Higher-Order Theories of Consciousness: An Overview

Rocco J. Gennaro

1. General Introduction and Terminology

Explaining the nature of consciousness is one of the more important and perplexing areas in philosophy. What is consciousness? How is the conscious mind related to the body? Can consciousness be explained in terms of brain activity? How can we explain the sensation of the smelling of a rose or a conscious visual experience? Given that philosophy is interdisciplinary by its very nature, the problem of consciousness is also explored in such related fields as psychology and neuroscience. One question that should be answered by any viable theory of consciousness is: What makes a mental state a conscious mental state? That is, what transforms a nonconscious mental state into a conscious one? There is a long tradition that has attempted to understand consciousness in terms of some kind of higher-order awareness. For example, John Locke (1689/1975) once said that “consciousness is the perception of what passes in a man’s own mind.” This intuition and attempt to explain consciousness has been recently revived by a number of philosophers (Rosenthal, 1986, 1990, 1993a, 2000, 2004; Armstrong, 1968, 1981; Lycan, 1996, 2001). In general, the idea is that what makes a mental state conscious is that it is the object of some kind of higher-order representation (HOR). A mental state M becomes conscious when there is a HOR of M. A HOR is a “meta-psychological” state, i.e. a mental state directed at another mental state. So, for example, my desire to write a good introduction becomes conscious when I become “aware” of the desire. Intuitively, it seems that conscious states, as opposed to nonconscious ones, are the mental states that I am “aware of” in some sense. Any theory which attempts to explain consciousness in terms of higher-order states is known as a higher-order (HO) theory of consciousness. It is best initially to use the more neutral term ‘representation’ because there are a number of different kinds of higher-order theory, depending upon how one characterizes the HOR-state in question. Moreover, to be clear, the sense of ‘conscious state’ that I have in mind is the same as Nagel's (1974) sense, i.e. there is
“something it is like to be in that state” from the subjective or first-person point of view. When I am, for example, having a conscious visual experience, there is something it “seems” or “feels” like from my first-person subjective perspective.

The key to HO theories is their hierarchical or iterative structure, and so they have been called “double-tiered” theories (Güzeldere, 1995). HO theories are also attractive to some philosophically inclined psychologists and neuroscientists partly because they suggest a very natural realization in the brain structure of humans and other animals (see e.g. Rolls, 1999; Weiskrantz, 2000). At the risk of oversimplification, if we think of the brain as developing layers upon layers corresponding to increasing sophistication in mental ability, then the idea is that mental states corresponding to various “higher” areas of the brain (e.g. cortex) are directed at various “lower” states rendering them conscious. It is important to note, however, that HO theories do not attempt to reduce consciousness directly to neurophysiological states. Unlike some other theories of consciousness (Crick and Koch, 1990; Crick, 1994), they are not reductionist in the sense that they attempt to explain consciousness directly in physicalistic (e.g. neurophysiological) terms. Instead, HO theories attempt to explain consciousness in mentalistic terms, that is, by reference to such notions as ‘thoughts’ and ‘awareness.’ Thus, conscious mental states arise when two nonconscious mental states are related in a certain specific way; namely, that one of them (the HOR) is directed at the other (M). HO theorists are normally of the belief that such mental states are identical with brain states, but they tend to treat this matter as a further second step reduction for empirical science. Moreover, it is important to keep in mind that all HO theorists are united in the belief that their approach can better explain consciousness than any purely first-order representational (FOR) theory, such as those presented by Dretske (1995) and Tye (1995).

Two terminological distinctions should also be made at this point (see e.g. Rosenthal 1990; 1993a). First, a distinction is often made between creature consciousness and state consciousness. The former recognizes that we often speak of whole organisms as conscious or even as simply “awake.” The latter recognizes that we also speak of individual mental states as
conscious. Explaining state consciousness is the primary focus for most researchers, though there are also, no doubt, some interesting connections between state and creature consciousness. Second, some authors also utilize a distinction between intransitive and transitive consciousness. We sometimes use the word ‘conscious’ as in our being ‘conscious of’ something. This is the transitive use. On the other hand, we also have the “x is conscious” locution and, due to the lack of a direct object, this is called the intransitive use. Thus, analyzing state consciousness in terms of this distinction leads to the idea that a mental state is intransitively conscious just in case we are transitively conscious of it. We must be careful to guard against any suggestion of circularity here (see Section 3A below); it is important to keep in mind that the HO (transitively conscious) state is not normally itself intransitively conscious. There is the potential for confusion, however, so some HO theorists prefer to speak of the HO (nonconscious) “awareness of” the lower-order state instead of invoking the transitive use of ‘consciousness.’

2. Different Kinds of HO Theory

There are various kinds of HO theory depending on how one understands the HOR. The most common division is between higher-order thought (HOT) theories and higher-order perception (HOP) or higher-order experience (HOE) theories. HOT theorists, such as David M. Rosenthal, think it is better to understand the HOR as a thought of some kind. HOTs are here treated more like cognitive states involving some kind of conceptual component. The latter argue that the HOR is closer to a perceptual or experiential state of some kind (e.g. Lycan, 1996) which does not require the kind of conceptual content invoked by HOT theorists. Due to Kant (1781/1965), HOP theory is also sometimes referred to as “inner sense theory” as a way of emphasizing its sensory or perceptual aspect. Although HOT and HOP theorists agree on the need for a HOR theory of consciousness, they are sometimes concerned to argue for the superiority of their respective positions (such as in Rosenthal, this volume; Lycan, this volume). Some philosophers, however, have argued that the HOP theory ultimately reduces to the HOT theory (Güzeldere, 1995) Others have argued that the difference between these theories is perhaps not as important or as clear as some think it is (Gennaro, 1996; Van Gulick, 2000).
Finally, Peter Carruthers (2000) has recently proposed that it is better to think of HOTs as *dispositional* states instead of the standard view that the HOTs are *actual*, though he also understands his dispositional HOT theory to be a form of HOP theory (Carruthers, this volume). His overall basic idea, however, is that the conscious status of an experience is due to its *availability* to higher-order thought. A key idea is that “conscious experience occurs when perceptual contents are fed into a special short-term buffer memory store, whose function is to make those contents available to cause HOTs about themselves.” (Carruthers, 2000: 228). So some first-order perceptual contents are available to a higher-order “theory of mind mechanism,” which transforms those representational contents into conscious contents. Thus, no actual HOT occurs. Instead, some perceptual outputs acquire a dual intentional content; for example, a conscious experience of red not only has a first-order (analog) content of ‘red,’ but also has the higher-order content ‘seems red’ or ‘experience of red.’ Carruthers also makes interesting use of so-called “consumer semantics,” such as teleosemantics (Millikan, 1984) and inferential role semantics (e.g. Peacocke, 1992) in order to fill out his theory of phenomenal consciousness. The basic idea is that the content of a mental state depends, in part, on the powers of the organisms which “consume” that state, e.g. the kinds of inferences which the organism can make when it is in that state. Daniel Dennett (1991) is sometimes credited with an earlier, though somewhat different, version of a dispositional account (see Carruthers, 2000, chapter ten). Carruthers’ dispositional theory is criticized, in this volume, by Gennaro and Rosenthal.

One other source of disagreement within HO theories concerns the issue of whether or not the HO state should be understood as *extrinsic* to (i.e. entirely distinct from) its target mental state. This is the view defended by David Rosenthal. On the other hand, several authors have recently challenged this assumption. For example, following in the tradition of Brentano (1874/1973), Gennaro (1996: 21-30) has argued that, when one has a first-order conscious state, the HOT is better viewed as *intrinsic* to the target state, so that we have a complex conscious state with parts. Gennaro calls this the “wide intrinsicality view” (WIV) and he has also recently argued that Jean-Paul Sartre’s theory of consciousness can be understood in this way (Gennaro,
2002). Robert Van Gulick (2000, this volume) has also explored the alternative that the HO state is part of the “global” conscious state. He calls such states “HOGS” (= higher-order global states) within which the lower-order state is “recruited” and becomes conscious. Both Gennaro and Van Gulick have suggested that conscious states can be understood materialistically as global states of the brain, and it would be better to treat the first-order state as part of the larger complex brain state. This idea was also briefly explored by Thomas Metzinger who focused on the fact that consciousness “is something that unifies or synthesizes experience.” (Metzinger, 1995: 454; cf. Gennaro, 1996: chapter three) Metzinger calls this the process of “higher-order binding” and thus uses the acronym ‘HOB.’ A point of emphasis in all of these alternatives is on the concept of global meta-representation within a complex brain state.¹

3. Some Important Issues

There are a number of reoccurring themes in the literature on HO theories, some more important and controversial than others. I cannot hope to do all of them justice in a short introduction, but it is worth at least briefly mentioning the following six.²

A. Circularity and Regress

A common initial objection to HOR theories is that they are circular and lead to an infinite regress. For example, it might seem that the HOT theory results in circularity, i.e. by defining consciousness in terms of HOTs. It also might seem that an infinite regress results because a conscious mental state must be accompanied by a HOT, which, in turn, must be accompanied by another HOT ad infinitum. However, the standard reply from the HOT theorist is that when a conscious mental state is a first-order world-directed state the higher-order thought (HOT) is not itself conscious; otherwise, circularity and an infinite regress would follow. Moreover, when the HOT is itself conscious, there is a yet higher-order (or third-order) thought directed at the second-order state. In this case, we have introspection which involves a conscious HOT directed at an inner mental state. When one introspects, one's attention is directed back into one's mind. For example, what makes my desire to write a good introduction
a conscious first-order desire is that there is a (nonconscious) HOT directed at the desire. In such a case, my conscious focus is directed at the introduction and so I am not consciously aware of having the HOT from the first-person point of view. When I introspect that desire, however, I then have a conscious HOT (accompanied by a yet higher, third-order, HOT) directed at the desire itself (on this point see Rosenthal, 1986: 337-8). So instead of threatening the HOT theory, this issue actually brings out an important subtlety of any viable HOR theory. Although most opponents of HO theories are inclined to accept the standard reply to this objection, some are still not so easily satisfied (Rowlands, 2001).

B. Animal and Infant Consciousness

Perhaps the most common objection to HO (especially HOT) theories is that various animals are not likely to have to the conceptual sophistication required for HOTs, and so that would render animal (and infant) consciousness very unlikely (Dretske, 1995; see also Seager, this volume). Are cats and dogs capable of having such complex higher-order thoughts to the effect that “I am in mental state M.”? This is normally treated as a problem for HOT theory because the vast majority of us believe that animals and infants are clearly capable of having conscious states. Although most who bring forth this objection are not HO theorists, one notable HO theorist actually embraces the conclusion that animals and infants do not have phenomenal consciousness (Carruthers, 1989). Gennaro (1993, 1996) has replied to Carruthers on this point; for example, it is argued that the HOTs need not be as sophisticated as it might initially appear and that there is evolutionary and comparative neurophysiological evidence supporting the conclusion that animals have conscious mental states. It seems fair to say that most HO theorists do not wish to accept the absence of animal or infant consciousness as a consequence of holding the theory. And even the mere perception that HOT theories of consciousness can lead to this consequence clearly causes many to shy away from any form of HO theory. The debate continues, however, in Carruthers (2000) and Gennaro (this volume). This issue also brings out a host of related important questions: Must all conscious mental states have conceptual content? How should concepts be understood and defined in the context of HOT theory? What exactly is
the difference between perception and thought? How “rich” in content are our conscious states? Are there degrees of self-consciousness?

C. The Problem of the Rock

Another objection to HO theories has been referred to as the “problem of the rock” (Stubenberg, 1998) and the “generality problem” (Van Gulick, 2000, this volume), but it is perhaps originally due to Alvin Goldman (Goldman, 1993). When I have a thought about a rock, it is certainly not true that the rock thereby becomes conscious. So why should I suppose that when I think about a lower-order mental state (M), it becomes conscious? Indeed, why should being the intentional object of a meta-state confer consciousness on M? This seems puzzling to many and the objection forces HO theorists to explain how adding the HO state changes a nonconscious state to a conscious one since having a similar state directed at outer objects does not render them conscious. There have been, however, a number of responses to this kind of objection (Rosenthal, 1990/1997; Lycan, 1996; Van Gulick, 2000; Gennaro, unpublished). A common theme is to remind the objector that there is a principled difference between the objects of the HO states in question. Rocks and the like are not mental states in the first place, and so HO theorists are first and foremost trying to explain how a mental state becomes conscious. Nonetheless, the prospect of a damaging reductio of HO theories looms if this objection cannot be met properly.

D. The Hard Problem of Phenomenal Consciousness

The above objection leads somewhat naturally to the following problem: In the spirit of David Chalmers’ (1995; cf. Shear, 1997) discussion of what he calls the “hard problem of consciousness,” it might be asked just how exactly any HO theory really explains the subjective aspect of conscious experience (Stubenberg, 1998; Siewert, 1998). How or why does a mental state come to have a first-person qualitative “what it is like” aspect by virtue of the presence of a HOR directed at it? I think it is fair to say that some HO theorists have been slow to address this problem, though a number of overlapping responses have emerged. Some have argued that this objection simply misconstrues the main and more modest purpose of (at least, their) HO theories.
The claim here is that HO theories are theories of consciousness only in the sense that they are attempting to explain what differentiates conscious from nonconscious states, i.e. in terms of a higher-order awareness of some kind. A full account of ‘qualitative properties’ or ‘sensory qualities’ (which can themselves be nonconscious) can be found elsewhere in their work, but is independent of their theory of consciousness (Rosenthal, 1991; Lycan, 1996, 2001). Thus, a full explanation of phenomenal consciousness does require more than a HO theory, but that is no objection to HO theories as such. Another response is that proponents of the hard problem unjustly raise the bar as to what would count as a viable explanation of consciousness so that any such reductivist attempt would inevitably fall short (Carruthers, 2000). Part of the problem also, then, is a lack of clarity about and disagreement on what would even count as an explanation of consciousness (Van Gulick, 1995). Anyone familiar with the literature knows that there are also significant terminological difficulties in the use of various crucial terms (see Byrne, this volume), which also sometimes inhibits progress and mutual understanding on this matter.³

E. Misrepresentation

A further important objection to HO approaches is the question of how such theories can explain cases where the HO state might misrepresent the lower-order (LO) mental state (Byrne, 1997; Neander, 1998; Levine; 2001). After all, if we are dealing with a representational relation between two states, it seems possible for misrepresentation or malfunction to occur. If it does, then what explanation can be offered by the HO theorist? If my LO state registers a red percept and my HO state registers a thought about something green due, say, to some neural misfiring, then what happens? It seems that problems loom for any answer from a HO theorist and the cause of the problem has to do with the very nature of the HO theorist’s belief that there is a representational relation between the LO and HO states. For example, if the HO theorist takes the option that the resulting conscious experience is reddish, then it seems that the HO state plays no role in determining the qualitative character of the experience. This is an objection that must be taken seriously and it forces HO theorists to be clearer about just how to view the relationship between the LO and HO states. A reply to this objection is offered by Gennaro (in this volume).
F. The causal and/or inferential relation between the lower and higher-order states

A final pair of related issues concerns other aspects of the relationship between the LO and HO states. First, it is agreed upon by all HO theorists that the HO state must become aware of the LO state non-inferentially. We might say, then, that the HOR must be caused non-inferentially by the LO state in order to make the LO state conscious. The point of this condition is mainly to rule out alleged counterexamples to HO theory, such as cases where I become aware of my nonconscious desire to kill my boss because I have consciously inferred it from a session with a psychiatrist or where my envy becomes conscious after making inferences about my own behavior. The characteristic feel of such a conscious desire or envy may be absent in these cases, but, since awareness of them arose via conscious inference, the HO theorist attempts to account for them by adding this non-inferential condition. Of course, some still ask: why should it matter so much how I become aware of the LO state?

Second, and perhaps more controversial, is whether or not to understand the LO state as the only or primary “cause” of the HO state, in some important sense of that term (see e.g. Rosenthal, 1993b; Gennaro, 1996: 73-75). HO (especially HOP) theorists sometimes speak that way and there are advantages to such a view; for example, it would be a natural way to explain how the HO state gets directed at or refers to the appropriate LO state. However, there are also reasons to shy away from such a straightforward causal connection and to adopt the weaker notion of “accompaniment” or “co-occurrence” between states. In addition to avoiding the more general vexed problems of causality and reference within the context of HO theory, mere accompaniment would seem sufficient to explain state consciousness (especially according to HOT theorists). In any case, HOT theorists, such as Rosenthal, tend to hold that the LO state is somehow implicated in the causing the HOT, but there must be other factors that figure in causing it as well. Additionally, it is possible that there is instead a reliable tracking condition (in the brain) which obtains between the LO state and the HOT directed at it. For example, the LO state and its HOT might have a common cause such that whenever the former is produced the latter is also (typically and reliably) caused as well. Needless to say, however, not everyone
is satisfied with the HO theorist’s treatment of these issues (Francescotti, 1995).

4. The Essays

This book is divided into two general parts. Part I contains essays by authors who have defended some form of HO theory. Although they are often concerned to discredit one or more of the other HO theories, it must be kept in mind that they are strongly united in their agreement in the superiority of HO theories over, for example, various first-order (FO) accounts of consciousness (Dretske, 1995; Tye, 1995). Part I begins with a chapter entitled “Varieties of Higher-Order Theory” (Chapter 2) by David M. Rosenthal who argues for his preferred version of “extrinsic HOT theory” by systematically reviewing and critiquing alternative HO theories, including HOP and dispositional HOT theory. In the following chapter entitled “Higher-Order Thoughts, Animal Consciousness, and Misrepresentation: A Reply to Carruthers and Levine” (Chapter 3), Rocco J. Gennaro defends HOT theory at length against two of the key problems explained in the previous section: animal consciousness and misrepresentation. In “Higher-Order Global States (HOGS): An Alternative Higher-Order Model of Consciousness” (Chapter 4), Robert Van Gulick further develops his HOGS version of HO theory. For example, Van Gulick attempts to explain how HOGS theory can avoid some of the difficulties with standard HOP and HOT theory, and why his theory should indeed be understood as a kind of higher-order theory instead of a FO theory. In Chapter 5, William G. Lycan argues for “The Superiority of HOP to HOT.” In addition to responding to critics, he argues that HOP theory is preferable on a number of grounds, such as its ability to account for so-called “recognitional concepts” and the voluntary control we find in consciousness. Peter Carruthers’ cleverly titled “HOP over FOR, HOT Theory” (Chapter 6) argues for what he calls “the dispositionalist HOT version of HOP theory.” Carruthers argues that FOR and actualist HOT theories cannot give an adequate account of purely recognitional concepts of experience; nor can they properly explain the distinction between conscious and nonconscious perceptual states, especially when we examine such phenomena as blindsight and visual agnosia. In Chapter 7, Edmund T. Rolls presents a
somewhat different form of HOT theory which he calls “A Higher-Order Syntactic Thought (HOST) Theory of Consciousness” primarily against the background of a theory of emotion. Citing a wealth of empirical literature, he argues, for example, that it is best to take an information processing and brain design approach to consciousness such that the HOTs in question are directed at semantically based thoughts and that HOTs have important evolutionary adaptive value. Finally, in “Assumptions of a Subjective Measure of Consciousness: Three Mappings” (Chapter 8), Zoltán Dienes and Josef Perner use the HOT theory as a tool by which they analyze the appropriate use of various subjective measures of conscious awareness, such as the so-called “zero-correlation criterion.” They focus on the confidence-accuracy relationship and then use the zero-correlation criterion in both subliminal perception and implicit learning.

Part II contain papers by authors who do not subscribe to any form of HO theory. Thus, their papers not only criticize one or more HO theory, but often also urge us to look for an altogether different solution to the problem of consciousness. Part II begins with an essay entitled “What Phenomenal Consciousness is Like” by Alex Byrne (Chapter 9). He first does us the enormous favor of helping us to navigate through the terminological jungle that we find in the literature, especially regarding various attempts by HOR theorists to distinguish between “experiences” and “conscious experience.” Byrne then argues that HOR theories are mistaken primarily on the basis of rebutting those attempts. Robert Lurz argues that the assumed dichotomy between FOR and HOR theories of consciousness is a false one. In “Either FOR or HOR: A False Dichotomy” (Chapter 10), Lurz rejects both of these alternatives, and then argues for a third option he calls “same-order representationalism” (SOR) whereby what makes a mental state (M) conscious is that one is aware of the intentional content of M. William Seager takes “A Cold Look at HOT Theory” (Chapter 11) by focusing on whether or not animals can attribute mental states to other animals. Citing much important empirical literature, he argues that since most animals cannot do so, they therefore cannot also self-attribute mental states and, hence, on HOT theories, they cannot be conscious beings. As we saw earlier (section 3B), this is a conclusion which most HO theorists wish to avoid. In Chapter 12, Valerie Gray Hardcastle
pulls no punches against HO theory, and particularly against David Rosenthal’s arguments for HOT theory. As the title suggests, her “HOT Theories of Consciousness: More Sad Tales of Philosophical Intuitions Gone Astray,” primarily aims at exposing weaknesses underlying the entire HO approach, such as the “intuition” that we are “aware” that we are in conscious states, that HOT theory is empirically adequate, and the preference for reducing consciousness in mentalistic terms. In “A Few Thoughts Too Many?” (Chapter 13), William S. Robinson uses the fact that we often engage in subvocal speech to argue that, amongst various explanations, a simpler view not including any HORs should be preferred on grounds of parsimony. Significant doubts are thus raised about the claim that HORs are required for consciousness. Donelson E. Dulany, in “Higher Order Representation in a Mentalistic Metatheory” (Chapter 14), critiques various HO theories and offers an alternative “mentalistic” metatheory of consciousness. Dulany also challenges the current orthodoxy that there are unconscious mental states, such as unconscious perceptions, which form a key basis for any HO theory. Finally, in “Ouch! An Essay on Pain” (Chapter 15), Christopher S. Hill presents a theory of pain such that awareness of pain is akin to such paradigmatic forms of perceptual awareness as vision and hearing. Hill ultimately considers his theory in relation to HOP theory, but, despite some apparent similarities, argues that they are importantly different and for the superiority of the former over the latter.

NOTES:
1. Thomas Natsoulas also has a continuing series of papers defending intrinsic theory, beginning with Natsoulas, 1996. See also Kriegel (2003) and many of the essays in Kriegel and Williford (forthcoming).
2. Virtually all of them are addressed to some extent in Byrne, 1997.
3. For another HO theorist’s attempt to address the hard problem, see Gennaro (unpublished).
4. For more on this issue, see the exchange between Natsoulas (1993) and Rosenthal (1993b).
REFERENCES:


Gennaro, R. (unpublished). The HOT theory of consciousness: between a rock and a hard place?


