexicality on its own does not entail subjectivity. Rather, the presence of consciousness is a further fact beyond indexicality. For example, John Perry (1979) uses the example of the proposition, "I am making a mess", which is an indexical proposition. This indexical proposition centres Perry as the speaker, and so in this case there happens to be a conscious subject, namely Perry, associated with that centre. However, it is also conceivable that a nonconscious system could utter an indexical proposition such as "I am making a mess". In such a case, the indexical proposition centres the nonconscious system as the source of the utterance, but there is no conscious subject associated with that centre. Therefore, the suggestion that the first-person individuation of consciousness can be reduced to indexicality is false. Given all the indexical facts about how an utterance is centred, whether that centre is associated with consciousness remains a further fact beyond the indexical facts.

Dualism verified

The arguments considered above all involve the notion of logical supervenience. Physical facts about structure and dynamics only yield further structural and dynamical facts, but they do not entail the phenomenal fact about the subjective character of consciousness. Therefore, consciousness is not logically supervenient on the

physical. That is to say, the existence of consciousness remains a further fact beyond the physical facts.

The conceivability argument shows that given the complete physical facts about a system, the presence of consciousness is still an extra fact to consider. Nothing in the structural and dynamical properties of an organism's nervous system entails the presence of conscious experience. Thus, while organisms are actually conscious, there is no logical contradiction in conceiving of a realisation of the organism's organisation that lacks experiences entirely. The organism's activity can be explained fully in physical terms of structure and dynamics, and so the presence of consciousness is a further fact over and above these physical properties.

Similarly, the knowledge argument and the explanatory gap argument show that facts about conscious experiences are further facts over and above the physical facts. These facts pertain to the qualitative characters of experiences. Nothing in the structural and dynamical properties of a brain state can tell us why it is accompanied by a particular experience only or why that particular experience feels the way it does rather than like something else. As noted by Chalmers (1996), structural and dynamical facts only yield further structural and dynamical facts. They cannot account for the qualitative character of experience. From this, the knowledge argument states that one cannot know what an experience is like solely from physical knowledge, while the explanatory gap argument states that even if we know what an experience is like, any relation between it and the physical facts about the accompanying neural mechanism is contingent and inexplicable.

Although the subjectivity argument focuses on the first-person individuation of consciousness rather than on the qualitative character of experience, it also demonstrates the failure of logical supervenience. The third-person facts about the physical world do not entail the first-person fact about the subjectivity of consciousness. Hence, the first-person subjectivity of consciousness does not logically supervene on the third-person facts about the physical world.

The first implication of the failure of logical supervenience, as noted in chapter three, is epistemic. The physical facts do not entail facts about consciousness, and so consciousness cannot be reductively explained by science. By contrast, many of the features we experience in the universe are all logically supervenient on the physical, and so can be explained in physical terms. They all share the common physical parameters of structure and dynamics, and

these can be reduced further to even more basic structural and dynamical facts, and so on. It follows from this that the lower-level physical facts determine the higher-level facts, and so the higher-level facts can be reductively explained in terms of the lower-level facts. However, this is not the case with consciousness. The failure of logical supervenience implies that the existence of consciousness is not entailed by any physical facts, including the most basic ones. Structures and dynamics only yield further structures and dynamics. They do not capture conscious experience. Therefore, consciousness is irreducible. It is an extra fact beyond the physical facts, and so any attempt to explain it in terms of physical facts would fail.

The second implication of the failure of logical supervenience is ontological. The physical facts do not entail the existence of consciousness, and so consciousness is an extra fact over and above them. This failure of logical supervenience entails that physicalism is false. Given the complete physical facts about the world, the existence of consciousness is still a further fact to consider. And so, it must be taken as true that consciousness is a nonphysical entity that is ontologically separate from physical matter. Given the existence of consciousness and its nonentailment from the physical facts, dualism is necessarily true.

What sort of dualism am I advocating? As I have stated, any acceptable account of consciousness must acknowledge consciousness for what it is and must not conflate it with other features. Psychological features of the mind, such as awareness, perception, and reportability, are structural and dynamical processes, and so there is no reason why these cannot be fully explained in physical terms. Consciousness, however, refers to the phenomenon of first-person subjective experience. It is this for which physicalism cannot account, and so it is this which is relevant to the dualism I am advocating. Furthermore, as noted in chapter one, what is essential to consciousness is its irreducible first-person ontology, whereas the objective world has a third-person ontology. Subjective experience and the objective world are of fundamentally different kinds, and so cannot be identified with, reduced to, or entailed from each other. The dualism I am advocating, therefore, is the philosophical thesis that it is true that the first-person subjective existence that is consciousness is an ontologically separate entity from the third-person objective world.

The position I advocate could perhaps be considered a form of idealistic dualism. In addition to acknowledging that consciousness is an ontologically separate entity from the objective world, it

acknowledges that consciousness is epistemically foundational. As noted in chapter one, consciousness is first-person subjective existence. It is essentially what it is to be. Only when experienced as experience in consciousness does the objective world becomes like something. On its own, it is not like anything, but is a potential that is only realised when it manifests as experience in consciousness.

Therefore, the nonentailment from physicality to phenomenality proves that dualism is true. Consciousness exists as a *sui generis* fundamental entity that is ontologically separate from physical matter. I shall briefly expound this thesis further in chapter five, but now I consider some objections and show that these objections fail to undermine dualism.

Objections and replies

A question that is commonly asked about dualism is how it can explain the interaction between mind and matter. I have noted that subjective experiences are associated with certain physical processes, namely brain states. Furthermore, different subjective qualities are associated with different brain states, such as the experience of pain with the firing of C-fibres and visual qualia with activity in the visual cortex. Therefore, there is a robust correlation between experiences and physical processes. How can dualism explain this correlation? If consciousness and the physical world are ontologically distinct from each other, how do they interact? In reply to this, the dualist can accept that although consciousness is ontologically separate from the physical world, its experiences are, to some degree, nomologically related with events in the physical world. This position assumes a contingent association without any metaphysical necessity. Consciousness is not logically supervenient on the physical, but its contents are correlated with certain physical events in a lawlike yet contingent manner. In chapter seven, I shall be more specific and defend Chalmers' (1996) suggestion that there are contingent psychophysical laws that correlate subjective experiences with certain physical processes.

Of course, dualism is a philosophical position that is most famously associated with René Descartes (1641). While the dualist view which I am advocating accords with Descartes' view with respect to the ontological distinction between the physical and the mental, the aforementioned nomological correlation between the physical and the mental can be seen as an advancement on

Descartes' view. Notably, Descartes suggested that mind causally interacts with matter. However, it has been suggested that this runs into problems.

First, it has been suggested that Descartes' picture violates the causal closure of the physical world as postulated by science. If the mind has a causal influence on physical processes, then it would interfere with basic physical laws. To some, this is an unacceptable consequence, since it undermines the integrity of our scientific knowledge about the physical world we experience.

This objection can be overcome by denying the causal closure of the physical world. Indeed, as I shall argue in chapter nine, a scientifically informed nondeterministic view of the universe and a regularistic approach to the laws of nature suggest that the world is not causally closed. Nonetheless, for those who remain concerned about causal closure, the view I am advocating may seem more acceptable than Descartes' view. A nomological relation between the physical and the phenomenal poses no problem for causal closure, insofar as it does not interfere causally with physical processes, but merely occasions correlations between these processes and phenomenal qualities. Thus, the assumption of causal closure can be maintained and consciousness remains an extra fact.

Second, it has been suggested that Descartes' picture does not specify a mechanism through which the nonphysical mind acts on physical matter. To some, the idea of something nonphysical having a causal influence on something physical seems to makes little sense. The suggestion here is that for something to have a causal influence on the structure and dynamics of physical processes, it itself must also possess structure and dynamics, and so it must also be physical.

The objection can be overcome by endorsing a regularity view of causation, such as that suggested by David Hume (1748). Consider that *C* causes *E*. According to the regularity view, causation is merely the instantiation of a regular, yet contingent, association between *C* and *E*. There is no need to posit any powers or mechanisms underpinning the regularity between *C* and *E*. Hence, under such a view, mental-physical causation is no more problematic than physical-physical causation. All that needs to obtain for a nonphysical mind to exert a causal influence on physical matter is for there to be a regular association between them. No mechanism needs to be posited. Nonetheless, for those who remain concerned about the mechanism of causation, the view I am advocating may seem more acceptable than Descartes' view. The dualism I am proposing does not necessarily assume a causal interaction between mind and

matter, but a robust correlation between the phenomenal and the physical. This correlation obtains in virtue of a nomological relation, but there is no need to assume that the correlation that ensues amounts to causation.

At this point, one might ask if such a position I am defending can, in fact, be called dualism at all. Specifically, one might suggest that my acceptance of a correlation between physical processes and subjective experiences suggests that I am, in fact, proposing a weaker form of physicalism, but I argue that this is mistaken. Such an objection is made by John Searle (1992), who holds that the experiences of consciousness are correlated with, but not logically supervenient on, the physical, but denies that his position is a version of dualism. After all, how can one claim to be a dualist if one accepts that subjective experiences are correlated with physical processes?

I argue that this terminological objection is unsound, because it involves a semantic error that arises from a misunderstanding of what a dualist position entails. Physicalism claims that subjective experience has a physical ontology, whereas dualism claims that it is ontologically different from the physical. This captures a philosophically significant distinction between the two positions. Indeed, throughout this chapter, I have aimed to refute the physicalist position as it is defined above by arguing that consciousness is an ontologically separate entity from the physical world. Therefore, it is true that what I am affirming in this book is necessarily a dualist philosophy, because it acknowledges this ontological distinction between the phenomenal and the physical. Although I accept that there is a correlation between physical processes and subjective experiences, it is the acceptance of this ontological difference between the phenomenal and the physical that defines my position as a dualist one. One may try to assume a different interpretation of physicalism so that it includes natural supervenience without logical supervenience, but this would amount to a false definition of physicalism. Simply giving something a different label does not change the relevant conceptual distinction. In fact, to extend the definition of physicalism would be unhelpful and unjustified, since it would neglect the relevant conceptual distinction on which the established definitions of physicalism and dualism are based. That is to say, it would nullify the concept of physicalism by dissolving the distinction between it and other positions in the philosophy of mind. Hence, to try to claim that the position defended herein is not dualist but physicalist would be to assume a false definition of physicalism. Under a true definition of dualism that captures the relevant

philosophical distinction, the philosophical position I am affirming herein is truly dualist in virtue of the fact that it acknowledges that consciousness is ontologically separate from the physical world.

A common objection to the dualist position is the claim that it is inconsistent with science. Such a view is held by Patricia Churchland (1988). Her first objection is that the acceptance of dualism would be to reject the principles postulated by evolutionary biology, modern physics, and chemistry. A reply to this objection is offered by Chalmers (1996), who argues that Churchland's claim is false. Even if the interactionism of Descartes undermines the scientific principle of causal closure, nothing in the dualism advocated herein suggests that our scientific knowledge should be undermined. As I argued earlier, in the dualism I am proposing, subjective experience does not interfere causally with the physical world, but exists as a further fact. Any assumed causal closure is conserved and our scientific theories about the physical world are not undermined.

To make her case, Churchland appeals to results from scientific research which she considers to provide evidence against dualism. First, she appeals to the success of neuroscience in explaining the structure and activity of the human nervous system. She argues that neuroscience can provide a perfectly adequate explanation of human behaviour in physical terms, without having to appeal to a nonphysical mind, and so she argues that we have no need to resort to a dualist position. Second, she appeals to the advances in computer science, and argues that complex cognitive processes can be performed by a machine without a nonphysical mind. From this, she argues, once again, that a dualist position is unnecessary.

In reply, I argue that Churchland has misunderstood what dualism actually entails and has also erroneously conflated consciousness with various psychological capacities. As I have already mentioned, a dualist philosopher who takes science seriously could accept that psychological capacities can be explained in physical terms. However, the fact remains that over and above this physical picture, the existence of consciousness itself is a further feature for which the physical facts cannot account. Thus, the evidence Churchland fails to undermine dualism. In fact, if anything, the evidence provided by Churchland can even be interpreted as providing support for dualism. The suggestion that cognitive processes can be realised without any appeal to a nonphysical mind, for example, is reminiscent of the conceivability argument. Although these cognitive processes can be realised and explained entirely in physical terms, there is still the issue of whether these processes are accompanied by subjective

experience or not. This is an open question, for it is just as logically conceivable for a cognitive system to be nonconscious as it is for it to be conscious. The existence of consciousness is not logically entailed by the structural and dynamical facts, and so it is an extra fact over and above the physical world.

A related objection, raised by Daniel Dennett (1991), is that the acceptance of dualism would be like "giving up" on explanation altogether. He suggests that to endorse a dualist position would be to disregard the possibility of a future physical account. Moreover, he claims that dualism "wallows in mystery" and that to accept it would be to abandon hope of further knowledge.

My reply to this is twofold. First, the failure of logical supervenience between the physical and the phenomenal indicates that a physical explanation of consciousness is, in fact, impossible. Therefore, to understand consciousness, it is essential that we move away from physical explanation. However, this is not to be interpreted as "giving up". Rather, it is the realisation that physical accounts necessarily fail, and so we are required to look elsewhere for an account. Second, all explanation stops somewhere. As noted by Chalmers (1996), even the physical sciences postulate a set of basic physical properties, such as mass, spin, and charge. These physical properties are taken as basic, and so there is no attempt to reduce them any further. They are properties that cannot be explained by one another within a theoretical framework, but can be said to be related by a set of given laws. Although the analogy is only partial, it nonetheless shows that accepting consciousness as a fundamental and irreducible entity is not "giving up". Indeed, consciousness is even more ontologically basic than the aforementioned physical properties of mass, spin, and charge, insofar as it is a further fundamental fact that exists beyond the structural and dynamical facts that pertain to these physical properties. It is a unique phenomenon in virtue of its essential first-person ontology, and so it cannot be reductively explained in terms of third-person properties. It is the fundamental fact of first-person subjective existence and is ontologically novel.

Another objection, raised by Colin McGinn (1989) and Robert van Gulick (1993), is that the explanatory gap between the physical and phenomenal arises due to a cognitive limitation. This suggests a form of mysterianism, according to which there may be an *a priori* conceptual implication from physicality to phenomenality of which one simply is incapable of conceiving due to a cognitive limitation, but I argue that this view is mistaken. To begin, we could ask what

sort of linking concept this would be. If it is a structural and dynamical concept, then we are confronted with the same problems as before. Structures and dynamics only yield further structures and dynamics, but they do not say anything about the subjective quality of experience. If it is not a structural or dynamical concept, then it cannot be logically entailed by structural and dynamical facts, since structural and dynamical facts only yield further structural and dynamical facts, but no more. Thus, the claim that there is an *a priori* link between physicality and phenomenality is false.

The explanatory gap between the physical and phenomenal has been suggested to be analogous to the inability to grasp certain mathematical theorems and the inability of an armadillo to grasp quantum mechanics, but I argue, again, that this is mistaken. Indeed, a cognitive limitation is the reason why an armadillo cannot grasp quantum mechanics. An armadillo does not have the appropriate capacities to process the theory. Similarly, the inability to grasp certain mathematical theorems may be due to a constraint set by the way in which the human cognitive system is embodied in the world. For example, we can conceive of beings whose cognitive capacities differ from ours in ways that enable them to grasp such mathematical theorems, much like how we can grasp some of the intricacies of quantum mechanics that an armadillo cannot grasp.

The analogy with the explanatory gap, however, is erroneous. Although facts about quantum mechanics or certain mathematical theorems are beyond the cognitive capacity of certain beings, there is no reason to believe that there is anything ontologically novel about them. Facts about quantum mechanics, despite their apparent strangeness, are still structural and dynamical facts. Similarly, certain mathematical theorems, although they cannot be grasped, are still only mathematical theorems, albeit more complex than the mathematical theorems that have been grasped. Thus, these facts may be facts about the objective world that are on a level of complexity than cannot be grasped by some beings, but they are still, nevertheless, facts about the objective world that have a third-person ontology. A cognitive system's ability to grasp them is contingent on the particular way in which that system is embodied in the world.

Consciousness, however, does not fall into this kind. Unlike quantum mechanical and mathematical facts, there is a good reason to believe that consciousness is ontologically novel. This is due to its first-person subjectivity. Consciousness is unique because, unlike quantum mechanical and mathematical facts, it is not a fact about the third-person objective world, but is the fact of first-person subjective

existence. Thus, the supposed analogy between the grasp of consciousness and the grasp of quantum mechanical or mathematical facts is false. In virtue of its first-person subjectivity, consciousness is ontologically unique in a way that quantum mechanical or mathematical facts are not.

Given this first-person subjectivity, one's access to consciousness is fundamentally different from one's access to quantum mechanical or mathematical facts. Since quantum mechanical and mathematical facts are objective, one's access to them is contingent on how one experiences the world. That is to say, the way in which one's cognitive system is embodied in the world will influence one's ability to grasp these quantum mechanical and mathematical facts. However, one's grasp of consciousness is certain, regardless of how one's cognitive apparatus is shaped, because consciousness is one's very first-person existence which one knows through acquaintance. Again, one's grasp of quantum mechanical or mathematical facts is disanalogous with one's knowledge of one's consciousness.

This indicates that mysterianism is false with respect to consciousness. Given the that consciousness is the first-person existence with which one is acquainted, it is true that knowledge of consciousness is foundational. Accordingly, the appeal to cognitive limitation is irrelevant with regard to consciousness, because consciousness is not grasped through cognising about the external world, but is grasped through direct acquaintance. Knowledge of consciousness marks the point at which skepticism is false.

There is similar objection to the above in the form of a debunking argument that claims that one's belief in dualism is a result of the structure of one's cognitive system. This form of argument has often been used to attempt to debunk certain spiritual beliefs. For example, Richard Dawkins (1976) attempts to undermine such beliefs by suggesting they are mere memes that perpetuate themselves. Also, Susan Blackmore (1982) attempts to demystify out-of-body experiences by explaining them as the products of certain neural processes. It is easy to see how this argument can be applied to the belief in dualism. As noted earlier, David Papineau (2002) suggests that dualism seems intuitive because of how one thinks about consciousness. Going further, Daniel Dennett (1991) suggests that consciousness involves a user illusion and that one's belief in it is the result of one's brain working in such a way that produces such a belief. However, I argue that this fails to discredit dualism.

The problem with the above line of argument is that it commits the genetic fallacy. This is the mistaken assumption that the validity

of a belief can be discredited by an explanation of its generation. I argue that it cannot. An explanation of a belief's generation does not discredit the belief's truth or the belief's justification. For example, imagine that I have the belief, "there is a dessert in the refrigerator", and that the generation of this belief was influenced by wishful thinking in the context of hunger. Indeed, there actually is a dessert in the refrigerator, and so my belief is true, regardless of how fanciful its generation is. Likewise, an explanation of how my cognitive system produced my belief in dualism does not provide any information about the truth value of the belief itself. Such explanations only elucidate the mechanisms behind the acquisition of a belief, but they do not undermine the truth or justifiability of the belief's content. Thus, regardless of how it was acquired, my belief in dualism can still be true. Similarly, Elliott Sober (1994) notes that some of our beliefs about the world might be explained by socialisation and adaptation, but this neither shows that these beliefs about the world are not true, nor shows that socialisation and adaptation are not truth conducive. If socialisation and adaptation are truth conducive, then the beliefs may be true and justified. In such a scenario, the debunking argument's conclusion is false.

Ideally, perhaps, the justification of a belief would have a role in the explanation of its generation. For example, one might find it more sensible if I had acquired my belief "there is a dessert in the refrigerator" from the fact that I have recently looked inside the refrigerator and have seen a dessert inside it, rather than from wishful thinking in the context of hunger, despite the fact that the same true belief is produced in both of these scenarios. In the former scenario, the justification of my belief, specifically my having perceived that there is a dessert in the refrigerator, is the reason why I believe that there is a dessert in the refrigerator. In the latter scenario, the reason why I believe that there is a dessert in the refrigerator, specifically my wishful thinking, does not involve that justification from perception. Since both instances produce the same true belief, the truth value of the belief cannot be the reason for favouring one instance over another. Rather, I propose that the former instance is favoured over the latter due to other considerations. In the former instance, the explanation serves as a justification, whereas in the latter instance, the explanation does not serve as a justification, and so further justification is required.

With respect to the belief in dualism, I argue that the belief does, indeed, have a justification and, moreover, that the manner in which it is justified is relevant to the explanation of the belief's presence.

Specifically, I propose that my belief about consciousness is justified by my direct acquaintance with the subjectivity of consciousness. From this direct acquaintance, I know consciousness to be of an entirely different kind from the objective world. It has a first-person ontology rather than a third-person ontology and its subjectivity is not captured by physical facts about structures and dynamics. I suggest that this first-person acquaintance with consciousness, which justifies my belief in dualism, could also have a role, albeit a noncausal role, in the explanation of my belief in dualism. That is to say, it is in virtue of my first-person acquaintance with consciousness that my philosophical knowledge of the truth of dualism is sound.

Importantly, such first-person acquaintance with consciousness is not a mere intuition, but is the necessary foundation for the very discernment of knowledge. It is ontologically foundational, because my consciousness is the first-person existence that I *am*. It is epistemically foundational, because the discernment of what is sound and what is unsound is only done through consciousness. Thus, it is true that my first-person acquaintance with consciousness is sound, because the soundness of my first-person acquaintance with consciousness is necessary for the discernment of what can be sound and what can be unsound. This reflexivity reveals a core truth about consciousness, namely that it is a fundamental phenomenon that is only understood through itself. I know that consciousness exists by being a conscious subject. Indeed, this first-person acquaintance with consciousness is more fundamental than any intuition, because an intuition involves such discernment, and so presupposes the prior first-person acquaintance with consciousness through which such discernment occurs. Therefore, the claim that phenomenal knowledge is based on an intuition is false. Rather, it is necessarily true that my knowledge of the existence of consciousness is sound is in virtue of my first-person acquaintance with consciousness.

Of course, I do not deny that the psychological aspect of my belief was influenced by how my cognitive system is structured. However, this explanation of my belief's formation is a structural and dynamical explanation, and so it only accounts for the origin of my belief's structural and dynamical aspect. It cannot, and does not, provide any information about the subjective content of my belief, which is not based on structure and dynamics, but on first-person acquaintance. I propose that this subjective content of my belief, which cannot be accounted for by a physical explanation, is accounted for and justified by my first-person acquaintance with my consciousness. That is to say, the fact that I have first-person

acquaintance with consciousness is evidence for my belief about consciousness. Therefore, not only does a structural and dynamical explanation of my belief in dualism fail to undermine the possibility of the belief's truth, but it fails to account for the key datum that forms the content of my belief, namely my first-person acquaintance with consciousness, which also justifies my belief.

In response to Dennett's (1991) particular claim regarding experience as a user illusion, I argue that this claim is false for two reasons. First, as noted previously, Dennett's illusionist eliminativism is necessarily false, because an illusion is itself a kind of conscious experience, and so it necessarily presupposes the existence of consciousness. That is to say, the very existence of consciousness is necessary for the discernment of what is real and what is illusory. Second, Dennett's claim is empirically false. If, as Dennett suggests, there was not any phenomenality but there was only the cognitive disposition to believe in phenomenality, then there would only be a third-person neutral space, because a cognitive disposition to believe is a structural and dynamical property that can be analysed exclusively in the third-person. However, a mental state is not an impersonal event in some third-person neutral space, but is individuated to a first-person subjective viewpoint. Indeed, the existence of this first-person subjective viewpoint is given by my acquaintance with it. Importantly, this first-person viewpoint is different from the third-person objective space. For example, my first-person acquaintance with an experiential state is different from a neutral third-person description of my cognitive disposition to form a belief about such a state, for such a third-person description does not capture the distinctive first-person character of that acquaintance. What is wrong with Dennett's claim is that it fails to account for the fact that there is such a first-person subjective viewpoint which is different from the third-person objective space. Therefore, illusionist eliminativism is false, because it fails to account for this first-person datum and its difference from the third-person facts.

A final objection to dualism I shall reply to is that based on simplicity. This objection appeals to the principle of the principle of parsimony, which states that in the formulation of a theory, one should not multiply the number of properties beyond what is necessary to explain the data. Simpler theories are favoured over complex theories. It might be suggested that because physicalism only postulates one class of properties, specifically physical properties, it is simpler than dualism, which postulates two classes, namely mental properties and physical properties.

However, Chalmers (1996) notes that while the principle of parsimony states that one should not multiply properties unnecessarily, there is a necessity in the case of consciousness. The structural and dynamical facts posited by physicalism fail to capture the subjective quality of experience, and so the acknowledgment of a further fact is necessary. Indeed, while simplicity is consideration in theory selection, a more important consideration is the theory's ability to account successfully for the datum. Physicalism fails to account for the datum of consciousness, and so is false. Therefore, the principle of parsimony fails to undermine dualism, because the empirical inadequacy of physicalism regarding consciousness makes dualism the simplest theory that is empirically adequate.

Furthermore, although simplicity is an important virtue in theory selection, it is not necessarily truth conducive. As noted by Bas van Fraassen in *The Scientific Image* (1980), the assumption that simpler theories are more likely to be true is unjustified because it is based on the unwarranted assumption that the truth is simple. Rather, van Fraassen argues that simplicity is a pragmatic virtue. When presented with empirically equivalent theories, the scientist would choose the simpler theory because it has more utility. For example, a simpler theory is easier to understand, easier to communicate, and easier to apply. However, the theory's simplicity does not make the theory more likely to be true. Accordingly, the principle of parsimony does not undermine the truth of dualism, because simplicity is not truth conducive. Sometimes, the principle of parsimony may be false.

And so, in this section I have considered various objections and have shown that these objections fail to undermine dualism. Although these objections do not undermine dualism, they do reveal much about contemporary attitudes towards dualism. With respect to Patricia Churchland's (1988) objection based on consistency with science, this appears to be an objection to the interactionist claim that the mind has a causal influence on matter, rather than an objection to the ontological distinction between consciousness and the physical world. The dualism I am advocating here entails the ontological distinction between consciousness and the physical world, without any commitment to the interactionist claim that the mind has a causal influence on matter. Thus, dualism has been misunderstood and interpreted as something is not. The objection raised by Dennett (1991) regarding the apparent mystery of dualism reflects a reluctance to give up the physicalist worldview. However, Chalmers (1996) suggests that this reluctance reflects nothing more than unjustified "contemporary dogma". The fact that physicalism fails to

account for consciousness indicates that this "contemporary dogma" is unsound. Given the existence of consciousness and its nonentailment from the physical facts, physicalism is necessarily false. To acknowledge consciousness for what it truly is, we must accept that dualism is true.

Indeed, dualism has an established history of being acknowledged as a legitimate and respectable philosophical thesis. In addition to René Descartes (1641) in the early modern world and David Chalmers (1996) in the contemporary world, dualism was accepted in the ancient world by Plato (c. 360 BCE), as well as by the eastern traditions of Sāṃkhya philosophy and Jaina philosophy. These latter two philosophical traditions are respectively codified in the *Sāṅkhya Kārikā* (c. 350) of Iśvarakṛṣṇa and the *Tattvārtha Sūtra* (c. 100–400) of Umāsvāti. Notably, the ontological distinction between the first-person subjective existence of consciousness and the third-person objectivity of physical matter can be taken to correspond broadly to the distinction between *puruṣa* and *prakṛti* in Sāṃkhya philosophy and to the distinction between *jīva* and *ajīva* in Jaina philosophy. The sorts of dualism proposed by Sāṃkhya philosophy and Jaina philosophy are also nontheistic, and so in that respect they accord with the naturalistic dualism of Chalmers (1996), as well as with the form of dualism I am proposing. And so, the dualist philosophy I am affirming in this book can be acknowledged as philosophically respectable in light of the wider dualist program in philosophy with an established history wherein the work is situated.

Importantly, accepting the truth of dualism does not undermine our current scientific theories, but merely requires us to acknowledge that the phenomenon of consciousness is beyond these theories. In *Mind and Matter* (1958), Erwin Schrödinger observed that our entire scientific knowledge "rests entirely on immediate sense perception". That is to say, our theories are based on our observations of the world and these observations are derived from our experiences. The dualism I am advocating proposes that our subjective experiences cannot be explained by our theories, because our theories are derived from our subjective experiences. Hence, the claim that dualism amounts to science denialism is false. Because consciousness is beyond science, my dualist position cannot be undermined by science, nor can it undermine science. Under a dualist framework, our scientific theories about the physical world can be fully accepted. Moreover, by acknowledging its existence as fundamental, we can, under a dualist framework, take consciousness seriously as a foundation for philosophical enquiry.

V
A Philosophy of Consciousness

Consciousness as a fundamental entity

In this brief chapter, I present some key principles of the dualist philosophy of consciousness I am proposing. These follow from the issues discussed in previous chapters and have implications that I develop in later chapters. What I present herein comprises the core of the philosophical system I am advocating, which I take to correspond to an absolute truth about the nature of consciousness.

First, it is true that consciousness exists as a fundamental entity. My consciousness is my first-person subjective existence. Insofar as it is my existence, it is foundational. Accordingly, eliminativism is false with regard to consciousness, because the prior existence of consciousness is necessary for the very discernment of what exists. Thus, realism about consciousness is necessarily true.

Given its subjectivity, consciousness cannot be explained by physical facts about structure and dynamics. Therefore, physicalism is false. Indeed, any form of monism is false with regard to consciousness, because third-person facts about the objective world fail to account for the first-person subjectivity of consciousness. Thus, it must be taken as true that consciousness exists independently as an ungrounded entity.

In view of the above, it is true that consciousness is a separate entity from physical matter. Therefore, dualism is true. Indeed, in virtue of its first-person ontology, it is true that consciousness is a *sui generis* entity that exists separately from the third-person objective world. Specifically, consciousness is the first-person experiencer of the third-person objective world. Accordingly, dualism is necessarily true, because the necessary first-person existence of consciousness is a further fact beyond the third-person facts about the objective world.

The first-person ontology of consciousness entails that it is necessarily integral to selfhood, insofar as selfhood pertains to first-person identity. Thus, it is true that my consciousness is my self. My consciousness is the "I" that I *am*.

Insofar as consciousness is a basic first-person existence, it is true that each consciousness is mereologically simple. Consciousness is not comprised of qualities, but is the pure first-person existence wherein qualities manifest. Hence, the suggestion that consciousness

is composed of constituents is false. The presence of consciousness is an all-or-none phenomenon and is not a matter of degree.

It is also true that each consciousness exists as a discrete unit. The identity of a given consciousness is essentially determined by its unique first-person individuation. Thus, haecceitism is true with respect to consciousness. In virtue of its first-person individuation, it is true that each consciousness is essentially unique.

The eternity of consciousness

Second, it is true that consciousness exists eternally. Given that consciousness exists separately from the objective world, it follows that consciousness is unconditioned by the laws that obtain in the objective world. These laws do not pertain to consciousness, because consciousness exists beyond their domain. Rather, the workings of these laws in the objective world are experienced by consciousness.

Such laws include those involving the dimensions of space and time, which are formal features of the physical world. Objects we experience in this physical world have structures that manifest in space and dynamics that manifest in time. Consciousness, however, is separate from the physical world, and so is unconditioned by space and time. Hence, the doctrine of impermanence is false with respect to consciousness. Because consciousness is unconditioned by time, it is necessarily true that consciousness cannot be annihilated. Space and time are not properties of consciousness, but are properties of the world that is experienced by the timeless existence of consciousness.

Because it is unconditioned by time, it is true that consciousness is eternal. This does not mean that it persists for an indefinite length of time, but rather it means that consciousness exists beyond time. Given all the spatial and temporal facts about the physical world, the existence of consciousness remains a further fact beyond these spatial and temporal facts. Therefore, it is necessarily true that consciousness has no start. Likewise, it is necessarily true that consciousness has no end. Temporal notions of generation, change, and annihilation do not pertain to consciousness, for consciousness is outside time. Space and time do not govern consciousness, and so it is necessarily true that consciousness cannot change.

It is sometimes questioned when in our natural history consciousness originated, but I argue that this question is misguided, because it falsely conflates consciousness with a psychological property. It makes sense to ask when certain psychological capacities

originated, such as attention and introspection, for these are structural and dynamical features whose histories have been shaped by socialisation and evolution. However, to question the origin of consciousness is to make a category mistake, because consciousness has no origin. Given its timelessness, it is true that consciousness is not generated. This also indicates that classical theism is false. A singular creator god does not exist, because consciousnesses exist eternally, and so are not created. Thus, nontheism is true, because it is required to account for the timelessness of consciousness.

The eternity of consciousness can also be proved by appealing to existence. It is true that exists necessarily exists, for existence is *what is*, which exists by definition. Likewise, it is true that nothingness necessarily does not exist, for nothingness is nonexistence or *what is not*, which does not exist by definition. My consciousness is my first-person existence. It is, to me, what it is to exist. Therefore, is true that consciousness exists necessarily, because consciousness is what it is to exist, which exists by definition.

This underpins a transcendental argument. Given that my consciousness is my first-person existence, it is foundational to me. Its existence is necessary for the discernment of what exists and what does not, as this discernment is only done through consciousness. The claim that consciousness does not exist is necessarily false, because its nonexistence would preclude such discernment and negate the very possibility of its nonexistence. Therefore, ontological nihilism is necessarily false regarding consciousness.

This also underpins a modal argument. My consciousness is my first-person existence, and so its nonexistence is impossible to me. To me, there does not exist a possible scenario wherein it does not exist. Every possible scenario for what exists and what does not presupposes the existence of consciousness as a necessary condition for the very discernment of what exists and what does not. Therefore, ontological eternalism is necessarily true regarding consciousness.

The infinite plurality of consciousnesses

Third, it is true that there exist an infinite plurality of consciousnesses. Experience is what realises the world, and so my consciousness is my subjective realisation of existence. This existence is absolute, for it is *what is*. There are infinite ways to realise existence, of which my consciousness is one. Existence is experienced from a first-person subjective viewpoint and there are

infinite potential subjective viewpoints from which existence can be experienced. Indeed, the suggestion that there are a finite number of viewpoints is false, because any given point has infinite angles from which it can be approached. Thus, there are infinite consciousnesses, each of which is a different subjective realisation of existence.

The fact that there exist infinite consciousnesses is necessary for the very discernment of my consciousness as one specific subjective viewpoint out of infinite potential distinct subjective viewpoints. That is to say, intersubjectivity is necessary for subjectivity. The fact that other consciousnesses exist is necessary for the very discernment of *my* consciousness and *other* consciousnesses.

The above can also be couched analytically. In virtue of its first-person subjectivity, any fact about my consciousness is essentially indexical. An indexical is relational, insofar as it specifies its referent in contrast to other tokens which exist. For example, "this" presupposes there is a "that", while "here" presupposes there is a "there". Likewise, *my* consciousness entails that there exist *other* consciousnesses that are not mine. Therefore, solipsism is false.

In virtue of its first-person individuation, each consciousness from the infinite plurality of consciousnesses is a separate first-person existence with a unique ipseity. Different consciousnesses are ontologically separated from one another by their different ipseities. I exist as a discrete first-person unit with a unique ipseity, just as other consciousnesses exist as discrete units with unique ipseities. My consciousness is my subjective existence and other consciousnesses are different subjective existences.

Given the above, the claim that consciousnesses could undergo fission is necessarily false. Likewise, the claim that consciousnesses could undergo fusion is necessarily false. Each consciousness is a discrete first-person unit in virtue of its unique ipseity that essentially individuates it from other consciousnesses. And so, it is necessarily true that consciousnesses cannot undergo fission. Likewise, it is necessarily true that consciousnesses cannot undergo fusion.

I have, herein, proposed a dualist philosophy of consciousness. The totality of existence contains two separate fundamental kinds, which are: (1) the infinite plurality of consciousnesses that exist as distinct first-person subjects; and (2) the third-person objective world that is realised by them. What I have presented may seem like speculative theorising. Nonetheless, it is philosophically informed theorising based on my acquaintance with and understanding of my consciousness as my first-person existence, and so, to me, it secures a necessary truth about the existence of consciousness.

VI
Other Minds

The first-person subjectivity of consciousness presents a unique epistemic asymmetry. I have direct first-person acquaintance with my subjective experience, and so I know that my consciousness exists. Concerning the subjective experiences of others, however, I am left in the dark, because whereas I can experience their bodies and behaviours, I am not directly acquainted with their subjective experiences. In spite of this, I am certain that I know that other consciousnesses do exist. Given that I have no direct experiential access to these consciousnesses, can I have any justification for this belief in other minds? There are many accounts that provide sound arguments for the existence of other minds, but many of these tend to focus on the psychological and intentional aspects of the mind, while saying little about subjective experience. In this chapter, I shall give a brief overview of some of these accounts, and then propose that we can, in fact, justify the belief that other consciousnesses exist by appealing to the phenomenology of intersubjectivity and the understanding of consciousness as first-person existence. I argue that there exist an infinite plurality of separate consciousnesses, each of which is a distinct and discrete first-person existence.

A well known justification for the belief that other minds exist is John Stuart Mill's (1889) argument from analogy. This argument involves an inductive inference from one's own case to the cases of others. For example, when I listen to the music of my favourite composer W. A. Mozart, I have a sublime experience and I observe myself reacting in a particular way (I am thinking here of the simultaneous combination of three contrasting metres in the dance scene of *Don Giovanni*, wherein Mozart attains a level of contrapuntal profundity greater than even J. S. Bach and, perhaps, all other composers). Observing that others also react in a similar way when listening to Mozart, I make an inductive inference that they too have sublime experiences in this situation. This inference does not need to be based solely on the behavioural reactions. Suppose that I am having a neuroimaging scan while I am listening to Mozart. In addition to my having a sublime experience, I learn about the accompanying brain state. If I observe that others too have similar brain states when listening to Mozart, I can infer that they too have sublime experiences. Thus, the argument from analogy takes one's

own case and projects it onto the cases of others. That is to say, it assumes an analogy between oneself and others.

Some objections have been made to Mill's argument from analogy. Notably, the inductive inference in the argument from analogy is based only on evidence from a single instance. My justification for the belief that others have sublime experiences when listening to Mozart is based only on the fact that I have a sublime experience when listening to Mozart. Hence, the argument from analogy involves very weak inductive reasoning.

Another objection is made by Norman Malcolm (1958), who argues that either there is a criterion that we can use to determine whether one has a given experience, or there is not. If there is such a criterion, then we have no need for the argument from analogy, since we can simply rely on the criterion to determine whether one has an experience or not. However, if there is no such criterion, then we cannot determine whether one has an experience or not, and so we cannot truly understand what it means when we conjecture that one has that sensation.

The trouble with Malcolm's objection is that it assumes a narrowly empiricist reading of the verification principle, which is commonly regarded as problematic. This is a doctrine associated with logical positivism and expounded by A. J. Ayer (1936). It suggests that a statement only has meaning if it can be verified empirically. The following is a brief overview of two criticisms that could be raised against the empiricist verification principle.

First, there are several statements that we make about the external world that cannot be verified, but still have meaning. Notable examples are universal generalisations. This is because our access to the external world is limited to our subjective experiences of it, and so the features we can empirically verify are derived from the data we observe. Universal generalisations we make about unobserved features of the world involve inferential leaps made from our experiences, and, since we cannot access these unobserved features directly, we cannot verify these statements in the manner required by the verification principle. Therefore, according to the verification principle, some universal generalisations we make about the external world are meaningless. However, this is an untenable claim. It is quite clear to me that the statements I make about the external world indeed *do* have meaning, for otherwise they would be of no use in how I engage with the world.

Second, the verification principle, itself, cannot be empirically verified, and so, by its own standard, is meaningless. Hence,

Malcolm's objection based on an empiricist verification principle is unsound. The fact that we cannot empirically verify whether one has an experience does not entail that the statement that one has an experience has no meaning. I, in fact, know what I mean when I say that one has a sublime experience when listening to Mozart.

Interestingly, Peter Strawson (1959) also touches on the question of whether there are criteria for others' experiential states in his own transcendental argument for the existence of other minds. He suggests that one's behaviour can either be a criterion which determines that one is having an experience, or merely be a sign that requires an inductive generalisation for it to suggest that one is having an experience. If the former obtains, then I can determine whether one is having an experience by simply observing one's behaviour. No inductive generalisation is required. If the latter obtains, then I need to make an inductive generalisation from my own case to the case of this other person. That is to say, I need to apply the argument from analogy. Concerning the latter scenario, however, Strawson argues that a necessary condition of my being able to attribute an experience to myself is my acknowledgement that experiences can also be ascribed to others. That is to say, the fact that I refer to it not just as an experience, but as my experience, shows that I also possess the concept of experiences which are not mine, namely others' experiences. Therefore, it follows that I had already known that others have experiences even before I could make the inductive generalisation from my own case to the cases of others. It seems that I do not need to make inductive generalisations after all.

It is worth noting that the argument from analogy seems to presuppose that there is a correlation between behaviour and experience. If, instead of subjective experience, we focus on a psychological property that is characterised by its role in the causation of behaviour, such as a sensation, then the argument from analogy is sound. We can use one's physical characteristics, such as one's behaviour and brain states, to infer that one has a given sensation, because a sensation is a structural and dynamical state which is logically supervenient on the physical. With respect to subjective experience, however, the situation becomes more problematic. Subjective experience is not logically supervenient on the physical, and so there is no logical entailment from observing one's physical characteristics to the conclusion that one has subjective experience. When I observe a person listening to Mozart, I observe that the person displays certain behaviour and I infer from this that the person has the sensation of listening to Mozart.

However, I cannot demonstrate that the person has the subjective experience of listening to Mozart, even though the person actually does have such an experience.

I argue that this is not too much of a problem for the argument from analogy. Although subjective experiences are not logically supervenient on the physical, there is no reason why we should not assume that they supervene naturally. I know that my subjective experience is intimately correlated with the physical activity in my body. That is to say, my body appears to act as an interface between my consciousness and the physical world. It is not unreasonable to infer, from this, that the physical activities in other bodies are also correlated with the subjective experiences of others. Indeed, as David Chalmers (1996) suggests, it would be incredibly arbitrary and counterintuitive to assume otherwise. If my body can act as an interface between my consciousness and the physical world, then plausibly other physically similar bodies too can act as interfaces between other consciousnesses and the physical world. As argued by Paul Ziff (1965), the hypothesis that only I have subjective experience supposes that I must differ significantly from others in some further physiological respect. However, since I do no differ significantly from others in this further respect, it follows that this solipsistic hypothesis is false.

Another problem with the argument from analogy is raised by Ludwig Wittgenstein in his *Philosophical Investigations* (1953). He alludes that it is difficult to imagine someone else experiencing pain because, in order to do so, "I have to imagine pain which I *do not feel* on the model of pain which I *do feel*". Since my only knowledge of pain is from when I feel it, then it appears that I should, from this, infer that pain only occurs when I feel it.

Wittgenstein provides his own argument for the existence of other minds which does not rely on induction from one's own case to the cases of others. For this, he appeals to his argument against the notion of a private language. That is to say, one's linguistic concepts are not acquired by abstraction from one's own case, but are necessarily social. They are learned from and used to communicate with others. What this suggests is that the existence of other minds is implied by the fact that we use language, because language necessarily relies on others. Therefore, instead of an extrapolation from one's own case to the cases of others, Wittgenstein suggests that this be inverted, so that one is applying the concepts learned from the cases of others to one's own case. For example, I know that one is in pain when one is wincing, not because I have made an

inductive generalisation from my own case, but because I had learned, from the cases of others, the concept of pain through reference to such behavioural reactions as wincing.

It is commonly objected that conceptualising one's mental states as concepts learned from observing the behaviours of others amounts to a form of behaviourism. To be clear, it is controversial whether or not this is Wittgenstein's intention. Nonetheless, the problem with such behaviourism is that it focuses entirely on one's third-person behaviour, while ignoring one's first-person subjective experience. The claim that one learns the concept of pain by observing others does not imply that all there is to pain is a behavioural reaction. Indeed, I may learn about the behaviour associated with pain by observing how others behave in certain situations and I may even display this behaviour in similar situations, but I also have a first-person subjective experience of pain. This experience accompanies my behaviour, but is a distinct feature from it. However, given that I have privileged access to this experience, one's observation of my behaviour does not secure one's knowledge of my experience.

Another justification for the belief in other minds, suggested by Hilary Putnam (1975), is that the claim that there are other minds involves an inference to the best explanation. According to Putnam, we ascribe mental states to others, because their having mental states is the best explanation we have for their behaviours. Indeed, it appears that any alternative hypothesis would be more complex and less pragmatic than the hypothesis that others have mental states. For example, consider an alternative hypothesis which makes no reference to mental states at all, but, rather, attempts to explain the one's behaviour in terms of the neural circuitry of one's brain and the dynamics of the environment wherein one is embedded. Although this hypothesis may, in principle, be able to explain one's behaviour in the specific context, the explanation would be convoluted, difficult to communicate, and unlikely to generalise to other social contexts. By contrast, the hypothesis that one has mental states would explain one's behaviour by ascribing to the person beliefs, desires, and other attitudes. This hypothesis also successfully explains one's behaviour, but is simpler, more comprehensible, and more generalisable than the alternative hypothesis.

First, Putnam's account appears not to be an argument for the existence of other minds, but, rather, an explanation of why we believe that others have minds. The reason, Putnam argues, is that the assumption that others have minds is the best explanation that we have for their behaviours. However, as noted above, the reason we

have for rejecting the alternative hypothesis is not that it does not explain the empirical data, but that its explanation is far more complicated than the explanation provided by the hypothesis that others have minds. That is to say, we favour the hypothesis that others have minds over the alternative hypothesis because it is simpler and more pragmatic. These are important considerations in hypothesis selection, but they are not necessarily truth conducive.

Second, Putnam's account appears only to focus on the psychological aspects of the mind that are involved in the generation of behaviour. Indeed, one's having psychological properties is a good explanation for one's behaviour, but this is because these structural and dynamical properties are logically supervenient on the physical, and so have causal roles. However, the same cannot be said for subjective experience. Given its irreducible subjectivity, experience is not logically supervenient on the physical, and so does not have a causal role. Therefore, Putnam's argument from inference to the best explanation can only justify the act of ascribing psychological states to others, but it cannot justify the act of ascribing phenomenal states to others.

Having considered some of the current arguments for other minds, I would now like to leave my own account. Specifically, I argue that there exist an infinite plurality of consciousnesses. The account I shall give is of a very different kind from the arguments we have seen in this chapter. Instead of relying on inferences from our behaviours and linguistic practices, it reflects on the ontology of consciousness itself as first-person existence. More specifically, I argue that we can prove that other consciousnesses exist, first, by appealing to the intersubjective phenomenology of consciousness and, second, by appealing to the absolute nature of existence.

Regarding the intersubjective phenomenology of consciousness, the fact that there are consciousnesses other than mine is entailed by and accounts for the form of my conscious experience. As noted by Edmund Husserl (1931), an integral feature of my experience of the world is that the world is also experienceable by others. I am directly acquainted with myself as an individual subject with a particular viewpoint on the world, but this discernment of myself as an individual subject is transcendentally dependent on there being other subjects from whom I can distinguish myself as an individual. Therefore, intersubjectivity is a necessary condition for subjectivity. Likewise, Jean-Paul Sartre (1943) argues that knowledge of other minds is *a priori*, as our relations with one another are basic features of our engagements with the world as subjects. From the fact that I

exist, it necessarily follows that others exist, because the relations between me and others are necessary for my discernment of myself as a distinct being. I know that there exist other subjective viewpoints, because these viewpoints are integral to my viewpoint.

And so, the above conveys a transcendental argument. The fact that other consciousnesses exist is a necessary condition for my acknowledgement of the existence of my consciousness as a distinct subjective viewpoint. Moreover, the fact that there exist other consciousnesses is necessary for the very discernment of *my* consciousness and *other* consciousnesses. Indeed, acknowledging the subjectivities of others could be foundational to an egalitarian moral philosophy that values social justice, as it recognises that it is true that we are all coequals as conscious subjects.

The above is also supported analytically. As noted earlier, Strawson (1959) observes that my ability to ascribe an experience to myself necessitates that I acknowledge that experiences can be ascribed to others. Given its first-person subjectivity, a fact about consciousness is essentially indexical. Consciousness is always someone's consciousness. Indexicality is relational, insofar as an indexical specifies its referent in contrast to other tokens. Hence, the meaning of an indexical presupposes that these other tokens exist. For example, "this" presupposes there is a "that". Likewise, the existence of *my* consciousness entails that there exist *other* consciousnesses that are not mine. Therefore, solipsism is false.

Regarding the absolute nature of existence, recall that the world is only realised through the subjective experience of it by consciousness. Thus, my consciousness is a subjective realisation of existence. Furthermore, it is true that the totality of existence is absolute because, by definition, existence is *what is*. Given that it is absolute, there are infinite possible ways for existence to be subjectively realised, of which my consciousness is one. Indeed, any given point has infinite angles from which it can be approached. Accordingly, existence is experienced from a first-person viewpoint and there are infinite potential viewpoints from which this absolute existence could be experienced, which respectively correspond to different consciousnesses. Moreover, in light of what is noted above, I know that these infinite consciousnesses exist, because the fact that there exist infinite consciousnesses is necessary for my discernment of my consciousness as one specific subjective viewpoint out of an infinite plurality of potential distinct subjective viewpoints.

In some respect, this partly evokes Gottfried Wilhelm von Leibniz's (1714) suggestion that there exists an infinite plurality of

separate monads, each being an individual soul which mirrors the universe from its own point of view. Herein, I am proposing that there are an infinite plurality of consciousnesses, each of which is a real and distinct first-person existence. However, as noted in chapter four, Leibniz's monadological monism is problematic because it assumes a necessary connection between the physical and the mental, which fails to account for the conceivability of modal variation between these two domains. And so, such monist panpsychism is false. Instead, my account has a fundamentally dualist ontology, which accounts for the conceivability of this modal variation. There exist an infinite plurality of consciousnesses, which are ontologically distinct from the physical world.

The notion of existence as absolute may also appear to recall Georg Wilhelm Friedrich Hegel's (1816) absolute idealism, insofar as it appears to suggest that individual persons are aspects of an absolute unity, but I argue that this monist interpretation of absolute idealism is false. Given the first-person ontology of consciousness, each consciousness is essentially distinct from other consciousnesses. As I noted in chapter two, the "I" that is my consciousness exists as a discrete and uniquely individuated first-person unit. Likewise, other consciousnesses also exist as discrete first-person units. The identity or haecceity of any given consciousness is determined by its unique first-person individuation, which is essentially different from the first-person individuation of any other consciousness. Hence, it must be taken as true that each consciousness is ontologically different from other consciousnesses in virtue of such individuation. Even if, in a possible world, there is a consciousness associated with the entire universe, this would be a discrete first-person unit that is distinct from the plurality of other consciousnesses that also exist as discrete first-person units.

In light of the unique first-person individuation that determines the haecceity of a given consciousness, it is true that consciousnesses exist as ontologically discrete units that are essentially separate from one another. The claim that consciousnesses could undergo fission is necessarily false. Likewise, the claim that consciousnesses could undergo fusion is necessarily false. Each consciousness is a discrete first-person unit in virtue of its unique ipseity that essentially individuates it from other consciousnesses. Accordingly, it is false to characterise the absolute as a fusion of consciousnesses. Rather, the absolute totality of existence contains an infinite plurality of discrete consciousnesses that are separate from one another, in conjunction with the objective world that these consciousnesses experience. It is

necessarily true that consciousnesses cannot undergo fission, just as it is necessarily true that consciousnesses cannot undergo fusion.

What I have presented above proves that solipsism is necessarily false. There are infinite potential subjective realisations of existence. Indeed, the suggestion that existence only has a finite number of potential subjective realisations is false, because any given point has infinite angles from which it can be approached. Existence is *what is* and there are infinite potential viewpoints from which *what is* could be realised. Moreover, the fact that these infinite subjective realisations actually exist is entailed *a priori* by the intersubjective phenomenology of consciousness and by the relationality of the essential indexicality of consciousness. That is to say, the fact that there exist infinite consciousnesses is a necessary condition for the very discernment of my consciousness as one specific subjective viewpoint out of an infinite plurality of potential distinct subjective viewpoints. I know that there exist other consciousnesses, because the fact that there exist other consciousnesses is necessary for the very discernment of *my* consciousness and *other* consciousnesses. Thus, it is necessarily true that an infinite plurality of consciousnesses exist.

Some metaphorical resemblance might be noted with David Lewis' (1986) modal realism, Hugh Everett's (1957) "many worlds" interpretation of quantum mechanics, and multiverse theory. Modal realism is the philosophical theory that it is true that all possible worlds are real, the "many worlds" interpretation is a model for interpreting some aspects of quantum mechanics, and multiverse theory is the cosmological theory that there are multiple concrete universes. These views appeal to the realisation of multiple possibilities through their manifestations in different worlds. However, these views do not correspond to the view I am proposing, which concerns the plurality of subjective realisations of existence and not the plurality of concrete physical worlds. Whether there are many concrete physical worlds is irrelevant to the issue of whether there are many subjective experiencers, and so it is false to suppose that the "many worlds" interpretation affects the fact that there exist an infinite plurality of consciousnesses. Hence, the view I am proposing in this chapter is about a fundamentally different issue from the views of Lewis and Everett. Nonetheless, in a metaphorical sense, the realisation of multiple possibilities in Lewis' and Everett's views can be considered to evoke the multiple possible realisations of existence I am proposing. The infinite plurality of consciousnesses reflect the infinite potential ways existence can be experienced.

VII
Psychophysical Interaction

It is apparent that the subjective qualities that are experienced by my consciousness are robustly correlated with certain happenings in the objective world, specifically the processes that occur in the vicinity of my body. In the reality that I experience, I assume the viewpoint of a person, and it is the happenings in the body of this person that appear to evoke the subjective qualities that I experience. Therefore, although consciousness exists as a separate entity from the objective world, it is reasonable to assume that there is a certain psychophysical interface between the two. Subjective experience is associated with the objective world in such a way that certain events in certain parts of the objective world are correlated with certain subjective qualities in consciousnesses. In the case of my own consciousness, this interface appears to be most strongly associated with a certain bodily system, that is, my central nervous system. In this chapter, I shall explore the nature of this interface.

Embodiment

There seems to be a correlation between what happens in my body and what I experience in my consciousness. For example, my visual, auditory, olfactory, gustatory, tactile, thermal, and painful qualia appear to be correlated with the stimulation of certain receptors by their respective stimuli. With respect to visual qualia, it is the stimulation of photoreceptors on the retina by photons. With respect to auditory qualia, it is the mechanical stimulation of cochlear cells by vibrations. With respect to olfactory and gustatory qualia, it is the chemical stimulation of chemoreceptors on the nasal mucosa and lingual mucosa, respectively. With respect to tactile experiences, it is the mechanical stimulation of mechanoreceptors on the skin. With respect to thermal experiences, it is the thermal stimulation of hot and cold receptors on the skin. With respect to painful experiences, it is the stimulation of C-fibres and Aδ-fibres by noxious stimuli.

The stimulation of these receptors causes the selective permeability of the receptor membrane to change via the opening and closing of certain intramembrane ion channels. This produces an electric current by enabling the flow of ions across the receptor

membrane in process called transduction. This electric current is then propagated as action potentials along neurones to certain areas of the brain, via synapses in the thalamic nuclei and elsewhere. The activation of certain areas of the brain appears to be correlated with the experience of certain qualia. Such areas include the primary visual cortex in the occipital lobe for visual experiences, the primary auditory cortex in the temporal lobe for auditory experiences, the uncus and parahippocampal gyrus in the temporal lobe for olfactory experiences, the insula in the depths of the lateral sulcus for gustatory experiences, and the primary somatosensory cortex in the parietal lobe for tactile, thermal, and painful experiences.

Other areas of my brain are also associated with other modalities of experience. For example, the activity of certain areas of my association cortex may be correlated with the qualia associated with cognition, imagination, and memory recall, while frontal lobe activity may be correlated with the qualia associated with the planning and performance of voluntary action. Subcortical structures too may be associated with certain qualia. For example, the activity of the limbic system, including the amygdala, is thought to be correlated with emotional qualia, while the activity of the nuclei of the reticular formation may be correlated with qualia relating to certain states of mind, such as arousal and somnolence.

And so, the psychophysical interaction between the physical world and my consciousness seems to be most strongly concentrated in the activity of my nervous system. This partly accounts for why I experience the world from an embodied perspective. Events in the environment occasion changes in my body, which are detected by receptors, which then transmit signals to my brain. My brain processes these signals and transmits signals back to the rest of my body which produce physiological and behavioural responses. In addition, these processes in my body and brain are also accompanied by the subjective experience of qualia in my consciousness.

Psychophysical laws

This correlation between neural activity and the experience of qualia appears to suggest that there is something special about the brain that allows it to act as an interface between the physical world and consciousness. Consequently, there have been many theories that have attempted to unlock this apparent property. Among them are intricate accounts of the brain's biochemistry, electromagnetic

activity, causal organisation, and even its quantum microstructure. However, I argue that any such account is insufficient, insofar as it is a physical account about the structure and dynamics of the brain. As I noted in chapter four, structural and dynamical facts can only yield further structural and dynamical facts, but they do not entail anything about the presence of first-person subjectivity. Therefore, research into neural mechanisms may tell us about how the brain processes stimuli to produce behavioural outputs, but it cannot tell us why it acts as an interface between the physical world and consciousness.

Accordingly, I argue that there is nothing necessarily special about the physical structure or activity of the brain that allows it to act as a psychophysical interface. Moreover, I argue that the brain is not the only kind of structure that can act as a psychophysical interface. Despite its intricacy, the brain is effectively just a collection of matter arranged in a certain configuration. This configuration allows it to operate in a certain way physically. However, nothing in its configuration entails that the brain should act as an interface with consciousness. And so, the brain is as metaphysically likely or unlikely as any other physical system to act as a psychophysical interface.

Instead of suggesting that there is something physically special about my brain that allows it to act as an interface with my consciousness, I propose that it is a contingent fact about this world that there is a certain correlation between physical events and subjective experiences, such that physical events in various parts of this world are mirrored by the subjective experiences of consciousnesses. It just so happens that the physical events that are mirrored by the subjective experiences in my consciousness are those in a particular region of this world, namely my body. Furthermore, it is reasonable to suggest that physical events in other parts of this world are mirrored by subjective experiences in other consciousnesses. What I am proposing is a form of regularism. The physical facts about the world do not entail that certain systems should be accompanied by consciousnesses, and so there is no necessary connection between physicality and phenomenality. Rather, it is a contingent fact about this world that some physical events are mirrored by subjective qualities.

This may initially seem to recall Gottfried Wilhelm von Leibniz's (1686) parallelism, which suggests that physical events and mental events accord with each other like perfectly synchronised clocks. Under such a view, there would be no need to postulate any form of necessary connection between physicality and phenomenality.

Instead, there would just be a regular yet contingent correlation between the two. Furthermore, there would be no need to assume a causal relationship between physical events and subjective qualities. If we consider Leibniz's analogy with the synchronised clocks, the clocks coincide with each other, but neither one causally influences the other. Similarly, under such a view, physical events may coincide with subjective qualities, but it cannot be said that the physical events cause the subjective qualities to arise, or that the subjective qualities cause the physical events to occur.

However, Leibniz's parallelism is unsound, for it suggests that the correlation between physicality and phenomenality is merely accidental, and so appears to deny any genuine form of interface between the physical world and consciousness. This idea of accidental synchronicity without any interface is extremely arbitrary. Thus, in the form of regularism I am advocating, there is a nomological relation between the mental and the physical, rather than mere synchronicity. That is to say, physicality and phenomenality are not correlated by chance, but are coordinated in ways that are robust but contingent. This is not a mechanistic relation, but a relation of regularity. One does not push or pull the other in a temporally dependent fashion, but both influence each other in a regular manner. Again, there is no need to posit any necessary connection. There is, in this world, a regularity between certain physical events and certain subjective experiences, but this is just a contingent property of this world. There is no logical reason why this regularity must hold across all possible worlds.

This complements the view endorsed by David Chalmers (1996), who proposes that there are, in this world, psychophysical laws that establish robust correlations between physical events and subjective qualities. I am broadly in agreement with this view and I argue that it can be made compatible with the regularism I am advocating, specifically if we acknowledge these laws as being descriptive and not prescriptive. That is to say, these psychophysical laws do not dictate that physical events cause subjective qualities to arise, but describe the correlations between these physical events and subjective experiences. They capture the relations between events in the physical world and the contents of consciousness. Furthermore, these laws are contingent. Certain physical events may be correlated with certain phenomenal qualities in this world, but they may not be in other worlds. In this world, it so happens that the events in the vicinity of my body are correlated with the phenomenal qualities that are experienced by my consciousness. However, in a possible

zombie world, there may not be this correlation, and so bodies may not be associated with consciousnesses in that world. Alternatively, in a phenomenally inverted world, the correlation may be different. While bodies may be associated with consciousnesses in that world, the bodily states may be correlated with different qualities to those with which the analogous bodily states in this world are correlated.

In summary, while the existence of my consciousness is necessary for my experience, the manner in which my experience is currently embodied is a contingent feature of this world. My body acts as an interface between the world and my consciousness because certain psychophysical laws obtain in this world that occasion correlations between certain bodily events and certain phenomenal qualities. There is no necessary connection between embodiment and subjective experience, but they do, in this world, reflect each other in regular ways. Hence, while physicalism is false because it fails to account for subjective experience, idealistic monism is false because it fails to account for the intersubjective reliability and regularity of the events that we experience. To account for the existence of first-person subjective experience and the regularity of the third-person objective world, it is necessary to acknowledge that dualism is true.

Other consciousnesses

In chapter six, I proposed an infinite plurality of separate consciousnesses that exist as discrete and distinct beings. In this chapter, I am proposing that certain physical events and certain subjective qualities are nomologically and contingently correlated. A synthesis of the above would imply that individual consciousnesses, from the infinite plurality of consciousnesses, are associated with certain systems in this world and other worlds. In this section, I shall speculate on what the picture I have presented can say about which things in this world are actually associated with consciousnesses. What features are conscious and what features are nonconscious?

As noted earlier, nothing in the physical facts about a system entails whether or not it should be accompanied by consciousness. My body is just as metaphysically likely or unlikely to act as an interface with consciousness as any other system in this world. Given that my body is capable of acting as a psychophysical interface, there seems to be no reason why any other systems should not be capable.

However, this does not warrant panpsychism. Indeed, all systems are metaphysically as likely or unlikely to be associated with

consciousnesses as my body is, but this does not mean that all systems are naturally associated with consciousnesses in the actual world. There is no necessary connection between physicality and phenomenality. Rather, there are correlations between certain physical events and certain phenomenal qualities due to a nomological relation between physicality and phenomenality in this world. This nomological relation is contingent. It holds in this world for my brain and my consciousness, but it does not have to hold for another brain that is physically indistinguishable from mine in another world that is physically indistinguishable from this world.

Given that the correlation between physicality and phenomenality is contingent, it is plausible that naturally, in this specific world, only certain physical events are correlated with phenomenal qualities. Thus, although every physical system is as metaphysically capable of acting as a psychophysical interface as my brain is, it may be the case that only some physical systems naturally act as psychophysical interfaces in this world, whereas others do not. Of course, due to the subjectivity of consciousness, we cannot empirically demonstrate which physical systems do and do not act as interfaces in this world. However, given the intersubjective phenomenology of our relations with one another, it is reasonable to acknowledge, in this world, that other humans are conscious. It is also reasonable to accept that other animals, plants, fungi, protozoa, bacteria, viruses, and even certain artificial systems are also conscious. Conversely, it is reasonable to suppose that many inanimate objects in this world, such as tables and chairs, are nonconscious. Also, while organisms are conscious, it is reasonable to suppose that certain integrants associated with their bodies, such as the viscera of animals, the fruits of trees, and the endotoxins of bacteria, are nonconscious.

As well as intersubjective phenomenology, the above also involves an inference from my own case. My consciousness is associated with my body, and so it is reasonable to suppose, in this world, that other systems which share a similar kind of organic basis or a similar kind of causal organisation are also associated with consciousnesses and that inanimate objects that do not share a similar kind of organic basis or a similar kind of causal organisation are nonconscious. This may seem like a weak inference, but it is reasonable to accept. One can conceive that there is, as Thomas Nagel (1974) notes, “something it is like” to be an autonomous system that is different from but, in some respect, similar to oneself.

In other possible worlds, or in other universes of the multiverse, different systems may act as interfaces. Perhaps the psychophysical

laws between physical events and subjective qualities are different in other possible worlds or other universes. Accordingly, although inanimate objects such as chess pieces are nonconscious in this world, there could be a possible world wherein chess pieces are associated with consciousnesses. Likewise, there could be worlds wherein there are immaterial beings associated with consciousnesses, worlds wherein fictional characters are associated with consciousnesses, and worlds wherein mythological characters are associated with consciousnesses. There could even be a possible world wherein elementary particles are conscious and a possible world wherein the entire universe as a whole is conscious. These speculations may be extravagant, but insofar as these psychophysical laws are contingent, it is logically conceivable that such structures could be associated with consciousnesses in other possible worlds.

The systems which act as interfaces in other worlds could be geometrically different from those in this world. This is because consciousness is not constrained by facts about geometry. Indeed, the claim that consciousness is constrained by mathematics is false. Given all the mathematical facts about the world, the existence of consciousness remains a further fact to consider. Therefore, it is true that the existence of consciousness is beyond the mathematical facts about the world. Accordingly, there could be worlds wherein structures with different spatial dimensions are associated with consciousnesses. For example, there could be a world, such as that in Edwin Abbott's *Flatland* (1844), wherein consciousnesses are associated with beings on a two-dimensional surface.

One may object that there is "nowhere in a chess piece for a consciousness to fit", but I argue that this objection involves a category mistake. Indeed, consciousness cannot be found in a chess piece, but consciousness also cannot be found in the brain. This is because consciousness is not located in the objective world, but is the first-person subjective existence that experiences the world. That is to say, consciousness is not located in the brain, but is a separate existence that is contingently associated with the brain. Hence, in the aforementioned possible world, consciousness is not located in the chess piece, but is a separate existence that is contingently associated with the chess piece. Likewise, in a possible world where fictional realism obtains, the claim that a fictional character is conscious could mean that a consciousness is associated with the abstract individual that is the fictional character.

I am, therefore, proposing a constrained form of psychophysical liberalism. There is no necessary connection between physicality and

phenomenality, and so every system is metaphysically capable of being associated with consciousness. Nonetheless, the contingency of psychophysical correlation means that although every system is metaphysically capable of acting as a psychophysical interface, not every system must be naturally capable in any given world.

This is compatible with acknowledging, as I did in chapter six, that there exist an infinite plurality of consciousnesses. Indeed, it is true that there exist an infinite plurality of consciousnesses across the totality of existing worlds or across the infinite expanse of existence. However, in any given world, which systems are associated with consciousnesses will depend on the psychophysical laws in that world. While a given system may not act as a psychophysical interface in this world, it could be associated with a consciousness in another world. And so, across the totality of worlds, different systems may turn out to be associated with consciousnesses. There may even be a possible world where all systems are associated with consciousnesses. Throughout the infinite totality of existence, there may even be consciousnesses that do not have psychophysical interfaces. Yet, in the world wherein we currently reside, it is plausible that the psychophysical laws operate in such ways that only some sorts of systems turn out to be associated with consciousnesses.

Some speculation

There may, at this point, be some questions concerning what happens when a psychophysical interface is formed, changed, or destroyed, such as in reproduction, development, and death. We cannot establish the answers for sure, but there are hypotheses that are highly speculative yet reasonable to accept. What I suggest in the rest of this chapter should not be interpreted as definitive answers to these questions, but merely as loose speculations that are conceivable under the framework presented in this book. The assumption underpinning these speculations is that the psychophysical laws in our world are such that certain biological systems and some artificial systems are contingently associated with consciousnesses.

New psychophysical interfaces can be thought to be formed in reproduction. In vertebrate sexual reproduction, an ovum is fertilised by a spermatozoan to produce an embryo. It is reasonable to suppose that the both of the parents themselves are associated with consciousnesses, although the ovum and spermatozoan they produce are nonconscious, and that another consciousness becomes

associated with the resulting foetus after a sufficient period of gestation has taken place to enable the foetus to develop into a system that is capable of acting as a psychophysical interface. In cases of multiple gestations, the different foetuses become associated with different consciousnesses. For example, as noted in chapter two, a pair of twins, whether they are unconjoined or conjoined, have different brainstems, and so are associated with two different consciousnesses. By contrast, a person with chimerism, where the body is formed by the aggregation of cells with different genotypes, has a single brain, and so is associated with a single consciousness.

In the process of asexual reproduction, such as the splitting of the planarian *Girardia tigrina*, I suspect that one resulting organism would continue to be associated with the same consciousness with which the organism was associated before splitting, while a different consciousness would become associated with the other resulting organism. Importantly, this is not the "fission" of a consciousness, which is impossible given the discrete first-person individuation of consciousness. Rather, the distinct consciousness that was associated with the original organism remains associated with one resulting organism, while another distinct consciousness becomes associated with the other resulting organism. The above may also apply to the asexual reproduction of a plant such as *Populus tremuloides*, which produces clones whose roots grow from a rhizome. Here, it is reasonable to suppose that the original plant would continue to be associated with the same consciousness, while the clones become associated with different consciousnesses.

Some organisms have more complex methods of reproducing. In parasitic flukes of the species *Fasciola hepatica*, a miracidium develops into a sporocyte in the intermediate host. Through asexual reproduction, the sporocyte produces rediae and cercariae, which then develop into metacercariae in the definitive host. The metacercariae are capable of sexually reproducing in the definitive host, while the sporocyte remains in the intermediate host. Here, it is reasonable to suppose that the miracidium and the sporocyte are associated with the same consciousness, while the rediae and cercariae that are produced and develop into metacercariae are associated with different consciousnesses. In multicellular fungi of the species *Agaricus bisporus*, two haploid hyphae of different mating strains meet to produce a dikaryotic mycelium, which then grows into a basidiocarp. Within the basidiocarp, karyogamy and meiosis occur, generating new haploid basidiospores, which then germinate into new hyphae. I argue that it is reasonable to suppose

that the two original haploid organisms form interfaces with two different consciousnesses. I would also suppose that the dikaryotic organism that results from the merging of their hyphae also forms an interface with a different consciousness. Considering this, I would suppose that the new haploid organisms that are produced from karyogamy and meiosis in the basidiocarp then form interfaces that become associated with different consciousnesses.

In the sexual reproduction of the unicellular fungi of the species *Saccharomyces cerevisiae*, two haploid cells merge to form a diploid cell. It could be supposed that the two consciousnesses that were associated with the original haploid organisms are no longer associated with them after the cells merge, while a different consciousness becomes associated with the resulting diploid organism. This is not the "fusion" of consciousnesses, which is impossible given the first-person individuation of consciousness. Rather, a distinct consciousness was associated with one haploid organism, another distinct consciousness was associated with the other haploid organism, and another distinct consciousness becomes associated with the diploid organism. Such cellular merging is different from the phagocytosis and degradation, for example, of a *Paramecium aurelia* by an *Amoeba proteus*. Here, it could be supposed that the *Amoeba proteus* continues to be associated with the same consciousness, while the consciousness that had previously been associated with the *Paramecium aurelia* is no longer associated with it after phagocytosis and degradation.

A structural change in a psychophysical interface may occur in the development of an organism. An example is the metamorphosis of a caterpillar into a butterfly. Here, it is plausible that the organism remains associated with the same consciousness throughout metamorphosis, regardless of the change in morphology.

Sometimes, several individual organisms gather to form a colony, such as with slime moulds of the species *Fuligo septica*. The colony has a single membrane but the distinct nuclei of the individual organisms are maintained. Colony formation is different from the aforementioned process of cellular merging where the result is a single merged nucleus instead of the two original nuclei. In colony formation, it could be supposed that the individual organisms or nuclei remain interfaces with their respective consciousnesses, while the colony as a whole remains nonconscious. This recalls the thought experiment by Ned Block (1978), wherein the entire population of a country is arranged into an isomorphic realisation of the brain. While each member of the population is conscious, it is reasonable to

suppose that the collection as a whole is nonconscious, insofar as the individual members are still independent and autonomous. A similar analysis may also apply to the process of inosculation, which is where the branches of two different trees that are in prolonged contact during growth eventually join together. Here, it is reasonable to suppose that the two organisms, which had different embryonic origins, remain associated with their two respective consciousnesses, while the combined system they make up remains nonconscious.

The above process of colony formation is different from the process of growth through mitosis in a multicellular organism. While a colony is comprised of a group of organisms that could still subsist independently of one another to certain extents, the cells in a multicellular organism are dependent on one another and on the activity of the whole system. Hence, in the case of a genuinely multicellular organism, the whole organism may be associated with a consciousness, but the cellular components may be nonconscious. By contrast, in a colony, the individual organisms may be associated with consciousnesses, but the whole colony may be nonconscious.

As noted earlier, these speculations are reasonable given that the psychophysical laws in this world work in certain ways. However, in a different world where the psychophysical laws work differently, a large-scale system may be associated with its own consciousness and the small-scale components of the system may also be associated with their own consciousnesses. For example, in such a possible world, there may be a distinct consciousness associated with a macroscopic organism whose bodily components are microscopic organisms that are also associated with distinct consciousnesses.

As with the example of a colony, it is plausible that in ectosymbiosis, the individual organisms remain interfaces with their respective consciousnesses, while the collective system remains nonconscious. For example, the human body contains many commensal and mutualistic microorganisms. It is reasonable to suppose that the human is conscious and the individual microorganisms are conscious, but that the combined aggregate of the human and microorganisms is nonconscious. However, with respect to endosymbiosis, the process may later result in structures derived from endosymbiotic organisms being integrated into the hosts' cells as organelles. Here, it is plausible that the past symbiotic organisms from which the present organelles are historically derived were associated with consciousnesses, but the derived organelles that have been integrated into the hosts' cells are not. For example, we can suppose that present mitochondria are nonconscious because

they are fully integrated into the hosts' cells, but that the past autonomous organisms from which mitochondria are historically derived were associated with consciousnesses.

Somewhat similar reasoning might be applied to other cases where nonconscious components of systems that are or were otherwise associated with consciousnesses are removed and instead integrated into the other systems that are associated with different consciousnesses. For example, the process of grafting, where a scion from a donor plant is grafted onto the rootstock of a recipient plant. Here, it is reasonable to suppose that the two plants, namely the donor and the recipient, which remain associated with their respective roots, continue to be associated with their two respective consciousnesses. The scion itself is nonconscious, but becomes integrated into the system that is the recipient plant.

New psychophysical interfaces can also be thought to be formed in artificial processes. For example, in the manufacturing of a conscious artificial intelligence, it could be assumed that a consciousness becomes associated with a system or software that can act as such an interface. Importantly, this is not the "generation" of a consciousness, which is impossible given the timelessness of consciousness. Rather, an existing consciousness becomes associated with a system that happens to act as a psychophysical interface.

Finally, there is case of the destruction of an interface, such as in death. Importantly, as I shall argue in chapter ten, this is not the "annihilation" of a consciousness, which is impossible given the timelessness of consciousness. Rather, consciousness remains a further fact beyond the physical facts about the system. And so, while a given consciousness would no longer be associated with the destroyed interface, it is reasonable to suppose that this same consciousness could then become associated with another interface.

What I have suggested in this section is not vitalism. I am not suggesting that organisms act as psychophysical interfaces because of some "life force" that guides their workings. The dualism I am proposing accepts that life is a physical process that can be explained in structural and dynamical terms, and so we can accept that vitalism has been scientifically falsified. Rather, what I have suggested is that consciousnesses are contingently associated with certain biological and artificial systems in this world in view of the psychophysical laws that operate in this world. Given that consciousness is itself nonphysical, it is not required to explain life. Thus, we can accept that consciousness is a further fact that is contingently associated with life and accept that life itself can be physically explained.

VIII
Constructing Reality

Reality is constructed from experience. One's image of the world is built from the qualities experienced within the first-person existence of one's consciousness and from one's intersubjective interactions with other experiencers. Furthermore, our inferences about this reality, from everyday classifications of objects to abstract scientific theories, are derived from our enquiries into the patterns which we discern in our experiences.

These classifications, in turn, influence how reality is perceived. For example, at this moment in my visual field, there is a patch of black. Upon my reaching out and touching it, it feels solid. However, I do not perceive this as a mere bundle of qualia. Rather, I classify combinations of qualia into objects. In this particular case, the combination of qualia is classified as a table. Just as I classify combinations of qualia into objects, scientists make inferences about the perceived objects and formulate theories to explain them. Thus, we tend to view our reality as composed of objects, such as tables and chairs, and a scientist may say that these are composed of theoretical constituents, such as atoms and molecules.

Nonetheless, these objects are not the basic constituents that make up reality, but are constructs derived from them. Rather, the basic constituents are the experiences that occur within our consciousnesses. The categories and theories that we use are means of organising these experiences and the patterns they form. What are these experiences like? Neuroscience can detail the neural processes involved in sensations, but these are not what concern me here. As I have said, theories about the physical world are constructs derived from the patterns found within experiences. What I am examining here are the phenomenal qualities of the experiences themselves.

A phenomenology of conscious experience

The following is a classification of some of the qualia that constitute the reality that I experience. The list is by no means exhaustive. How I experience the world is influenced by how my interface with the world is embodied, and so there may be other sorts of qualia that can be experienced by subjects who are embodied differently.

Nevertheless, it may help to illustrate how some of my experiences combine and interact to form my subjective reality. As I have said, the following descriptions are not accounts based on neurophysiology or psychophysics, but are subjective reflections on the qualitative nature of the experiences themselves.

Visual qualia: In everyday life, we use many different terms to describe visual qualia, such as hue, intensity, brightness, and contrast. Furthermore, our scientific knowledge has a lot to say about different properties of light, and the neural mechanisms of visual perception. However, phenomenologically, visual qualia have two main parameters, which are colour and space. My visual qualia are composed of various colours at various spatial locations. Hue, intensity, brightness, and contrast, are terms used to classify different kinds of visual qualia and the relations between them, but the qualia themselves are composed of colours on a visual space.

Auditory qualia: Phenomenologically, auditory qualia have two main parameters, which are pitch and volume. They are composed of combinations of different pitches at different volumes. Furthermore, a third parameter of space can be added, since my auditory qualia are binaural. Properties, such as timbre, are due to the fact that I can experience more than one pitch at a time, so different combinations of pitches result in different qualities, from the crystalline attack of the piano to the unrefined rumble of thunder.

Olfactory qualia: It is reasonable to think of olfactory qualia as having two parameters of quality and intensity. Unlike auditory qualities, with which there is a gradual continuum of pitch, olfactory qualities can be arbitrary and discontinuous. There is not, with olfactory qualities, an obvious one-dimensional scale, as there is with pitch. Rather, there is a diverse variety of olfactory qualities.

Gustatory qualia: Like olfactory qualia, gustatory qualia involve the parameters of quality and intensity. We commonly recognise five basic gustatory qualities, namely sweetness, acidity, salinity, bitterness, and umami, with umami being the gustatory experience associated with the taste of monosodium glutamate. Furthermore, gustatory qualia combine with olfactory qualia to yield an extraordinarily vast spectrum of possible flavours, from the freshness and crisp acidity of a Touraine Sauvignon Blanc to the richness and bitterness of a Uganda Robusta coffee.

Tactile qualia: Phenomenologically, tactile qualia can be analysed as having two main parameters of intensity and space. They are composed of sensations of various intensities, and at various locations on a space that I associate with the area of my body. These

are capable of combining to produce a subtle variety of tactile qualities, such as softness, stickiness, wetness, and so on.

Thermal qualia: Evidence from neuroscience suggest that there are two different kinds of receptor for heat and cold, respectively, but phenomenally, qualia associated with temperature could be interpreted as lying along a single one-dimensional continuum of warmth, ranging from very cold to very hot. As well as this, thermal qualia can often be located on a space, much like tactile qualia.

Proprioceptive qualia: These refer to qualia associated with the spatial position and movement of the body. Such qualia include those concerning spatial orientation, joint position, motion, and balance. They also include the disruptive qualities of dizziness and disorientation.

Visceral qualia: This is a generic term I am using to describe the diverse qualia that are associated with internal bodily sensations. These vary greatly in quality. Examples of this group of qualia include, amongst others, those associated with hunger, thirst, satiety, nausea, breathlessness, arousal, and orgasm.

Painful qualia: In a broad sense, these are qualia associated with aversive stimuli and with actual or potential tissue damage. Painful qualia vary greatly in quality and intensity, and in addition, they can sometimes be localised in space, such as the pain that can be associated with tactile, thermal, and visceral qualia.

Affective qualia: I am referring here to the diverse group of qualia associated with various mood states. These vary in quality, and are usually prolonged, indistinct, and nonlocal. They often present as a homogenous background atmosphere that permeates throughout the vicinity of my experience. Examples include the qualia associated with attention, arousal, and somnolence. I also include, amongst these, the qualia associated with emotions. Examples are those qualia associated with happiness, sadness, anger, fear, desire, and aversion. Affective states are also socially influenced and exhibit diverse variations across cultural contexts.

Ideational qualia: These refer to the qualia that are associated with the deliberative cognitive processes, such as imagination, reasoning, recall, and prediction, amongst others. The nature of the qualia associated with these processes often reflects the nature of the other sorts of qualia, although ideas tend to be fainter and less distinct. For example, the idea that I experience upon recalling the taste of an apple is much weaker than the actual gustatory quality associated with the tasting an apple. More vivid ideas include the qualia associated with dreams.

Temporal qualia: While many of the qualia I have mentioned above give an impression of space, it is clear to me that I also experience, in my reality, the flow of time. This is provided by the immediate qualia of memory and anticipation, which respectively give the impressions of the past that has just gone and the future that is yet to come. These qualia give my reality a sense of continuity along what we think of as the dimension of time, rather than it being composed of fragmented and discontinuous bundles of perceptions.

Agential qualia: In addition to the temporal qualia of memory and anticipation, the dynamics of my actions and thoughts are accompanied by my experience of the fact that I control them. This sense of agency is associated with the concept of free will, which I discuss in greater detail in chapter nine.

As noted earlier, this list is neither exhaustive nor definitive. There are certainly many more kinds of quality that I have omitted from the above classification. There are also different ways of classifying the qualities that make up my experience. Furthermore, it covers only the qualities I am capable of experiencing in my current embodiment as a human organism. The possibility of other sorts of qualia available to other conscious beings has not been considered here. My aim, now, is to analyse how it is that my reality is constructed from these qualia.

The construction of reality

As I have noted, my reality is composed of intricate combinations of the above qualities that are experienced in the first-person existence that is my consciousness. With reference to how I experience these qualities, I partition my reality into my environment, my body, and my psyche. I now examine how these compartments of my reality are constructed from the aforementioned qualities.

If I focus on the qualities of visual, auditory, olfactory, gustatory, tactile, and thermal qualia, I notice that they are vivid and immediate. Furthermore, an impression of space is provided by the spatial qualities of the visual, auditory, tactile, and thermal experiences. The combinations of these qualia provide the structure of my environment. These combinations, however, cannot be random, because the structure of the environment, as I experience it, is ordered and coherent. There appears, therefore, to be some sort of pattern behind these combinations of qualia. Certain qualia in my visual space tend to concur with certain auditory, olfactory,

gustatory, tactile, or thermal qualia, forming unique combinations that tend to recur. Through my engagement with the world and my interactions with people who also engage with this world, it becomes apparent that some combinations are taken to be more salient than others. Some of these combinations are classified as and are taken to represent objects. For example, the dark brown patch in my visual space concurs with a fragrant olfactory quality. Also, upon my picking up and tasting it, it evokes a hot thermal quality and a bitter gustatory quality. I learn from people in my community of speakers and shared practices that this combination of qualia represents an item called a cup of coffee.

I also notice that the locations of some qualia in my visual space are correlated with the locations of qualia in my auditory and tactile spaces. Thus, objects in my environment are assigned locations in the dimension of space. For example, if a patch in the centre of my visual space concurs with an auditory quality in the centre of my auditory space, I can localise this object as being in front of me.

In addition to a spatial structure, my environment has dynamics, which are experienced as having an ordered continuity in the dimension of time. The qualia of memory and anticipation give me the impression that the reality that I experience is changing, for each moment of perception appears to follow from a preceding moment and to lead onto a succeeding moment. Thus, the reality I experience appears to be in a state of temporal change. Furthermore, this change appears to follow an ordered and predictable pattern, which I call the dynamics of my reality. I do not experience my reality as a series of discontinuous episodes, but as a continuous flow.

I experience, in this reality, a physical structure which I call my body. This body, I experience as belonging to me, for not only do I experience myself controlling its actions through the sense of agency, but many of the qualia that I have are strongly correlated with the stimulation of certain parts of my body. For example, visual qualia are associated with the presence of objects in front of my eyes, auditory qualia with the stimulation of the cochlear by sound waves, olfactory qualia with the chemical stimulation of the nose, gustatory qualia with the chemical stimulation of the tongue, and tactile qualia with an object's contact with my skin. Thus, since qualia are closely linked to my experience of the stimulation of certain parts of the structure I call my body, my environment appears to be experienced from the perspective of my body.

Other qualia I experience, such as visceral qualia and painful qualia, can also often be roughly localised in the space of my body,

but are not always associated with the stimulation of the body by objects in what I experience as the external world. Thus, these qualia are recognised as being associated with processes occurring in the space that I experience as my body. For example, the experience of pain is associated with what I experience to be potential or actual damage to my body, the experience of hunger is associated with the body's need for nourishment, the experience of orgasm is associated with sexual pleasure, and so on. From these experiences, and from the people in my community of speakers and shared practices who teach me about the meanings of these qualities, I develop a body image, which helps me interpret and navigate my bodily activity.

The vivid qualia that I associate with the external world and my bodily processes are what David Hume (1748) calls the impressions. In addition to these impressions, there are the duller and more indistinct qualia commonly referred to as ideas, as well as the prolonged and nonlocal qualia associated with affective states. These qualia are recognised neither as events located in the external world nor as straightforward bodily processes, but as the activities of something else, which can be called my psyche. Although these states are not experienced as being located in my environment, they sometimes influenced how I perceive my environment. For example, the state of intoxication from the consumption of alcohol is likely to dampen the impressions, whereas the state of stimulation from the consumption of caffeine is likely to make them more vivid. My psyche is also shaped and supported by the social context wherein I am embedded, including the developmental interactions with others that influence my dispositions and the cultural conventions that influence the affordances that are salient to me. Again, I experience my psyche as belonging to me, because I experience myself having some agency and control over its activity.

These compartments that I construct from the nature of my basic qualities make up what I call my reality. I construct an environment wherein physical objects manifest in the dimensions of space and time, a body which is part of this environment but is more closely linked to the qualities I experience, and a psyche which influences and is influenced by the activity of my body and the events which I experience in my environment. Accordingly, the three compartments are not independent, but are integrated. They interact in complex ways to form what I experience as my reality. Events I experience in the environment occasion changes in my bodily processes and my psychological states, my bodily processes respond to events in the environment and influence my psychological states, and my

psychological states influence how I experience and act on events in the environment and my bodily processes.

Intersubjectivity

So far in this chapter, I have presented a conception of reality as being based in the phenomena we experience. Consciousness is my first-person existence, and so it is what I experience within it that forms the basis of what I call real. One may, at this stage, ask the following question. If my reality is constructed from my experience, then what happens to the parts of the world I am not currently experiencing? I see a table ahead of me at this current moment. If I am to close my eyes, does the table lose its reality?

One may reply to this with reference to other consciousnesses. Although I am not experiencing the table, there are other consciousnesses experiencing it, and so it stays real in these other consciousnesses. Hence the table is real in an intersubjective reality. Indeed, Edmund Husserl, in his *Cartesian Meditations* (1931), emphasised that one experiences the world as being experienced by others. This is an promising solution, because it acknowledges some kind of congruence between the experiences of different consciousnesses, and so does not deny the subsistence of some kind of objective world that is responsible for this congruence. Furthermore, it acknowledges that consciousnesses other than mine also exist. As I proposed in chapter one, this objective world, itself, has no reality, but it is the potential that is realised and given quality when experienced by consciousness. Although I may not be experiencing the table at the moment, the objective potential that is realised as a table in my consciousness when I am experiencing it still persists, and so it may be realised as a table in the other consciousnesses that experience it. Thus, the table stays real in these other consciousnesses and its objective potential still subsists.

Although this solution considers the intersubjective reality of the table in other consciousnesses and the subsistence of its objective potential while I am not experiencing it, it says nothing about its subjective reality in my consciousness. Is the table still real to me when I do not experience it? I argue that it is. When I perceive the table, I am experiencing it as part of what I call the external world. Upon closing my eyes, I am not longer experiencing the table in such a way. However, it still manifests as some form of experience in the existence that is my consciousness. This experience is the idea of the

table. Having previously experienced the table through visual and tactile experiences, I form, in my mind, the idea of this particular table in this particular region in space. Due to memory and anticipation, this idea continues to subsist in my mind, even when I am no longer having the visual and tactile experiences of the table. Furthermore, due to the subsistence of this idea in my memory, I expect, upon reopening my eyes, to again have the visual experience of the same table. Thus, the ideas I form from prior experiences provide some continuity of reality during periods when I am not perceiving the object in the external world. In a way, they fill in gaps in my reality that are due to the constraints of my immediate senses.

Another reason to acknowledge an intersubjective reality is that it allows us to come to a consensus about what is veridical and what is nonveridical. If one's experience of the world is compatible with others' experiences of the same world, then there is reason to believe that it is veridical, but if one's experience of the world is incompatible with others' experiences of the same world, then there is reason to believe that it is nonveridical. For example, if a person develops the persecutory belief that I am a police officer who is trying to arrest the person, then we can conclude that the person's belief is false on the basis that I know that I am not a police officer and others around me know that I am not a police officer. Likewise, if the person claims that there is water flooding the room, then we can conclude that the person's claim is false on the basis that I and others around me perceive that there is no water flooding the room.

In psychiatry, a rigid belief about the world that occasions harm and is inconsistent with the intersubjective reality of others is often considered a symptom of psychosis. However, can we appeal to this intersubjective reality to dismiss the meaning of one's personal encounters with, for example, deceased beloveds or mythological characters in one's dreams or visions? I argue that we cannot. Whereas the persecutory belief described earlier is a false belief about the world that is inconsistent with others' perceptions of the same world, the phenomena in one's dreams or visions may not be interpreted as occurrences in this world, but as one's personal encounters with mystical phenomena that are not occurring in this world. Given that they do not represent occurrences in this world, they cannot be scrutinised with reference to others' perceptions of this world. And so, we cannot appeal to the intersubjective reality of this world to invalidate the individual's belief that these phenomena may be real in other ways, such as with respect to their meanings in the person's phantasy world or with respect to the roles that the

mythological characters have in our shared cultural discourses. There could even be metaphysically possible worlds where such mythological characters, as abstract individuals realised through patterns in our cultural discourses, are associated with consciousnesses. Accordingly, R. D. Laing (1967) and, more recently, Simon Dein (2004) note that mystical and spiritual phenomena can be meaningful and illuminating for the people experiencing them. This suggests that the criterion that determines whether a belief is pathological is not its truth value, but a value judgement about whether it has a harmful impact on the person's wellbeing and engagement with the intersubjectively shared world.

Importantly, this intersubjectivity does not amount to relativism, because there is still a crucial role for input from the world. Rather, it indicates a form of pluralism about truth, whereby different domains may warrant different criteria for truth. In the domain of philosophy, which is a discipline that aims for knowledge and understanding of what is foundational to existence, it is sometimes correct to accept a correspondence theory of truth, such as that proposed by Bertrand Russell (1912). For example, the philosophical proposition that dualism is true is an absolute truth that corresponds to the ontological fact that consciousness exists as a fundamental entity that is ontologically separate from the physical world. Here, the correspondence theory of truth is sound with respect to the ontological status of consciousness. Also, insofar as this domain concerns a concrete fact about ontology, it cannot be marked by the sort of dialetheic contradiction suggested by Graham Priest (1987). Dialetheism is false with respect to the ontological status of consciousness, because the ontological status of consciousness pertains to a concrete fact about the fundamental existence of a specific entity. Likewise, given that this domain concerns a concrete fact, a probabilistic analysis of verisimilitude such as that based on the theorem by Thomas Bayes (1763) is inapplicable. Such a probabilistic account of verisimilitude is false with respect to the ontological status of consciousness, because ontological status of consciousness pertains to a concrete fact and not to a probability distribution. Given that such a concrete fact about the fundamental existence of a specific entity is exclusively either true or false, it follows that if "dualism is true" is true, then it is exclusively and absolutely true. Of course, in other domains, such as cultural, scientific, and mathematical domains, there are certainly roles for coherence, pragmatic, and deflationary theories. Nonetheless, even here, truth is still importantly informed by input from

correspondence with facts about the observed world and intersubjective agreement.

An epistemology of science

Although what we access as reality is constructed from our subjective experiences, there is also a need to discuss how the objective world relates to our experiences. As I noted in chapter one, the fact that we have privileged access only to our experiences does not imply skepticism about the objective world. The objective world certainly does subsist on its own, for it is what induces the experiences in our consciousnesses. In chapter seven, I proposed that there is a psychophysical regularism between the physical events in the objective world and our subjective experiences. That is to say, our subjective experiences are realisations of the objective world.

Accordingly, I am not proposing a thoroughgoing rationalism that claims that the intrinsic nature of the objective world can be established independently of experience. Such a thoroughgoing rationalism cannot be correct, because our access to what we know as reality is ultimately experiential and intersubjective. Only our subjective experiences have actual qualities, for it is these that manifest in the existences that are our consciousnesses. Therefore, there is certainly a role for empiricism, whereby we learn about and form beliefs about the world through our experiences.

The above suggest the need for a sort of synthesis of rationalism and empiricism. Such a synthesis is provided by Immanuel Kant (1781), whose system of transcendental idealism distinguished the noumena of the external world from the phenomena of our experiences. Drawing on this synthesis, I propose that the objective world subsists on its own, but on its own it is not like anything. On its own, the objective world is only a mere potential. Only when realised as experience in consciousness does it gain any reality and become like something. That is to say, consciousness, through the act of experience, gives the potential of the objective world quality and brings it into reality.

Earlier in this chapter, I mentioned that due to my experiences of memory and anticipation, my experience of the structure and dynamics of this reality is such that they form regular and ordered patterns. These patterns tend to recur consistently. For example, objects tend to fall towards the ground when dropped, salt tends to dissolve when placed in water, wax tends to melt when warmed, and

so on. Thus, there appears to be an underlying pattern to the dynamics of my reality. Furthermore, since the reality I experience is a realisation of the potential of the objective world, it is reasonable to assume that these patterns are derived from this potential.

Our method of scientific enquiry involves the detailed analysis of these patterns. Laboratory experiments are performed wherein these patterns are tested and observed, inductive inferences are made wherein the patterns observed in the empirical data are generalised to predict the properties of unobserved cases, the generalised patterns are described and communicated in a mathematical language, and theories are constructed to account for the patterns described. This enquiry serves to facilitate our perception of and predictions about the reality we experience. Rather than viewing reality as a series of coincidences, science constructs a comprehensible framework through which our reality can be interpreted.

Do the theories that we construct aim to be literal representations of the objective world behind what we experience? Scientific realism suggests that they do. However, I argue that this is incorrect. Theories cannot be literal representations of the objective world, because the objective world, on its own, is not like anything. Therefore, it would be misleading to suggest that the objective world is actually composed of the theoretical concepts that we postulate, such as particles and forces, because the objective world, on its own, has no reality. It subsists only as a potential, which is only realised and given quality through experience by consciousness.

What epistemic roles do our scientific theories have, if they are not literal descriptions of the nature of the objective world? Given that the features that we experience are realisations of the potential that is the objective world, it is reasonable to assume that our theories about these features do provide at least some sort of knowledge of the objective world. However, the data on which our theories are based do not come directly from the objective world, but from our experiences of it. After all, as noted above, the objective world, on its own, is a mere potential. Only our realisations of it in our experiences have the substantial qualities that can inform our theories. Therefore, our scientific theories are not literal representations of the objective world itself, but are models that help us to interpret, explain, predict, and intervene on the features that it occasions in our experiences.

This suggests that a key role of scientific theories is pragmatic. Theories do not have to reflect any metaphysical reality, but are constructs that account for and organise the features we experience.

They are not even literal accounts of these features. That is to say, theories are not mere descriptions of the features that present in our experiences, but are generalisations and inferences that we make about these features and the relations between them. As I have argued, the features that we know to be real are what manifest as experiences our consciousnesses. The theoretical concepts that we postulate in science do not fall into this kind. As David Hume (1748) notes, when one observes a causal interaction between two objects, one perceives a succession of events, but one does not perceive any necessary connection linking these events. There is no distinct impression that corresponds to causation, except for the expectation that a particular event will be followed by another. Hence, a theoretical concept, such as causation, is not a mere description of a feature that is experienced, but is a construct that is inferred from what is experienced.

Even methods whose reliability we take for granted, such as induction, involve unjustified assumptions being made. I have observed so far that the sun has risen every day and, from this, I assume that the sun will rise tomorrow. What justification do I have for my assumption? I have not yet observed whether the sun will rise tomorrow or not, and so it is not entailed by my immediate experience. Furthermore, there is no logical entailment from the sun having risen previously to the sun rising tomorrow. Although the sun has risen every day so far, it is logically conceivable that the sun will not rise tomorrow.

One might try and justify induction by claiming that whenever I had previously induced that the sun would rise the following day, my experience of the sun actually rising the following day confirmed the validity of my induction. Hence, I continue to assume that the sun will rise tomorrow, because this assumption has been correct so far. However, Hume argues that this is inadequate. This is because the statement, "induction has been correct in the past, and so it will be correct in the future", is itself an inductive inference. Such a statement is trying to use induction to justify induction, and so is circular. Therefore, Hume's argument seems to suggest that we lack a global justification for induction.

If there is no global justification for induction, why do I continue infer that the sun will rise tomorrow rather than assume the frightening alternative that it will not? According to Hume, it is due to custom or habit. While we may lack a global justification for induction, we are inclined to form associations and generalisations after observing repeated events. Moreover, from a pragmatic

perspective, such associations and generalisations are instrumentally valuable. I infer, from my observation that the sun has risen every day, that the sun will rise every day, because this generalisation is useful for navigating myself in the world. The act of making sense of the world is facilitated by the use of induction.

This also applies to theory formation in science. Traditionally, theory formation is suggested to involve two key steps. The first step is the gathering of empirical data acquired through observation. The second step is the inferential process of constructing the theory that accounts for this empirical data. Of course, this is a simplification, as the theoretical assumptions influence what data are considered to be salient and how the data are interpreted, but it is still a useful characterisation of the process of theory formation. Importantly, theory formation is generally a nondeductive procedure. The theory is not entailed by the data, but goes beyond the data by positing further theoretical concepts in order to account for the data. This involves abductive reasoning, or inference to the best explanation.

An implication of the above is the underdetermination of theory by data. Because the empirical data does not entail the theory, there could be several different theories that all successfully account for the data. Accordingly, the scientist cannot select one theory from other empirically equivalent theories based solely on the empirical data, because the data is successfully accommodated by all the theories. Rather, further criteria are required to select one theory over others. These are known as superempirical virtues.

In *The Scientific Image* (1980), Bas van Fraassen argues that these superempirical virtues that are used to select one theory over other empirically equivalent theories do not reflect how true the theory is, but reflects its pragmatic utility. Thus, van Fraassen suggests that the choice between empirically equivalent theories is based more on utility than on truth.

This becomes apparent when we consider some of the superempirical virtues that determine theory choice. For example, simplicity is obviously a criterion that influences theory choice. For example, if a physicist is considering two theories, one which is mathematically simple and the other which is mathematically complex, then the physicist is likely to choose the former theory, provided that both are empirically adequate. Another example, presented by William Lycan (1998), is the practice of fitting a curve on a graph. If provided with a set of data points that lie approximately along a straight line, it is often considered better practice for one to draw a straight line through them, rather than to

draw a convoluted curve that meets all the data points. Clearly, pragmatic considerations influence this practice. A straight line is chosen, even though it does not meet all the data points, because it is quicker and easier to interpret, draw conclusions from, and make predictions from a straight line than from a convoluted curve. This reflects the aim for pragmatic utility in science.

A scientific realist may respond to this by suggesting that simpler theories tend to be more successful than complex theories and that this success indicates that these simpler theories are closer to the truth. However, van Fraassen notes that truth is not necessary for success. Rather that the success of a theory is measured by its ability to accommodate and account for the empirical data. That is to say, empirical adequacy is the mark of a theory's success, but empirical adequacy does not require the theory to be a literally true representation of the objective world.

Furthermore, there is the question of what comprises simplicity in a scientific theory. Scientists are often said to favour simpler theories, but what is defined as simple may vary from one scientific practice to another. For example, in one domain simplicity may involve drawing straight lines on graphs, whereas in another domain it may involve postulating the fewest possible number of variables. Different ideas of simplicity may appear quite distinct from one another, and so there may be no single criterion that makes a theory simple. Rather, I argue that simplicity may, in fact, be defined by pragmatic considerations, which may vary across different domains and contexts. Therefore, simplicity is itself a pragmatic virtue.

Another superempirical virtue criticised by van Fraassen is explanatory power. This is a virtue that is often placed in priority of others in the process of theory selection. One theory is considered to be better than another if it explains the data in a more comprehensible way. Indeed, as Peter Lipton (1991) notes, the "loveliest" explanation is "the one which would, if correct, be the most explanatory or provide the most understanding". Again, van Fraassen argues that such explanatory power is a virtue only for pragmatic and aesthetic reasons, in that it makes a theory more useful and attractive, but not necessarily truth conducive. A theory which explains the data more thoroughly is favoured over an empirically equivalent theory which explains the data less thoroughly, because the former theory is pragmatically more useful and aesthetically more satisfying.

Characteristic of an explanation are the unification of data and the communication of the outcome. By explaining something, one

accounts for it in terms of other ideas and expresses it in a coherent manner. This, I argue, has a pragmatic goal. Unification attempts to link together potentially unrelated ideas, and so aims to achieve parsimony and simplicity. Communication attempts to express ideas in a coherent manner, and so, again, aims to achieve pragmatic utility through simplicity.

I have, therefore, in this chapter, endorsed a modest antirealist position that scientific theories are not representative of objective truth, but provide a means of accounting for data that we observe in a convenient and comprehensible way. The virtues that influence the choice between empirically equivalent theories are not truth conducive, but pragmatic. Nevertheless, I would like to stress that this does not, in any way, make them any less legitimate as reasons for choosing a theory. The theory's reception, understanding, and application are greatly aided by these virtues. Theories, while they may be unable to provide us with a literal account of objective truth, serve to facilitate our judgements about the world, and so pragmatic utility is an important theoretical virtue.

This antirealism does not in any way undermine the epistemic value of science. As noted above, our scientific theories are highly valuable tools for interpreting, explaining, predicting, and intervening on events and patterns in the reality we observe. Accordingly, the acceptance of science is epistemically warranted. Science denialism is unsound and ought to be rejected. Nonetheless, while these theories are useful and reliable models that account for the features we experience, they do not provide us with literal representations of the intrinsic natures of these features. Science aims to facilitate our judgements about the world we experience by organising them under a comprehensible framework, and so it is unsurprising that the theories that are parsimonious and comprehensive tend to get selected.

This antirealism also does imply skepticism about the objective world, for our experiences are realisations of this objective world. However, it is only as experience in the first-person existence of consciousness that the objective world has any phenomenal reality. On its own, the objective world subsists only as a noumenal potential. Our theories, therefore, are not direct representations of the objective world. Rather, out theories are tools which we use to describe, explain, predict, and intervene on the patterns apparent in our experiences of the objective world.

IX
Free Will

I noted, in chapter eight, that I experience the dynamics of my reality as following ordered and intelligible patterns. Furthermore, since the reality that I experience is a subjective realisation of the potential that is the objective world, it is reasonable to assume that these patterns present in my reality are derived from this objective potential. From these recurrent patterns that I experience, theories are constructed, which aim to account for these patterns. These theories do not, as I have argued, aim for a literal representation of the objective world, for the objective world on its own has no reality. It is a potential that is only realised as experience in the first-person subjective existence of consciousness. Nonetheless, our scientific theories provide reliable models which characterise these patterns in a comprehensible manner and, furthermore, are capable of powerful predictions of the dynamics of the reality we experience.

These theories sometimes postulate laws, which are theoretical statements about the patterns in our reality that are deemed to be projectable. They are inductive generalisations about the dynamics of the world communicated within theoretical frameworks. For example, the laws of motion in classical physics are generalisations about the movements of objects that are communicated within a theoretical framework built from concepts such as masses and forces. Likewise, the laws of thermodynamics are generalisations about heat transfer that are communicated within a theoretical framework built from concepts such as energy and entropy.

Given the explanatory and predictive successes of our scientific theories, it appears that the patterns that I experience in my reality follow the laws of nature that are postulated by these theories. Furthermore, I experience my body as part of this reality, and so it also follows these laws. This seems to suggest that all the choices I make, whether they are as trivial as choosing between red apples or green apples in the market or as significant as choosing to study philosophy as well as medicine at university, can be explained in terms of physical processes that follow the laws of nature.

Yet, from my direct acquaintance with myself as a conscious agent, I have the unequivocal sense that I have free will. I am the controller of my thoughts and actions. This presents a problem. The proposition that my behaviour is determined by physical laws

appears to contradict my experience that I am the agent who controls my behaviour. However, I argue that there is no contradiction here. Free will is entirely compatible with the laws of nature. In this chapter, I defend a form of compatibilism, which argues for a strong kind of free will, while conserving these laws of nature.

Against determinism

In classical physics, from the mechanics of Isaac Newton (1687) to the relativity of Albert Einstein (1916), the laws of nature tend to be posited as being strictly deterministic. They completely and unequivocally determine the dynamics of the physical matter of the universe. Hence, once the initial conditions of a physical system have been established, the rest of its history follows inevitably. Strong determinism is often associated with Pierre-Simon Laplace (1814), who suggested that complete knowledge of the present conditions of all the particles in the universe could make it theoretically possible to calculate all their past and future conditions.

The kind of free will proposed by René Descartes (1641) is one wherein the immaterial mind causally interacts with the physical world. This interactionist free will seems to be incompatible with determinism in classical physics. If the laws of nature are deterministic, then the physical world is causally closed. There would be no room for an immaterial mind to influence the course of the physical world. Thus, under classical physics, interactionist free would violate the laws of nature. For this reason, the neurophysiologist Roger Carpenter (1997) defends a form of dualism he calls "one-way Cartesianism", which accepts that consciousness is a separate entity from the physical world, but suggests that it does not exert a causal influence on the physical world.

The deterministic picture of the universe presented by classical physics was challenged by the emergence of quantum mechanics. Under quantum mechanics, the state of a physical object could no longer be described in definite terms, but only as a probabilistic superposition of states, or a wave function. There is a deterministic component of quantum theory, namely Erwin Schrödinger's wave equation, which predicts how the wave function evolves. However, the state of the object is still a probabilistic superposition of values within this wave function, and not a definite location in space. Thus, under this framework, the world is no longer deterministic, but indeterministic.

Even under classical physics, strong determinism had its flaws. Although one could compute the conditions of a particular physical object, one cannot compute the complete conditions of the world, because one is also part of this world, and cannot compute one's own conditions. However, under quantum mechanics, it appears that one cannot even compute the conditions of individual objects, for these objects are in an indeterminate superposition of states. And so, uncertainty appears to be an integral property of the physical world.

In *The Self and Its Brain* (1977), Karl Popper and John Eccles suggest that quantum indeterminism could provide room for interactionist free will. They endorse a form of dualism, whereby the immaterial mind causally interacts with the physical world. While this is a possibility, it is somewhat tenuous. Indeed, the developments in quantum mechanics may indicate that determinism is false, but indeterminism does not necessarily entail free will. For example, indeterminism could just indicate that the laws of nature at a subatomic level are partly influenced by random chance. Just because the universe is not deterministic, it does not mean that an immaterial mind influences its course. Furthermore, the suggestion that the mind can only influence the physical world at a subatomic level seems somewhat inadequate. If free will only obtains at this microscopic scale, then it is a very weak kind of free will indeed.

The above suggests that the indeterminism of the universe, as proposed by quantum mechanics, may not be the place to look for free will. In spite of quantum indeterminism at the microscopic scale, I still experience many of the dynamics of my macroscopic reality as following ordered and predictable patterns. Thus, the dynamics of the world still appear to follow the laws of nature. Quantum indeterminism, at most, indicates that these laws are not deterministic, but are probabilistic. While this may open up the possibility of libertarian free will, it does not secure it.

Phenomenal judgements

It has also been suggested that free will might be found in our phenomenal judgements. These are the judgements that we make about our subjective experiences, such as, "it is true that consciousness exists", "scientific explanation cannot capture the phenomenal redness of a red experience", or "consciousness is an ontologically separate entity from the physical world". The physicist Avshalom Elitzur (1989) has argued that our ability to report our

subjective experiences shows that they have a causal influence on our behaviour, and that this demonstrates a kind of interactionist free will. However, David Chalmers (1996), suggests that phenomenal judgements present some kind of paradox.

The paradox arises because subjective qualities cannot be reductively explained, yet our phenomenal judgements about these subjective qualities are behavioural acts, and so should be reductively explainable in structural and dynamical terms. How can it be that subjective qualities are not reductively explainable but our claims about them are? Does this suggest that our subjective qualities are explanatorily irrelevant to our claims about these subjective qualities? One solution would be to accept Elitzur's interactionist free will and suggest that our subjective qualities do have causal influences on our behaviours.

While this is a possibility, it does undermine a central claim of the conceivability argument, as presented in chapter four. According to the conceivability argument, there is no logical contradiction in the idea of a nonconscious physical replica of me, because the physical facts of a system do not entail the presence of consciousness. Since my zombie twin and I are physically indistinguishable, our behaviour and physiology will be indistinguishable. However, if we assume the interactionist free will as proposed by Elitzur, then my zombie twin and I will not display the same behaviour. My behaviour would be partly influenced by subjective experience, whereas my zombie twin's behaviour would not. Thus, it appears that interactionist free will might be incompatible with the conceivability argument.

Another option is to accept that subjective qualities are explanatorily irrelevant to the contents of phenomenal judgements. For some judgements, this may seem plausible. For example, one does not need to refer to the subjective quality of red to explain the judgement "this apple is red". Such a judgement can be explained by cognitive psychology with appeal to awareness, perception, and reportability. Sensory information from the retina is processed by the brain and this results in the output of an appropriate verbal response. Subjective experience does not need to be invoked such an explanation, and so it is conceivable that my zombie twin could make such an utterance when confronted with a red apple.

However, for other judgements, such as, "no amount of explanation can capture the phenomenal redness of a red experience", or "consciousness is an ontologically separate entity from the physical", it is more difficult to see how subjective qualities can be explanatorily irrelevant to them, since these judgements

appear to actually be about the subjective qualities, or even about consciousness itself.

Nevertheless, Chalmers suggests that these judgements may still be explained causally without any appealing to subjective qualities. He proposes that we imagine a nonconscious perceptual system that can distinguish between colours. Assume that the system is capable of introspection and has access to information about its own state, but not about its lower-level parts. Hence, it can tell what informational state it is in, but it cannot tell what constitutes its informational states, much like how a human with no knowledge of neuroscience can distinguish one colour from another, but may be unaware of the neural mechanisms underlying colour perception.

Upon showing the system a red object, it will report that it perceives red. However, when asked how it knows that it perceives red other than blue, the system cannot answer. Indeed, from our point of view, we know that red throws the system into one informational state and blue throws it into another, but, since the system itself has no access to the details of what constitutes its informational states, it cannot describe how red is different from blue. It can merely distinguish them. Thus, from the system's outlook, it is just a brute fact that red is different from blue. It may even describe them as simply "feeling" different. If we were to grant the system an even higher cognitive capacity, it may begin to wonder why red "feels" one way and blue another, and it may even refer to these "feelings" as "phenomenal qualities".

This scenario shows how a system that lacks consciousness altogether may still be capable of making phenomenal judgements. Our subjective qualities, it appears, may be explanatorily irrelevant to our judgements about them. In fact, as shown by Chalmers' hypothetical situation, a nonconscious system's judgements about what it calls "consciousness" are not about consciousness at all. Rather, they are about the properties of the system's informational states which are inaccessible to it.

However, this presents another peculiar situation. Are our phenomenal judgements any more justified than a zombie's? My zombie twin lacks consciousness altogether and has no knowledge of consciousness. Its judgements about "consciousness" are not about consciousness at all, but are due to a limitation in its capacity for self-awareness. My zombie twin is nonconscious, and so there is no possibility of consciousness having a role in the mechanisms that underpin the generation of its phenomenal judgements. If we assume the causal theory of knowledge, which suggests that for a belief to

constitute as knowledge about a feature there must be a causal connection between the feature in question and the formation of the belief, then it follows that my zombie twin's judgements about "consciousness" are unjustified, because there is no consciousness to have a causal role in the generation of its judgements.

Now, consider the fact that my zombie twin is physically indistinguishable from me. From this, it follows that the same underlying mechanisms that are responsible for the generation of my zombie twin's phenomenal judgements are also responsible for the generation of mine. It also follows, from this, that my consciousness has no causal role in the generation of my judgements, and so, according to the causal theory of knowledge, my phenomenal judgements are as unjustified as my zombie twin's.

However, I argue that my judgements about consciousness are justified on the basis that the causal theory of knowledge is not applicable to consciousness. As noted in chapter one, consciousness is ontologically and epistemically unique. Due to its irreducible first-person ontology, it cannot be accessed objectively in the same way that third-person features can. Yet, I know with absolute certainty that it is true that consciousness exists, because my consciousness is my very first-person subjective existence. My knowledge of it does not depend on any causal connection between it and the formation of my belief, but rather my knowledge of it is secured by my direct acquaintance with it. My consciousness is what I *am*.

Therefore, despite the fact that the same sorts of mechanisms that are responsible for my zombie twin's phenomenal judgement are also responsible for mine, the fact that I have certain knowledge of consciousness through direct acquaintance means that my phenomenal judgement is true and justified, whereas my zombie twin's is not. That is to say, we could say that a nonconscious zombie's judgement about what it calls "consciousness" is merely an empty claim about an inaccessible informational state, but we cannot say this about a conscious being's judgement about consciousness. Although the judgements in both cases are formed through the same sorts of mechanisms, the conscious being's direct acquaintance with consciousness verifies and justifies the conscious being's judgement about consciousness, and so it can actually be interpreted as a genuine claim about consciousness.

One may object to this on the basis that it seems suspiciously coincidental that a judgement, in whose generation consciousness had no explanatory role, can actually be about consciousness. However, I argue that this is not necessarily a coincidence, because

the mechanisms that are involved in the producing phenomenal judgements may also be related to the mechanisms that are correlated with conscious experiences. After all, many of our experiences appear to coincide with certain informational states. We experience red qualia when we process information about a red object and blue qualia when we process information about a blue object. Furthermore, phenomenal judgements appear to be claims about the same informational states. We talk about red qualities when we process information about a red object and blue qualities when we process information about a blue object. Thus, the same informational states that are involved in generating phenomenal judgements are those that are correlated with subjective qualities. Whereas my zombie twin's neural activity is not accompanied by consciousness, my neural activity is accompanied by consciousness, and so the presence of subjectivity in my case verifies my phenomenal judgement as a judgement not only about an informational state but also about the subjective quality that is correlated with the informational state, whereas the absence of subjectivity in the case of my zombie twin allows us to interpret its judgement only as being a claim about the informational state. Although our utterances may sound the same and be produced through the same processes, the presence of an extra phenomenal realm in my case enables my utterance to possess a further level of meaning specifically about this phenomenal realm.

What I hope to have shown, in this section, is that the generation of a phenomenal judgement may be causally explained without any reference to consciousness. This is despite the fact that the phenomenal judgement may, in the case of a conscious being, actually be both true and justified. Nonetheless, I have suggested that phenomenal judgements do not necessarily provide evidence for interactionist free will. As shown by Chalmers' hypothetical scenario, a structural and dynamical explanation of how phenomenal judgements are produced is possible. It appears, therefore, that we ought to look elsewhere to understand free will.

Laws of nature

As noted at the beginning of this chapter, the problem of free will arises because of two seemingly conflicting premises. First, my behaviour follows a set of physical laws. Second, I am the agent who freely chooses my behaviour.

Traditional interactionist free will is problematic because it denies the first premise. The suggestion that an immaterial mind has a free causal influence on physical processes seems to assume that it can violate of the laws of nature. Deterministic incompatibilism is also problematic because it denies the second premise. My aim for the rest of this chapter is to defend a compatibilist theory of free will that denies neither premise, but accepts both premises in a noncontradictory fashion.

How can these two premises be synthesised? In the epilogue to *What is Life?* (1944), Erwin Schrödinger proposes that they can be made compatible by suggesting that we are the authors of the laws of nature which our actions follow. We, as the infinite plurality of consciousnesses, choose our actions and these actions become engrained in the natural history of the universe that the laws of nature describe. This is not a form of traditional interactionism whereby the mind causally influences the physical world, but is a stronger form of free will whereby consciousness occasions the behaviour that becomes part of the reality that is described by the laws of nature.

According to the view known as necessitarianism, the laws of nature govern the universe. That is to say, they dictate necessarily how the universe must operate. This view appears to suggest the presence of some kind of physical necessity or necessary connection between events in the objective world. It suggests that things are the way they are because the laws of nature dictate that they have to be the way they are. However, many find necessitarianism unsatisfactory and, instead, accept regularism, which is the view that laws of nature are not necessary properties of the universe, but mere descriptions of it. That is to say, laws do not govern the course of nature, but describe it.

As David Hume argued in his *Enquiry Concerning Human Understanding* (1748), we experience, in this world, events following one another in a regular fashion, but we experience no necessary connection between these events. That is to say, we observe causes and their effects, but we do not observe any glue connecting the two. Thus, the idea of a connection between cause and effect does not come from anything we observe in the external world, but it comes from our own minds. Hume proposed that the impression that this idea originates from is the feeling of anticipation we have, upon observing a cause, for its effect to occur. What Hume's argument shows is that we really have no justification for believing that there are any physically necessary laws out there, since

we do not experience any necessary connection out there, and the idea of a connection which we do experience originates in the mind. Necessitarianism, it seems, is making an unsubstantiated claim about the world by proposing the presence of physically necessary laws.

Further to the above, I argue that necessitarianism is undermined by the fact that we can readily conceive of modal variation between different parameters. In chapter three, chapter four, and chapter seven, I discussed the conceivability of modal variation between physicality and phenomenality. It is in virtue of this conceivable modal variation that any form of monism is false with regard to consciousness, as it shows that physicality and phenomenality come apart metaphysically. Here, I propose that there is also conceivable modal variation between different parameters within the physical domain. Indeed, many insights in science have been yielded by asking how things might have been had the values of their parameters been different. For example, David Chalmers (1996) notes that we can conceive of a world wherein the law of gravitation is different, such that a stone moves upward instead of downward when one lets go of it. The fact that such counterfactual reasoning is possible indicates that the relevant relations are contingent. Therefore, necessitarianism regarding laws of nature is false, because it fails to account for the conceivability of such modal variation.

Given that necessitarianism is unsound, a move to regularism is warranted. According to regularism, laws of nature are not prescriptions, but are descriptions of the universe. For every occurrence in the universe, there will be a possible factual description of it. Some of these descriptions are generalisations about regularities in the universe that are informed by and, in turn, inform inductive inferences. A subset of these generalisations that have roles in scientific practice are considered to be laws. Hence, the state of affairs is the inverse of what is suggested by necessitarianism. Events do not follow a regular pattern because they are governed laws. Rather, events can be described by laws because they happen to follow a regular pattern.

It is important to note that regularism does not undermine the semantic meaning or the epistemic value of a law. For example, the law of gravitation should still be interpreted as stating that any two objects with mass exert a gravitational force of attraction on each other. Furthermore, this law can inform inductive inferences and predictions about the dynamics of a system. However, this is not because it is a physically necessary rule which governs the course of nature, but because it is a description of a regularity that has been

observed to occur reliably in the course of nature. Every event that occurs becomes engrained in the course of nature, and so can be analysed under the framework of laws that we use to describe the course of nature.

Of course, the data we empirically observe may not accord exactly with what the laws predict. As Nancy Cartwright (1983) notes, the laws we infer are idealisations that abstract away the messy details of the actual events that we observe to occur. Hence, they are not empirically adequate unless they employ *ceteris paribus* conditions and further *ad hoc* assumptions. This suggests that the laws of nature are not exceptionless rules that govern events in the world, but are imperfect generalisations about the regularities we observe in the world.

Compatibilism

In *The Concept of a Physical Law* (1985), Norman Swartz proposes that regularism can help to resolve the problem of free will. According to Swartz, if we consider natural laws as descriptions of the course of nature rather than prescriptions, then free will can be compatible with determinism. This, I argue, is the position we ought to accept in order to acknowledge the reality of free will while taking seriously the laws of nature. According to necessitarianism, the laws of nature govern the motions of my body, and so the idea of free will is not tenable. However, under regularism, the laws of nature only describe the motions of my body. This allows my free will to be the phenomenon that actually governs the motions of my body. I control the contents of my thoughts and the motions of my body, whereas the laws of nature describe them. Therefore, regularism allows us to endorse a form of compatibilism, whereby free will can be acknowledged and the laws of nature can be conserved.

To see how the idea of free will can be made compatible with the notion of physical determinism, consider the following proposal. I can choose to do something or I can choose not to do something, but whatever I choose to do becomes engrained in the natural history of the universe. The laws of nature describe what happens in the natural history of the universe. Hence, the laws of nature accommodate whatever I choose to do. This also indicates that the laws of nature depend on what we choose to do. The actions we choose to perform become events in the natural history of the universe and the laws of nature are formulated to describe the natural history of the universe.

The above seems to suggest that one can come up with an empirically adequate scientific explanation for my behaviour, because my behaviour manifests as a structural and dynamical feature of the physical world that falls within the scope of scientific enquiry. Furthermore, one may use these scientific resources to predict my future behaviour. However, this does not undermine the fact that I am the agent who wills these actions. It is just the case that this "I", or my consciousness, is not part of the physical world, and so its free will cannot be observed scientifically. All we can observe are the consequences of its free will, namely my thoughts and actions, which do manifest themselves in the physical world as structural and dynamical events.

Given that our thoughts and actions are part of the observable physical world that falls within the scope of scientific enquiry but the subjective selves that will them are not, it is understandable that our scientific explanations of these thoughts and actions be purely physical, without any reference to our free will. For example, consider that I shine a light on a *Calliphora* larva, which results in it increasing its rate of movement and orientating its direction of travel away from the light source. Upon repeating the experiment several times, I find that the larva always displays this same kinesis and negative taxis in response to light. Having observed this, I formulate a law, "a *Calliphora* larva always reacts to light by speeding up and by travelling away from the light source".

Furthermore, I can explain, in physical terms, why this behaviour occurs. A causal explanation can appeal to the neural mechanisms that are involved in modifying the muscular activity of the larva in response to the detection of light by photoreceptors, while a functional explanation can appeal to the fact that moving quickly away from a light source into the dark makes the larva less vulnerable to danger. Free will does not appear anywhere in these scientific explanations and nor should it, because the subjective self that possesses free will is not part of the physical world that our scientific theories aim to describe. Also, from what I have observed so far, the law formulated has generally been correct, insofar as the larva has always reacted to light by displaying kinesis and negative taxis. Thus, the larva's behaviour is highly predictable. However, this does not mean that it does not have free will. Indeed, our scientific explanations and laws can account for its observed behaviour, but it is the larva's free will that directs it. After all, laws are just descriptions of the course of nature. So far, the larva has always reacted to the light by displaying kinesis and negative taxis,

but this is not because of any physical necessity that dictates that the larva must behave in this way. In a counterfactual scenario, the larva might have behaved otherwise. Rather, the larva can choose whether to react in such a way or not. There is no physical necessity that dictates this decision. However, in this case, it so happens that the larva has always chosen to react in such a way every time, and may continue to react in the same way in the future. The law presented earlier merely describes this pattern correctly.

The above also holds not only for a *Calliphora* larva, but for other conscious beings, including human beings, chimpanzees, octopuses, and so on. Consciousness wills events to occur. Because these events are part of the physical world, they can be analysed scientifically. However, the entity that wills them is not part of the physical world, and so will not feature these analyses. A scientific analysis of these events can only yield a structural and dynamical description of them, because only the structural and dynamical aspects of these events can be observed. Accordingly, what we observe as a biological mechanism, the larva experiences as the exercise of free will.

Recall that under the regularistic compatibilism that I am defending, the laws of nature are not prescriptions, but are descriptions of the course of nature. They do not govern our actions, but describe them. It follows, from this, that our actions are not consequences of these laws, but that these laws are consequences of our actions. We will events to occur, these events manifest as events in the physical world, and laws are formulated to describe these events. Thus, laws are mere portrayals of the structural and dynamical manifestations of our freely willed actions.

What this suggests is a strong kind of free will that is unconditioned by physical necessity. Our laws do not determine our actions, but rather the laws are derived from our actions. A framework of laws is not a set of instructions dictating the course that our actions must follow, but a set of descriptions recording the course that our actions do follow. Hence, we cannot violate the laws of nature, but this is not because the laws constrain our actions. Rather, it is because the laws accommodate our actions.

Although it is being presented here as a form of compatibilism, the account I am advocating could also be taken to be consistent with a form of libertarianism, insofar as it acknowledges that free will is unconstrained by physical necessity. Consciousness intervenes freely on the physical world. Whether this amounts to compatibilism or libertarianism depends on whether the subsequent description of the physical world is deterministic or indeterministic. If the model is

deterministic, then this can be considered compatibilism. If the model is indeterministic, then this can be considered libertarianism.

As noted above, this account explains why free will does not feature in our scientific analyses of behaviour. Each and every individual, from the infinite plurality of consciousnesses, influences the objective world by exercising free will. As noted earlier in this book, the objective world is no more than a potential that is only given shape when realised as experience in consciousness. Consciousness, by exercising free will, influences the potential of the objective world and alters the way it is realised in our experiences. However, because consciousness exists separately beyond the objective world, it is not realised alongside the objective world to form part of the reality we experience, and so it does not feature in any of our scientific analyses of this reality. All that we experience, and hence all that is accessible to our scientific enquiry, are the structural and dynamical events in nature that are the consequences of freely willed actions on the objective world.

It should be noted that the position I am advocating does not undermine the central claim of the conceivability argument in the way that traditional interactionism does. According to traditional interactionism, the immaterial mind has a causal influence on physical processes. This suggests that despite my being physically indistinguishable from my zombie twin, I would behave differently to it in any given situation, since I possess a mind that can influence my behaviour, whereas my zombie twin does not. And so, the conceivability argument would seem to be undermined.

However, if we accept the regularistic compatibilism I am advocating, this contradiction does not arise. The problem with traditional interactionism is that its proposal that the mind has a free causal influence on physical processes suggests the violation of physical laws. It suggests that the laws hold strictly in the case of my nonconscious zombie twin, but not in the case of me as a conscious being. With the position I am advocating, there is no such violation of the physical laws. Since we are physically indistinguishable, my zombie twin and I would both operate in the same ways according to the laws of nature, despite the fact that I possess free will. Therefore, we can take it as true that the conceivability argument is sound.

This may seem peculiar, but it can be understood as follows. According to regularism, the laws of nature are not prescriptions, but descriptions of the occurrences in nature, and so their formulation is based on our observations of the occurrences in nature. For example, the formulation of the laws that describe my behaviour is based on

the observations of my behaviour, very much in the same way that the formulation of the law, "a *Calliphora* larva always reacts to light by speeding up and by travelling away from the light source", is based on the observations of the *Calliphora* larva's reactions to light. Furthermore, as I have already said, only the structure and dynamics of my physical body are described by these laws, since only these structures and dynamics are observed in the physical world. Due to its not being part of the physical world, my consciousness cannot be observed, and so does not feature in the laws about my behaviour. Hence, these laws are essentially descriptions of the structure and dynamics of the physical matter that constitutes my body. Regarding the conceivability argument, these physical laws, whose formulation was based on the observations of my behaviour, are applied to the hypothetical example of my zombie twin.

Before I finish, the following objection must be addressed. If my free will is unconditioned by physical necessity, then why is it that I cannot perform supernatural feats at will, such as teleportation, levitation, or telekinesis? My inability to perform these feats seems to suggest that my actions are constrained by the laws of nature.

In reply, I accept that my choices are constrained by circumstances outside my control. I cannot teleport or levitate at will. However, this is not because of any metaphysical necessity regarding the laws of nature. Rather, it is because of the constraints on action that are set, first, by intersubjectivity and, second, by embodiment.

With respect to intersubjectivity, each and every one of the infinite plurality of conscious subjects is an agent with free will. Hence, in addition to the influence of my free will, other conscious subjects also influence the world with their freely willed actions. Furthermore, the free will of any given conscious subject is constrained by the freely willed actions of other conscious subjects. Consider the *Calliphora* larva that chooses to move away from the light. While I can influence the movement by picking up the larva or by changing the orientation of the light source, my influence on the outcome is constrained by the larva's own free will. I cannot, for example, make the larva decide to creep towards the light instead, nor can I make the larva suddenly grow legs and learn to walk. And so, the scope of my free will requires me to acknowledge that others also have wills that are free.

With respect to embodiment, the form that one's interface with the world takes constrains how one acts upon the world. At present, I assume the embodied perspective of a person and I experience the rest of my reality from the perspective of this embodied perspective.

That is to say, this embodied perspective acts as a psychophysical interface with the experiences in my consciousness. In virtue of this embodied perspective, limits are set on the way I experience the world. For example, I can only experience the qualities that are correlated with the spatiotemporal happenings in my body. Thus, my access to the objective world is confined to the small part of it that is realised as my experience and, as a consequence, so is the action of my free will. I can only influence what I can access.

Moreover, from this perspective, I influence the world by exercising my free will, but the way in which I influence it constrains my further activity. As an analogy, I can decide to mould a lump of wax into a certain shape and leave it to solidify, but the act of leaving it to solidify prevents me from further modifying its form. Additionally, other conscious beings also influence the world through their freely willed actions and the ways in which they influence it also set constraints on my further activity.

And so, my influence on the world is restricted, but this is not because of a deficiency in the power of my free will. Rather, it is because constraints are set on my actions, not only through the way I experience and shape the world, but also through the freely willed actions of others. As conscious subjects, we act on the objective world through our freely willed actions, construct our realities from it, and influence what we and others are capable of doing within it.

X
On Immortality

I have proposed, in this book, that consciousness, or the first-person subjective existence that equates to one's self, is a fundamental entity that is ontologically separate from the objective world. I would now, in this final chapter, like to speculate on whether the philosophical thesis proposed in this book can shed any light on a certain topic of existential significance, namely the topic of immortality. This is a controversial topic, partly because of the strength of peoples' spiritual beliefs concerning it and also because of the influence of the contemporary scientific worldview, which tends to dismiss such beliefs as superstitions. What I present here is not intended to be taken seriously as science. Rather, it is merely some speculation based on my philosophical analysis of the nature of my first-person subjectivity. Such speculation is highly tentative, but it might offer some helpful insight that could be taken forward.

The concept of immortality to which I am referring pertains to the eternity of the self. Often, immortality is confusingly described as "everlasting life" or "life after death". This suggests an erroneous conflation between the presence of life and the presence of consciousness. For example, when an organism is described as being alive, we often assume that the organism has some kind of subjective existence, as well as being alive in a biological sense. Further to the organism's biological properties, we attribute experience to the organism. Indeed, there is often a correlation between being alive and being conscious, such that it is reasonable to assume that living organisms in this world tend to be associated with consciousnesses. However, this correlation between life and consciousness is contingent. While they are correlated in this world, being conscious and being alive are separate features that are metaphysically independent of each other and can come apart. Consciousness refers to first-person subjectivity, whereas life refers to a biological process. Hence, it is a mistake to conflate life with consciousness.

A factor that could have contributed to the conflation of life with consciousness is the vagueness of life as a concept. Indeed, the notion of life is not precise and there is much contention about how to define it adequately. Nonetheless, despite this vagueness, we can accept that that life is a structural and dynamical property of biological organisms, and so it can be physically explained.

How might life be defined? An answer is sketched by Erwin Schrödinger (1944), who notes that a living organism has the ability to "keep going" for a much longer time than an inanimate object. To explain this, Schrödinger appeals to the second law of thermodynamics, which suggests that the entropy of the universe tends to increase with time. For example, imagine that a hot object is placed in a cold room. As time passes, heat conduction causes the object to cool down and the room to warm up, which eventually results in the temperature of the whole system becoming uniform. Thermodynamic equilibrium is reached.

Although the entropy of the universe as a whole tends to increase, what allows a living organism to "keep going", according to Schrödinger, is its ability to resist local increases in entropy. A living organism can, to some extent, evade the trend towards thermodynamic equilibrium. This is made possible by the organism's ability to metabolise energy from nutrients. With this energy, the organism can, among other things, maintain chemical and electrical gradients across membranes, produce movement, synthesise new parts, and remain structured. Therefore, the capacity to metabolise energy in a manner that resists local increases in entropy is a key feature of things that are alive. This is, of course, not a sufficient condition, for several things are able to do this that we would not consider as being alive, such as crystals and refrigerators. Nevertheless, it does provide an explanation of why even a simple living organism, such as *Amoeba proteus*, can "keep going" for a much longer time than, for example, a rolling ball on a level surface.

This ability to resist local increases in entropy, however, is not infallible. There comes a time when an organism's metabolism can no longer cope with the forces of the environment and can no longer resist local increases in entropy. This is the process we call death. The organism can no longer "keep going", but it falls into the drift toward thermodynamic equilibrium with the rest of the universe.

In a more complex organism, such as a human being, death is harder to define, because a person's body is composed of multiple interacting systems. In some contexts, death might be defined by irreversible cardiopulmonary arrest. After all, the heart and lungs are responsible for supplying the body with the oxygen that is required for respiration. However, in other contexts, death might be defined by the irreversible loss of activity of a particular structure in the body, namely the brainstem. This is because the brainstem produces respiratory activity, modulates cardiovascular activity, and induces wakefulness through the action of its reticular activating system.

The above comprises a somewhat simplified account of life and death. Life involves the capacity of an organism to metabolise energy in such a way that it resists local increases in entropy, while death marks the termination of life and the loss of this capacity to resist local increases in entropy. This suggests that life and death are structural and dynamical properties of biological organisms, and so are explainable in physical terms. Accordingly, the prospect of everlasting life is unlikely in the present, when we lack the capability to resist increases in entropy indefinitely. However, the question in which I am interested here is the following. What happens to the self after death? The trouble with various approaches to this question is that they are often confounded by the aforementioned mistaken conflation of the presence of life with the presence of consciousness.

As noted in chapter two, we acknowledge one another as subjective selves, but our embodied interactions with one another involve the features of us that are objectively accessible, such as our bodies and our behaviours. Hence, while we know that our consciousnesses are the essences of our personal identities, we become inclined to characterise one another in terms of our bodily and behavioural features. We also notice that these bodily and behavioural features are properties of living creatures, as the process of life is crucial for maintaining them. Hence, life is often considered to be to one's personhood, while death is often considered to mark the departure of this aspect of personhood from the body.

Throughout history and across cultures, it has widely been believed that this aspect of personhood is attributable to an immaterial soul, which was thought to inhabit the body during life and to leave the body at death. This soul is often characterised as being composed of one's consciousness in conjunction with one's personality and memories. Again, this reflects a common yet misleading conflation of the phenomenal and the psychological. Many spiritual beliefs about life after death, ancient and modern, are based on the supposed everlasting survival of the soul. After death, the soul is believed to leave the body for another destination. The details vary across cultures and religions. In ancient Egypt, it was believed that the *kꜣ*, or vital force, required sustenance after death, while the *bꜣ*, or soul, left the body and survived forever after death. Indeed, in *The Instruction* (c. 2350 BCE), Ptahhotep, who is perhaps the earliest philosopher historically known to us, hints at this dualism between the *kꜣ* and the *bꜣ* through the different ways these terms are used in the text. In some extant religions, the soul is thought to eventually reside in paradise. In other religions, the soul is thought to

become associated with a new body through reincarnation. Some also believe that incorporeal souls, or ghosts, continue to inhabit our world. Despite their different claims about the precise destination of the soul after its departure from the body, these beliefs share the idea that the basis of immortality is the soul's eternal survival after death.

This view was challenged by the scientific claim that the processes that occasion one's personality occur in the brain. If the brain occasions one's thoughts and actions, then it might seem like there is no longer room for a soul to do this work. Thus, skeptics deny the continuity of personality after death. If one's personality is occasioned by the brain and if brain activity ceases upon death, then it seems that one's death marks the cessation of one's personality.

Nonetheless, speculations about the soul and life after death have remained popular, as reflected by the reports of paranormal occurrences in popular media. These include cases suggestive of reincarnation, near-death experiences, mediums, hauntings, poltergeists, and possessions. Of course, skeptics are correct to claim that many of these involve hoaxes or illusions. However, there is also some parapsychological research that is considered respectable. For example, Ian Stevenson (1966) is known for his rigorous research on what many consider to be legitimate cases of reincarnation.

I am not, in this chapter, going to attempt to explain such occurrences. At most, legitimate cases might provide empirical evidence for the survival of one's psychological properties, such as one's memory and personality, after death. However, they do not tell us about the immortality of the self, which is one's first-person subjective existence, or one's consciousness. And so, to understand immortality, one must reflect on the nature of consciousness.

How can reflection on consciousness reveal anything about immortality? I suggest that it can reveal what is necessarily true about first-person existence. For example, it is true that existence necessarily exists, because existence is what exists, which exists by definition. Likewise, it is true that nothingness necessarily does not exist, because nothingness amounts to nonexistence, which does not exist by definition. Given that my consciousness is my first-person existence, it is true that my consciousness necessarily exists to me.

The necessity of existence was acknowledged at least as far back as Parmenides (c. 500 BCE), who suggested that the notion of nonexistence is conceptually incoherent. Of course, I argue that it is true that some things do not exist, as demonstrated by logical impossibilities. I also argue that the monist component of his poem is false because it fails to account for how experiences are individuated

to different experiencers. Nonetheless, I contend that his observation that existence itself is necessary is true. It is true in virtue of its meaning that existence exists, because existence is *what is*, which exists by definition. Moreover, existence is *what is*, regardless of the specific content of *what is*. Even if *what is* amounts to emptiness, there would still exist the existence wherein that emptiness manifests. Correspondingly, it is true in virtue of its meaning that nothingness does not exist, because nothingness is *what is not*, which does not exist by definition. Hence, the suggestion "existence does not exist" is false, because its subject is existence, which exists by definition, and so secures the truth that existence exists. Also, the suggestion "nothingness exists" is false, because its predicate indicates that something exists, which entails that it is not nothingness, and so secures the truth that nothingness does not exist. Therefore, it is necessarily true that existence exists. This is an analytic truth that obtains even under the metaphysical picture suggested by Alexius Meinong (1904), whereby there are some things do not exist. It is necessarily true that existence itself exists, because existence is what exists, which exists by definition.

As well as being a conceptual truth, the necessity of existence is an ontological truth, insofar as the discernment of what exists and what does not is only done within existence. That is to say, the existence is a necessary condition for the very discernment of what is existent and what is nonexistent. Nothingness would preclude such discernment, and so negates the very possibility of nothingness. In some respect, this recalls Jean-Paul Sartre's (1943) suggestion that "nothingness" is a notion that is only realised within existence, and so it is not nothingness in the sense of nonexistence. Indeed, even emptiness is not nothingness, because it presupposes an existence wherein such emptiness is discerned. Therefore, ontological nihilism is necessarily false. There is something rather than nothing, because something is necessary to discern what there is and what there is not.

The above indicates that ontological eternalism is necessarily true with respect to consciousness. My consciousness is my first-person subjective existence. It is what I *am*. Likewise, each of the infinite plurality of consciousnesses is also a separate subjective existence. Given that consciousness is what it is to exist, it is true that consciousness exists necessarily. Indeed, the nonexistence of consciousness is impossible, because the existence of consciousness is necessary for the very discernment of what exists and what does not, insofar as this discernment is only done through consciousness. Thus, the claim that consciousness could not exist is necessarily

false, because its nonexistence would preclude such discernment and negate the very possibility of its nonexistence. It is impossible for me not to exist, because my discernment of what exists and what does not presupposes my existence as a necessary condition. Even when I conceive of emptiness, this necessitates a first-person existence wherein such emptiness is conceived. That is to say, the realisation of emptiness is done through consciousness. Emptiness is usually only apperceived through experiential qualities, such as darkness and silence, which depend on consciousness. Nonetheless, consciousness would exist even without these qualities, as it would obtain as a pure individuated first-person existence wherein qualities could potentially manifest and which is necessary for the discernment of the presence or absence of such qualities.

Given that consciousness is necessary, it is true that the existence of consciousness is eternal. Each of the infinite plurality of consciousnesses exists eternally, because each consciousness is its own necessary first-person existence. Therefore, the necessary fact that I *am* proves to me the truth of the immortality of the self.

Such immortality is also confirmed by dualism. Because consciousness is ontologically separate from the physical world, it is unconditioned by the formal features of the physical world, such as space and time. Given the complete spatial and temporal facts about the world, the existence of consciousness remains a further fact beyond these spatial and temporal facts. Therefore, it is necessarily true that consciousness is immortal, because it is unconditioned by space and time. Rather, consciousness exists beyond space and time.

Indeed, for the experience of time to be possible, it is necessary that there is a constant subjective existence beyond time wherein the experience of time can present. It is only because consciousness exists outside time that one can reflect on relations across time and conceive of scenarios where time is transcended. These include the notion of skipping between moments in time or between timelines, the notion of multiple universes with different spatiotemporal dimensions, and J. M. E. McTaggart's (1908) notion of the unreality of time. Thus, the claim that consciousness occurs within space and time is false, because it fails to account for one's conceptual access to these transtemporal and atemporal features which involve viewing space and time from the outside. Consciousness is more fundamental than space and time, for space and time are only contingently realised through the necessary existence of consciousness.

The above reveals how extraneous the notions of life, death, and change are to the immortality of the self. A process theory that

emphasises continual change, such as that of Alfred North Whitehead (1933), may apply to the physical world, but such a process theory is false with respect to consciousness, because consciousness is nonphysical and unconditioned by time. The suggestion that consciousness could change is false, because consciousness is timeless. Hence, a form of substance theory is more true with respect to consciousness, albeit a form that acknowledges that consciousness exists as a timeless entity that is separate from the objective world. It is true that consciousness has no start and, likewise, it is true that consciousness has no end, because consciousness exists eternally beyond time. Generation, change, and annihilation do not pertain to consciousness, as these notions depend on time. Also, life and death do not pertain to consciousness, as these are processes in the physical world that consciousness transcends.

Accordingly, it is false to think of a comatose or anaesthetised period as marking a "cessation of consciousness". It is more accurate to say that consciousness continues to exist outside time and experiences the comatose or anaesthetised period as a discontinuity in time. Meanwhile, consciousness retains its identity across and beyond this discontinuity. Indeed, as noted in chapter five, the notions of before and after are memories and expectations that are experienced in the timeless existence of consciousness.

Nonetheless, the notions of life and death are not entirely out of place in the present discussion about the self. After all, I have the clear impression that my experience is correlated with the activity of a living body. As I noted in chapter seven, my body provides a psychophysical interface between my consciousness and the physical world. We can also suppose that other bodies act as interfaces between consciousnesses and physical events in the world. However, since one's body is only active when one is alive, it is reasonable to suppose that one's body ceases to act as such an interface when one dies. Life, therefore, can be considered to enable a psychophysical interface between consciousness and the physical world, while death can be considered to involve the removal of this interface.

Given its timelessness, it is true that consciousness is not generated. This seems to indicate that classical theism is false. We can accept that the historical figures associated with religions were real conscious individuals. We could even accept that there are possible worlds wherein the mythological characters associated with religions are real conscious individuals. However, a singular creator god does not exist, because consciousness exists eternally and is not created. The notion of creation does not pertain to consciousness,

because creation is a process in time, whereas consciousness exists beyond time. Thus, nontheism is true, because it is necessary to account for the fundamentality and timelessness of consciousness. We can acknowledge the cultural value and moral significance of religious practices, but we can rebut some of the associated ontology.

Can the timelessness of consciousness shed light on the findings of parapsychological research? It can neither confirm nor disconfirm them, because the ontology of consciousness is independent of the empirical facts about the world. The nature of the afterlife is underdetermined by the eternity of consciousness. However, this also means that the eternity of consciousness can accommodate various beliefs about the nature of the afterlife. It is true that the existence of consciousness is necessary. When that is acknowledged, there is room for further philosophically informed speculation.

As noted earlier, my body provides a psychophysical interface between my consciousness and the physical world. Not only are the qualities I experience correlated with the events within this interface, but the physical structure of the interface defines the limits of my experience. Upon death, the activity of my body ceases and, as I noted earlier, this amounts to the removal of the interface between my consciousness and the physical world. There are countless possibilities for what happens to my experience once this interface is removed. Perhaps a new interface is provided by another biological system in this world. This would be a literal interpretation of reincarnation. Perhaps an interface is provided by a different kind of system, in which case the ways that space, time, and logic manifest may be very different. Perhaps no new physical interface is provided and I exist as a pure consciousness, which is reminiscent of *nibbāna* in Buddhism or *mokṣa* in Sāṃkhya philosophy and Jaina philosophy.

As well as acknowledging the eternity of consciousness, any speculation about what happens to experience after death must also acknowledge the first-person individuation of consciousness. For example, the suggestion that we are all incarnations of the same soul would be false, because it fails to account for the fact that subjects are ontologically separate from one another in virtue of their being experientially individuated from one another. You and I are fundamentally different subjects, because my subjective experience has a first-person individuation unique to me and your subjective experience has a first-person individuation unique to you. Likewise, as noted in Sāṃkhya philosophy and Jaina philosophy, the suggestion that I could exist in a state of *nibbāna* or *mokṣa* as a pure consciousness after death must acknowledge that the I would still be

ontologically separate from the plurality of other subjects in virtue of my first-person individuation. And so, speculation about immortality must also acknowledge the infinite plurality of ontologically distinct consciousnesses with discretely unique ipseities.

In virtue of its timelessness, it is true that consciousness cannot be annihilated. What this indicates is that the existence of consciousness would be unaffected by a much larger dissolution of the psychophysical interface, such as the cessation of the physical universe. Consciousness exists beyond the spatiotemporal subsistence of the physical universe, and so consciousness would still exist if the physical universe ceased. As noted above, a subject might exist as a pure consciousness without a physical interface. We could also reasonably speculate that new interfaces for consciousnesses might be provided by structures in other universes of the infinite multiverse or in subsequent cycles of an infinitely cyclical universe. This is not necessarily a consolation, because we cannot tell what these other universes might be like. Nonetheless, the claim that the existence of consciousness would be affected by the cessation of the physical universe is false.

This raises the question of whether other features, such as personality and memory, are also conserved after death. Do our memories continue to be associated with our respective consciousnesses? Again, the eternity of consciousness can neither affirm nor preclude this. The issue remains open to philosophically informed speculation. It may be reasonable to suppose that personality and memory are conserved with the self. For example, in the sort of case of reincarnation studied by Ian Stevenson (1966), it may be true that the same consciousness is associated with the earlier incarnation and the later incarnation, such that the earlier incarnation and the later incarnation are incarnations of the same subjective self. However, even if we assume a case where psychological properties cease with the brain, it is impossible to negate the eternity of consciousness. Under the philosophical framework I have presented, it is necessarily true that consciousness is indestructible, because it transcends the structure and dynamics of the physical world.

Therefore, I have presented my admittedly speculative, yet philosophically informed, dualist account of the immortality of the self. My consciousness is my first-person subjective existence. This first-person subjective existence is a separate entity from the physical world, and so does not subsist within space and time. Rather, is the existence wherein space and time are realised. Accordingly, it is fundamental and eternal. Consciousness exists necessarily.

Bibliography

Abbott, E. A.	1844	*Flatland: A Romance of Many Dimensions*. London: Seeley and Company.
Anscombe, G. E. M.	1975	"The First Person". In S. D. Guttenplan (ed.), *Mind and Language*. Oxford: Clarendon Press.
Armstrong, D. M.	1968	*A Materialist Theory of the Mind*. London: Routledge.
Ayer, A. J.	1936	*Language, Truth, and Logic*. London: Victor Gollancz Limited.
Baars, B. J.	1988	*A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
Bayes, T.	1763	"An Essay Towards Solving a Problem in the Doctrine of Chances". *Philosophical Transactions of the Royal Society of London*, 53: 370–418.
Berkeley, G.	1710	*A Treatise Concerning the Principles of Human Knowledge*. New York: Oxford University Press, 1998.
Bhikkhu, Ṭ.	1993	*The Mind Like Fire Unbound*. Barre: Dhamma Dana Publications.
Black, M.	1952	"The Identity of Indiscernibles". *Mind*, 61: 153–164.
Blackmore, S. J.	1982	*Beyond the Body: An Investigation into Out-of-Body Experiences*. London: Heinemann.
Block, N.	1978	"Troubles with Functionalism". In C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundation of Psychology*. Minneapolis: University of Minnesota Press.
Bohr, N.	1922	*The Theory of Spectra and Atomic Constitution: Three Essays*. Cambridge: Cambridge University Press.
Brentano, F.	1874	*Psychology from an Empirical Standpoint*. London: Routledge and Kegan Paul, 1973.
Carpenter, R. H. S.	1997	"Anyone for Free Will?" *New Scientist*, 26th April 1997.
Carpenter, W. B.	1874	*Principles of Mental Physiology*. London: H. S. King and Company.
Cartwright, N.	1983	*How the Laws of Physics Lie*. Oxford: Clarendon Press.

Chalmers, D. J.	1996	*The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
Chisholm, R. M.	1957	*Perceiving*. Ithaca: Cornell University Press.
Churchland, P. M.	1985	"Reduction, Qualia, and the Direct Introspection of Brain States". *Journal of Philosophy*, 82: 8–28.
	1995	*The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge, MA: MIT Press.
Churchland, P. S.	1988	"The Significance of Neuroscience for Philosophy". *Trends in the Neurosciences*, 11: 304–307.
Crick, F. H. C. and Koch, C.	1990	"Towards a Neurobiological Theory of Consciousness". *Seminars in the Neurosciences*, 2: 263–275.
Damasio, A.	1999	*The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt, Brace, and Company.
Darwin, C.	1859	*On the Origin of Species*. London: John Murray.
Davidson, D.	1970	"Mental Events". In L. Foster and J. W. Swanson (eds.), *Experience and Theory*. Amherst: University of Massachusetts Press.
Dawkins, R.	1976	*The Selfish Gene*. Oxford: Oxford University Press.
Dein, S.	2004	"Working with Patients with Religious Beliefs". *Advances in Psychiatric Treatment*, 10: 287–294.
Dennett, D. C.	1991	*Consciousness Explained*. Boston: Little and Brown.
Descartes, R.	1641	*Meditations on First Philosophy*. Indianapolis: Hackett, 1993.
Edelman, G.	1989	*The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.
Einstein, A.	1916	*Relativity: The Special and General Theory*. New York: Holt and Company.
Elitzur, A.	1989	"Consciousness and the Incompleteness of the Physical Explanation of Behavior". *Journal of Mind and Behavior*, 10: 1–20.
Engels, F.	1893	"Letter to F. Mehring". In K. Marx and F. Engels, *Selected Works in Two Volumes*, volume II. Moscow: Foreign Languages Publishing House, 1949.

Everett, H.	1957	"Relative State Formulation of Quantum Mechanics". *Review of Modern Physics*, 29: 454–462.
Fodor, J.	1968	*Psychological Explanation*. New York: Random House.
Frege, F. L. G.	1892	"Über Sinn und Bedeutung". *Zeitschrift für Philosophie und Philosophische Kritik*, 100: 25–50.
Freud, S.	1900	*The Interpretation of Dreams*. London: Hogarth Press, 1953.
Hameroff, S. R.	1994	"Quantum Coherence in Microtubules: A Neural Basis for an Emergent Consciousness?" *Journal of Consciousness Studies*, 1: 91–118.
Hamilton, W.	1865	*Lectures on Metaphysics*, volume I. Boston: Gould and Lincoln.
Hardin, C. L.	1988	*Color for Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett.
Hegel, G. W. F.	1816	*The Science of Logic*. London: George Allen and Unwin, 1969.
Heisenberg, W.	1930	*The Physical Principles of the Quantum Theory*. New York: Dover.
Helmholtz, H. v.	1867	*Handbuch der Physiologischen Optik.* Leipzig: Voss.
Hobbes, T.	1655	*De Corpore*, part I. New York: Abaris Books, 1981.
Hodgson, D.	1991	*The Mind Matters*. Oxford: Oxford University Press.
Hofstadter, D. R.	1979	*Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.
Hume, D.	1740	*A Treatise of Human Nature*. Oxford: Clarendon Press, 1975.
	1748	*An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press, 1999.
Husserl E.	1921–1928	*Zur Phänomenologie der Intersubjektivität II.* The Hague: Martinus Nijhoff, 1973.
	1931	*Cartesian Meditations: An Introduction to Phenomenology*. The Hague: Martinus Nijhoff, 1960.
Ibn Sīnā	1027	*Avicenna's De Anima: Being the Psychological Part of Kitāb al-Shifā.* F. Rahman (ed.). London: Oxford University Press, 1959.

Iśvarakṛṣṇa	c. 350	*Sāṅkhya Kārikā*. H. T. Colebrooke (trans.). Oxford: Oxford University Press, 1837.
Jackson, F.	1982	"Epiphenomenal Qualia". *Philosophical Quarterly*, 32: 127–136.
	1986	"What Mary Didn't Know". *Journal of Philosophy*, 83: 291–295.
James, W.	1890	*The Principles of Psychology*. Cambridge, MA: Harvard University Press.
Kant, I.	1781	*A Critique of Pure Reason*. Cambridge: Cambridge University Press, 1998.
Kripke, S. A.	1980	*Naming and Necessity*. Cambridge, MA: Harvard University Press.
Laing, R. D.	1967	*The Politics of Experience* and *The Bird of Paradise*. Harmondsworth: Penguin.
Laplace, P. S.	1814	*A Philosophical Essay on Probabilities*. New York: Dover, 1951.
Leibniz, G. W. v.	1686	*Discourse on Metaphysics*. Manchester: Manchester University Press, 1953.
	1714	*Monadology*. Pittsburgh: University of Pittsburgh Press, 1991.
Levine, J.	1993	"On Leaving out what it's Like". In M. Davies and G. Humphreys (eds.), *Consciousness: A Mind and Language Reader*. Oxford: Blackwell.
Lewis, D. K.	1986	*On the Plurality of Worlds*. Oxford: Blackwell.
	1990	"What Experience Teaches". In W. Lycan (ed.), *Mind and Cognition: A Reader*. Oxford: Blackwell.
Lipton, P.	1991	*Inference to the Best Explanation*. London: Routledge, 2nd edition 2004.
Locke, J.	1689	*An Essay Concerning Human Understanding*. Oxford: Clarendon Press, 1975.
Lycan, W. G.	1998	"Theoretical (Epistemic) Virtues". In E. Craig (ed.), *Routledge Encyclopedia of Philosophy* [Online]. London: Routledge.
Mach, E.	1886	*The Analysis of Sensations and the Relation of Physical to the Psychical*. New York: Dover, 1959.

Malcolm, N.	1958	"Knowledge of Other Minds". In D. M. Rosenthal (ed.), *The Nature of Mind*. Oxford: Oxford University Press.
Marcel, A. J.	1988	"Phenomenal Experience and Functionalism". In A. J. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science*. Oxford: Clarendon Press.
McGinn, C.	1989	"Can We Solve the Mind-Body Problem?" *Mind*, 98: 349–366.
McTaggart, J. M. E.	1908	"The Unreality of Time". *Mind*, 17: 457–474.
Meinong, A.	1904	"Über Gegenstandstheorie". In *Untersuchungen zur Gegenstandstheorie und Psychologie*. Leipzing: J. A. Barth.
Mill, J. S.	1889	*An Examination of Sir William Hamilton's Philosophy*. London: Longman, Green, Longman, Roberts, and Green.
Nagel, T.	1974	"What Is It Like to Be a Bat?" *Philosophical Review*, 78: 79–89.
	1986	*The View from Nowhere*. New York: Oxford University Press.
Nelkin, N.	1993	"The Connection between Intentionality and Consciousness". In M. Davies and G. Humphreys (eds.), *Consciousness: A Mind and Language Reader*. Oxford: Blackwell.
Nemirow, L.	1990	"Physicalism and the Cognitive Role of Acquaintance". In W. G. Lycan (ed.), *Mind and Cognition: A Reader*. Oxford: Blackwell.
Newton, I.	1687	*Philosophiae Naturalis Principia Mathematica*. Cambridge: Cambridge University Press, 1972.
Papineau, D.	2002	*Thinking About Consciousness*. Oxford: Oxford University Press.
Parfit, D.	1971	"Personal Identity". *Philosophical Review*, 80: 3–27.
Parmenides	c. 500 BCE	*Parmenides of Elea: Fragments*. D. Gallop (trans.). Toronto: University of Toronto Press, 1984.
Penrose, R.	1989	*The Emperor's New Mind*. Oxford: Oxford University Press.
Perry, J.	1979	"The Problem of the Essential Indexical". *Noûs*, 13: 3–21.

Place, U. T.	1956	"Is Consciousness a Brain Process?" *British Journal of Psychology*, 47: 44-51.
Plato	c. 360 BCE	*Phaedo*. G. M. A. Grube (trans.). Indianapolis: Hackett, 1977.
Popper, K. R. and Eccles, J. C.	1977	*The Self and Its Brain: An Argument for Interactionism.* New York: Springer International.
Priest, G.	1987	*In Contradiction: A Study of the Transconsistent.* Dordrecht: Martinus Nijhoff.
Ptahhotep	c. 2350 BCE	*The Instruction of Ptahhotep*. B. G. Gunn (trans.). London: John Murray, 1909.
Putnam, H.	1975	"Other Minds". In *Mind, Language, and Reality: Philosophical Papers*, volume II. Cambridge: Cambridge University Press.
Quine, W. v. O.	1960	*Word and Object*. Cambridge, MA: MIT Press.
Russell, B.	1912	*The Problems of Philosophy*. London: Williams and Norgate.
	1927	*The Analysis of Matter*. London: Kegan Paul.
Ryle, G.	1949	*The Concept of Mind*. London: Hutchinson.
Sartre, J. P.	1943	*Being and Nothingness: An Essay on Ontological Phenomenology*. New York: Philosophical Library, 1956.
Schrödinger, E.	1944	*What is Life? The Physical Aspect of the Living Cell*. Cambridge: Cambridge University Press.
	1958	*Mind and Matter*. Cambridge: Cambridge University Press.
Searle, J. R.	1992	*The Rediscovery of the Mind*. Cambridge, MA: Harvard University Press.
Shoemaker, S.	1981	"The Inverted Spectrum". *Journal of Philosophy*, 74: 357–381.
Smart, J. J. C.	1959	"Sensations and Brain Processes". *Philosophical Review*, 68: 141–156.
Sober, E.	1994	*From a Biological Point of View: Essays in Evolutionary Philosophy*. Cambridge: Cambridge University Press.

Sperry, R. W.	1969	"A Modified Concept of Consciousness". *Psychological Review*, 76: 532–536.
Spinoza, B.	1677	*Ethics*. Indianapolis: Hackett, 1992.
Stevenson, I.	1966	*Twenty Cases Suggestive of Reincarnation*. Charlottesville: University Press of Virginia.
Strawson, P. F.	1959	*Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.
Sutherland, N. S.	1989	*The International Dictionary of Psychology*. New York: Continuum.
Swartz, N.	1985	*The Concept of a Physical Law*. New York: Cambridge University Press.
Swinburne, R.	1984	"Personal Identity: The Dualist Theory". In S. Shoemaker and R. Swinburne, *Personal Identity*. Oxford: Blackwell.
Turing, A.	1950	"Computing Machinery and Intelligence". *Mind*, 59: 433–460.
Umāsvāti	c. 100–400	*Tattvārtha Sūtra*. N. Tatia (trans.). San Francisco: Harper Collins, 1994.
van Fraassen, B. C.	1980	*The Scientific Image*. Oxford: Oxford University Press.
van Gulick, R.	1993	"Understanding the Phenomenal Mind: Are We All Just Armadillos?" In M. Davies and G. Humphreys (eds.), *Consciousness: A Mind and Language Reader*. Oxford: Blackwell.
Varela, F. J., Thompson, E., and Rosch, E.	1992	*The Embodied Mind: Cognitive Science and Human Experience.* Cambridge, MA: MIT Press.
Whitehead, A. N.	1933	*Adventures of Ideas.* New York: MacMillan.
Wittgenstein, L.	1953	*Philosophical Investigations*. Oxford: Blackwell.
Zahavi, D.	1999	*Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston: Northwest University Press.
Ziff, P.	1965	"The Simplicity of Other Minds". *Journal of Philosophy*, 62: 575–584.
Zohar, D.	1990	*The Quantum Self*. New York: Quill.


9 781446 680339