1 Introduction

1.1 The Problem

Consciousness is arguably the most important area within contemporary philosophy of mind, with an explosion of research over the past thirty years from philosophers, psychologists, and scientists.\(^1\) Consciousness is also perhaps the most puzzling aspect of the world, and yet it is so very familiar to each of us. Attempts to explain it in neurophysiological or even cognitive terms are still met with great resistance. It seems to many that conscious mental states simply cannot be reduced to, or explained in terms of, something less problematic. In this book, I defend and further develop a metapsychological reductive representational theory of consciousness and then apply it to several importantly related problems, including concept acquisition and animal consciousness.

Going back to my book *Consciousness and Self-Consciousness* (1996), I have defended a version of the higher-order thought (HOT) theory of consciousness, which says that what makes a mental state conscious is that there is a suitable higher-order thought directed at the mental state. Higher-order thoughts (HOTs) are metapsychological or metacognitive states, that is, mental states directed at other mental states. HOT theory is primarily concerned with explaining how conscious mental states differ from unconscious mental states. It seems reasonable to think that conscious mental states are states that we are “aware of” in some sense. Its best-known defender is David Rosenthal.\(^2\)

I called my version of HOT theory the wide intrinsicality view (WIV) for reasons we will see in due course. Moreover, in Gennaro 1996, I was chiefly concerned to argue for the more general Kantian thesis that consciousness entails self-consciousness. Defending HOT theory was therefore mainly a means to that end. Since that book, however, I have further developed my own version of HOT theory (Gennaro 2006a), including attention to issues...
such as animal consciousness (Gennaro 2004a, 2009) and the well-known “hard problem” of consciousness (Gennaro 2005b). In some cases, I have simply defended HOT theory against a specific objection (Gennaro 2003). In addition to further defending HOT theory, however, I am interested in solving what I take to be a larger underlying paradox that I will call the Consciousness Paradox, namely, how it is possible to hold the following set of apparently inconsistent yet independently plausible theses:

1. **The HOT Thesis**: A version of the HOT theory is true (and thus a version of reductive representationalism is true).
2. **The Hard Thesis**: The hard problem of consciousness, that is, the problem of explaining exactly how or why subjective experiences are produced from brain activity (or from any combination of unconscious mental activity), can be solved.
3. **The Conceptualism Thesis**: Conceptualism is true, that is, all conscious experience is structured by concepts possessed by the subject.
4. **The Acquisition Thesis**: The vast majority of concepts are acquired, though there is a core group of innate concepts.
5. **The Infants Thesis**: Infants have conscious mental states.
6. **The Animals Thesis**: Most animals have conscious mental states.
7. **The HOT-Brain Thesis**: There is a plausible account of how HOT theory, and especially the WIV, might be realized in the brain and can lead to an informative neurophysiological research agenda. Alternatively: HOT theory is interestingly related to and consistent with a number of leading empirical theories of consciousness.

Indeed, it is often claimed that HOT theory alone is inconsistent with several of the other theses in the foregoing list. For example, some have argued that the HOT Thesis conflicts with the Animals Thesis because animals cannot have what seem to be fairly sophisticated HOTs of the form “I am in mental state M now.” Much the same has been said about the Infants Thesis. Further, the Conceptualism Thesis has been thought by some to contradict the Acquisition and Animals Theses. Do animals even possess concepts? If all conscious perceptual experience is determined by already possessed concepts, then how could we acquire new concepts? Although I think that each of the theses is independently plausible and defensible, some are more controversial than others, and it is necessary to explain how they can all be mutually consistent. Thus my overall aim is to argue for a philosophical theory of consciousness while applying it to other significant issues such as concept possession and concept acquisition, topics not frequently found in the philosophical literature on consciousness. This book
also addresses interdisciplinary topics such as animal consciousness and how HOT theory might be realized in the brain. Thus I hope that it will be of interest to nonphilosophers as well as philosophers.

Most cognitive science and empirical works on, for example, concepts or animal consciousness do not address central philosophical theories of consciousness. Some are primarily experimental or scientific works by authors not necessarily interested in consciousness research as such. These works mostly focus on the nature of concepts, concept acquisition, and theories of mental state attribution, but without delving much into the philosophical problem of consciousness. On the other hand, many of the more philosophical works do not integrate a specific theory of consciousness with the cognitive science literature on the topics listed earlier.

There are of course numerous important and helpful anthologies in the field. Some are specifically on HOT theory or closely related theories of consciousness (Gennaro 2004b; Carruthers 2005; Rosenthal 2005; and Kriegel and Williford 2006), with the volumes by Rosenthal and Carruthers representing collections of their own writings. Finally, although there are many other excellent anthologies on consciousness, they are obviously not designed to put forth a single unified theory. The present work therefore addresses various problems in novel ways and in relation to HOT theory.

1.2 The Plan

It might be useful to think of this book as comprising two parts, chapters 1 through 5 and chapters 6 through 9. In the present chapter, I lay out the overall problem and make some key distinctions. In chapters 2 through 5, I defend the HOT Thesis. I also argue for the Hard Thesis in chapter 4. In chapters 6 through 9, I defend the remaining theses paying special attention to how each thesis is consistent with the others, including the HOT Thesis.

In chapter 2, I first defend representationalism, which is the thesis that phenomenal properties are identical to certain representational properties. I then argue that reductionist representationalism is most desirable if we are to explain consciousness. I defend the associated view that intentionality is more primitive than consciousness, and offer an account of mental content. Finally, I present an initial defense of HOT theory.

In chapter 3, I argue against several theories of consciousness that are somewhat close relatives to HOT theory. I reject first-order representationalism (FOR) such as the account offered by Michael Tye (1995, 2000). I argue that FOR cannot adequately explain the difference between conscious and
unconscious mental states. I also argue that HOT theory is preferable to the higher-order perception (HOP) alternative proposed by William Lycan (1996) and the dual-content (or dispositional HOT) theory of Peter Carruthers (2000, 2005).

In chapter 4, I first motivate the need for a modified version of Rosenthal’s HOT theory, namely, the wide intrinsicality view (WIV), focusing on two major objections to HOT theory: the so-called problem of the rock and the misrepresentation objection. I then take on the hard problem of consciousness and end with replies to several further objections to the WIV. I argue that HOT theory is immune to Chalmers’s (1995) criticisms of other attempted reductionist accounts of consciousness, but also that a similar version of the problem can be solved.

In chapter 5, I turn my attention to another related theory of consciousness. Although there is something right about the idea that conscious mental states represent themselves, I argue against the so-called self-representational theory of consciousness. Prominent in this chapter is critical discussion of recent work by Uriah Kriegel.²

At this point, I conclude that the HOT and Hard Theses have been tentatively established though much more will follow in later chapters especially with regard to how the HOT Thesis is consistent with the remaining theses.

Chapter 6 is devoted to defending the Conceptualism Thesis against some well-known objections and several prominent critics, such as Sean Kelly and Christopher Peacocke. The central issue is whether or not one can have conscious experience of objects or properties without having the corresponding concepts. In this chapter, I offer a more detailed account of concept possession and also explain crucial connections between HOT theory and conceptualism.

In chapter 7, I first defend the Acquisition Thesis. One objection to both HOT theory and conceptualism is that concept acquisition is impossible or extremely difficult to explain. For example, how can one acquire the concept of a novel (type of) object if having conscious perception of that object already presupposes that the subject possesses the concept? This chapter addresses this problem, which I take to be the “real hard problem” of consciousness. With the aid of a wealth of experimental evidence from the infant and child developmental literature, I argue that concept acquisition is indeed consistent with both conceptualism and HOT theory. I also defend the Infants Thesis and argue that it is consistent with the HOT and Acquisition Theses.

In chapter 8, I defend the Animals Thesis. The nature of concept possession and so-called “I-thoughts” plays an important role in the ever-
increasing animal cognition literature. In previously defending HOT theory, I have responded at length to the allegation that HOTs (along with their constituent concepts) are too sophisticated for many animals to have (Gennaro 1993, 2004a, 2009). I continue this defense in chapter 8 and then extend it to include discussion of conceptualism. It is also necessary to examine experimental results to determine whether or not animals have self-concepts, self-awareness, episodic memory, and an ability to attribute mental states to others. I argue that the Animals Thesis is consistent with both the Conceptualism and HOT Theses.

Chapter 9 contains a defense of the HOT-Brain Thesis. This chapter explores the neurophysiological evidence for HOT theory and, especially, the wide intrinsicality view (WIV). I argue that my theory can shed light on, and is consistent with, several empirical theories of consciousness. I examine the literature on the “neural correlates of consciousness” (NCCs), that is, the ongoing scientific project of determining the precise neural correlates of consciousness. Additional motivation for the HOT-Brain Thesis is, for example, to refute Antti Revonsuo’s recent charge that “these theories [i.e., HOT theories] have not had any major impact on the empirical study of consciousness” (2010, 189). The well-known “binding problem” must also be addressed in this context; namely, how does the varied incoming information to the brain result in a unified and coherent visual experience? The chapter also critically discusses the closely linked problem of the unity of consciousness.

Overall, I claim that the Consciousness Paradox can be solved.

1.3 Some Terminology and Distinctions

The concept of consciousness is notoriously ambiguous. It is important first to make several distinctions and to define key terms. The abstract noun “consciousness” is not often used in the contemporary literature, though it originally derives from the Latin con (with) and scire (to know). One can have knowledge of the external world or one’s own mental states through consciousness. The primary contemporary interest lies more in the use of the expressions “x is conscious” or “x is conscious of y.” Under the former category, perhaps most important is the distinction between state and creature consciousness (Rosenthal 1993a). We sometimes speak of an individual mental state, such as a pain or perception, as being conscious. On the other hand, we also often talk about organisms or creatures as conscious, such as when we say that “human beings are conscious” or “cats are conscious.” Creature consciousness is simply meant to refer to the fact that an organism
is awake, as opposed to sleeping or in a coma. However, some kind of state consciousness is normally implied by creature consciousness; that is, if a creature is conscious, then it must have conscious mental states.

Due to the lack of a direct object in the expression “x is conscious,” this is usually referred to as intransitive consciousness, in contrast to transitive consciousness, where the locution “x is conscious of y” is used (Rosenthal 1993a). Most contemporary theories of consciousness are aimed at explaining state consciousness, that is, what makes a mental state conscious. This is also the case for HOT theory in the sense that intransitive (state) consciousness is explained in terms of transitive consciousness.

It might seem that the term “conscious” is synonymous with, say, “awareness” or “experience” or “attention.” However, it is crucial to recognize that this is not generally accepted today. For example, though perhaps somewhat atypical, one might hold that there are unconscious experiences, depending, of course, on how the term “experience” is defined (Carruthers 2000). More common is the belief that we can be aware of external objects in some unconscious sense, such as during instances of subliminal perception. The expression “conscious awareness” does not therefore seem to be redundant. Finally, it is not clear that consciousness ought to be restricted to attention. It seems plausible to suppose, for example, that one is conscious of objects to some extent in one’s peripheral visual field even though one is attending to a more narrow (or focal) set of objects within that visual field. Needless to say, contemporary philosophers and psychologists are nearly unanimous in allowing for unconscious mental states or representations, though they differ as to whether this applies to all kinds of mental states including, say, pains and feelings.

Perhaps the most fundamental and commonly used notion of “conscious” is captured by Thomas Nagel’s famous “what it is like” sense (Nagel 1974). When I am in a conscious mental state, there is “something it is like” for me to be in that state from the subjective or first-person point of view. When I smell a rose or have a conscious visual experience, there is something it “seems” or “feels like” from my perspective. An organism such as a bat is conscious if it is able to experience the world through its echolocation senses. There is also something it is like to be a conscious creature, whereas there is nothing it is like to be a table or tree. This is primarily the sense of “conscious state” that I use throughout the book. “What it’s like” basically means “how a conscious state is for the subject.” When it comes to capturing the main phenomenon to be explained, it seems to me that we most often have Nagel’s “something it is like” sense in mind.
There are still, though, a cluster of expressions and technical terms associated with Nagel’s sense, and some authors simply stipulate the way that they use them. For example, philosophers often refer to conscious states as *phenomenal* or *qualitative* states. More technically, philosophers frequently describe such states as having qualitative properties called “qualia” (singular, * quale*). Chalmers explains that a “mental state is conscious if there is something it is like to be in that mental state. . . . We can say that a mental state is conscious if it has a qualitative feel. . . . These qualitative feels are also known as phenomenal qualities, or qualia for short” (1996, 4). There is significant disagreement over the nature, and even the existence, of qualia, but they are most often understood as the felt properties or qualities of conscious states. There is something it is like to have qualia or to be in a qualitative state. Most generally, perhaps, qualia are “introspectively accessible, phenomenal aspects of our mental lives” (Tye 2009a). But even this can be misleading if it is taken to imply that only introspected, or introspectible, states have qualia. Surely first-order, or world-directed, conscious states also have qualia. Kind (2008) explains that “qualia are subjective or qualitative properties of experiences. What it feels like, experientially, to see a red rose is different from what it feels like to see a yellow rose. Likewise for hearing a musical note played by a piano and hearing the same musical note played by a tuba. The qualia of these experiences are what give each of them its characteristic ‘feel’ and also what distinguish them from one another.” In any case, qualia are most often treated as properties of some mental states, though some use the term “qualia” in the more external sense of “the qualities of what is represented.” I will use it in the former sense.

One also finds closely allied expressions like “phenomenal character” and “subjective character” in the literature. Tye (2009a), for example, tells us that the “phenomenal character of an experience is what it is like subjectively to undergo the experience.” More explicitly, Kriegel (2009a) is at great pains to distinguish what he calls “qualitative character” from “subjective character” under the larger umbrella of “phenomenal character” because they play such a central role in his theory of consciousness. He explains that “a phenomenally conscious state’s qualitative character is what makes it the phenomenally conscious state it is, while its subjective character is what makes it a phenomenally conscious state at all” (Kriegel 2009a, 1). In his view, then, the *phenomenally conscious* experience of the blue sky should be divided into two components: (1) its *qualitative character*, which is the “bluish” component of the experience (or the *what* of the experience), and (2) its *subjective character*, which is what he calls the “for-me” component (or what determines *that* it is conscious). As we will see in chapter 5, I think
that Kriegel is mistaken in thinking that subjective character is itself phenomenally conscious, though I am more sympathetic with his account of qualitative character.

Finally, Ned Block (1995) makes an oft-cited distinction between phenomenal consciousness (or "phenomenality") and access consciousness. Phenomenal consciousness is very much in line with Nagel's notion described earlier. However, Block defines the quite different notion of access consciousness in terms of a mental state's relationship with other mental states, for example, a mental state's "availability for use in reasoning and rationality guiding speech and action" (Block 1995, 227). This view would, for example, count a visual perception as (access) conscious not because it has the "what it's likeness" of phenomenal states but because it carries visual information that is generally available for use by the organism, regardless of whether or not it has any qualitative properties. Access consciousness is therefore a functional notion concerned with what such states do. Although something like this idea is certainly important in cognitive science and philosophy of mind generally, not everyone agrees that access consciousness deserves to be called "consciousness" in any important sense. Block himself argues that neither sense of consciousness implies the other, while others urge that a more intimate connection holds between the two. For example, according to HOT theory, phenomenality would entail (higher-order) access consciousness, but not all access consciousness would have phenomenality.

My sense is that some authors only add to the terminological confusion by introducing new (or not so new) distinctions into the literature instead of clarifying existing meanings of "consciousness" or simply adopting a prior definition over others. Has, for example, Block's distinction between access and phenomenal consciousness really clarified the matter? It is important to resist the constant temptation to introduce our own special terminology, though we are all perhaps guilty to some extent. For example, here is a sample from my own 1996 book: "Qualia are the properties of phenomenal states that determine their qualitative character, i.e. 'what it is like' to have them" (Gennaro 1996, 7). A phenomenal state can occur unconsciously and is "a mental state which . . . typically has qualitative properties" (7), whereas a qualitative state is a "phenomenal state with a qualitative property" (8). Thus, according to my 1996 view, a phenomenal state can be unconscious, but a qualitative state must be conscious.

I make a plea for more uniform usage whenever possible. Unless I specifically indicate otherwise, I will from now on use the terms "phenomenal," "qualitative," and "experience" as conscious in Nagel's sense, but I will
allow for unconscious *awareness* and unconscious *representations* directed at the outer world or one’s own mental states. So, for me, there are no unconscious experiences and, in contrast to other higher-order theorists, no unconscious qualitative states (or unconscious qualia). There is little reason to have an unconscious counterpart for each of the terms I have listed. I will also avoid using as much of the foregoing technical jargon as possible throughout this book where I can do so without sacrificing rigor or accuracy. However, it will sometimes be necessary to get into the terminological weeds, especially when discussing other views. In any case, I now turn to a defense of the HOT thesis.