

# ASSC 10: Higher-Order Theories of Consciousness

© David M. Rosenthal  
City University of New York,  
Graduate Center  
Philosophy and Cognitive Science  
St Anne's Oxford, 23 June 2006

Association for the Scientific  
Study of Consciousness



ASSC 10  
June 23-26 2006, Oxford

## Organizational Outline

### FIRST HALF — Alternative Higher-Order Theories:

- Ia. Higher-order theories—as against *first-order* theories (Dretske, Searle).
- Ib. Higher-order perceiving (*inner sense*) and *dispositional* higher-order states.
- Ic. Distinct higher-order states versus *intrinsic higher-order content*.
-

## SECOND HALF — Applications of Higher-Order Theory:

- IIa. Qualitative consciousness: Can higher-order states result in there being something it's like for one?
- IIb. Function and origin: What role does consciousness play on higher-order theories? Is it adaptive? Why are there any higher-order states?
- IIc. Consciousness and speech and the timing of consciousness.

## Preliminary

- Each segment should (with any luck) be about 30 minutes—plus or minus.
- As I mentioned in an email, there's a bibliography of everything I'll mention in my presentation at  
 <http://web.gc.cuny.edu/cogsci/assc-bib.pdf>
- I sent email to all those whose email addresses I was given.  
If you didn't get that, please do send me an email so I'll have your address:  
[davidrosenthal@nyu.edu](mailto:davidrosenthal@nyu.edu)

## Ia. Higher-Order versus First-Order Theories

- Higher-order theories are all theories about what it is for a mental state to be conscious—what it is for, e.g., thoughts, sensations, perceptions, desires, and emotions to be conscious.
- Assuming that some mental states occur without being conscious, that means explaining the difference between mental states that are conscious and those which are not conscious.

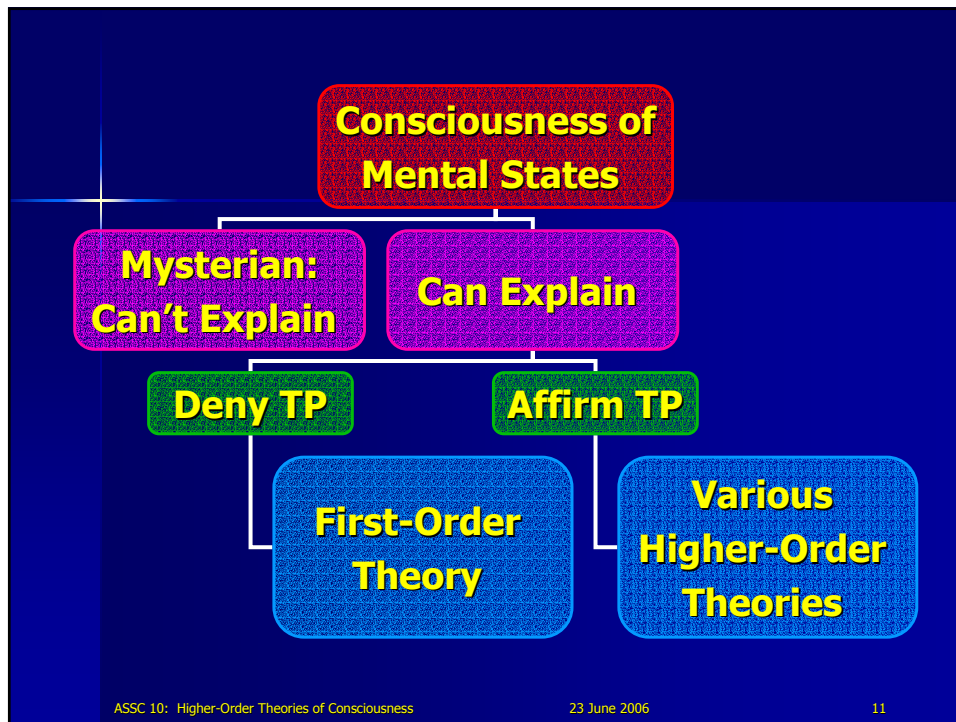
- Some hold that a state's being conscious goes with a creature's being conscious.
- But some mental states occur in waking, conscious life without those states' being conscious;  
so a state's being conscious can't be just its being a state of a conscious creature.
- Similarly, thoughts and perceptions make one conscious (= aware) of things (what I'll call transitive consciousness).
- So, since those states can occur without being conscious, a state's being conscious cannot be the same as a state's making one conscious of something. (More soon.)

- One natural line about what it is for a mental state to be conscious is that  
**A mental state is conscious only if one is in some way aware of it.**
- I'll call that the transitivity principle (TP)
- That offers a necessary condition for a state to be conscious ("state consciousness").
- TP seems right, since when an individual is not in any way conscious of a state, we do not regard that state as being a conscious state.

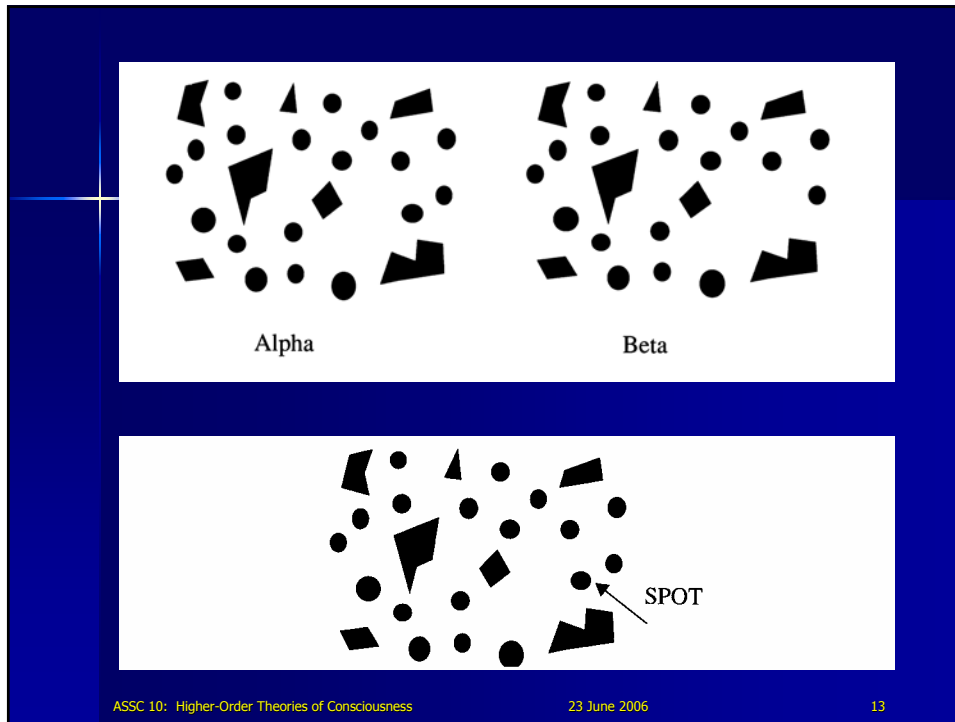
- A quick second argument in support of TP: Conscious states are those we can report being in—in some intuitively immediate way.
- Moreover, reports express thoughts— in this case, thoughts that one is in the state in question— thoughts that make one aware of the state being reported.
- Since states are conscious when a report could occur—i.e., when one has a thought that such a report could express—states are conscious when one has such a (higher-order) thought.

- TP by itself plainly cannot provide a sufficient condition for a state to be conscious:
  - We're sometimes conscious (= aware) of our own states even when those states fail to be conscious.
- We may learn that we're happy or angry from somebody's telling us—without that state's thereby coming to be conscious.
- And one may learn that one sees something by experimental means, and by applying standard theories (say, about masked priming) to oneself.

- So TP at best provides a necessary, not a sufficient condition for state consciousness—for a state's being conscious.
- TP says that a state is conscious only if one is conscious of that state.
- But if we can specify the way one is conscious of a state when that state is conscious, that should give us a condition that's both necessary and sufficient—and thus, in effect, a theory of consciousness.
- The variety of higher-order theories differ in how they seek to specify the way one is conscious of one's mental states when those states are conscious.



- John Searle and Fred Dretske have contested TP.
- Searle denies that we're conscious of our mental states because we never observe them (Rediscovery 95-6).
- But observing isn't the only way to be conscious of something.
- Dretske (1993; cf. 1995, ch. 4) has a more elaborate, and influential, argument. Suppose that one consciously sees two scenes that are exactly alike except for the presence of something in one scene that's missing in the other:



- Dretske assumes that, since we see each scene consciously, we see each part of each consciously. That seems right.
- So, not only is the experience of each scene conscious; each part of each experience is conscious.
- Still, since we don't (consciously) notice the difference between the two scenes, we aren't, he argues, conscious of the part of the experience of alpha that is an experience of "spot":
- So that's a conscious experience of which we aren't conscious—and hence a counterexample to TP.





- This kind of example may recall *change blindness*—about which something in a moment.
- But the case itself is *not at all esoteric*: We consciously see scenes that change all the time, without (consciously) noticing the absence or presence of something in a subsequent conscious viewing.
- Since a state's being conscious cannot, on Dretske's view, consist in one's being conscious of that state, Dretske needs *an alternative line about what it is for a mental state to be conscious*.

- His line is that a state is conscious just in case it's *a state in virtue of being in which one is conscious of something*.
- This has come to be called a *first-order* theory of state consciousness (not "same order"; that's a label for intrinsicism).
- This is the most serious challenge to higher-order theories—maybe the only.
- It echoes somewhat the more specific claim that, in the case of qualitative states, having qualitative character is of a piece with being conscious (e.g., Levine 2001, Chalmers 1996, Neander 1998, Jackson 1982, 1986)—about which more in Parts Ic and IIa.



- But this theory is arguably wrong:  
*Perceptions and (suitable) thoughts make us aware of things even when those states are not conscious states.*  
We're aware of such things—*but not consciously aware of them.*
- How else could such states have the effects they have on our psychological lives—e.g., through priming and the like?
- So Dretske's theory has a downside:  
If a state's being conscious consists in its making one conscious of something, *then perceptions and (suitable) thoughts cannot occur without being conscious.*

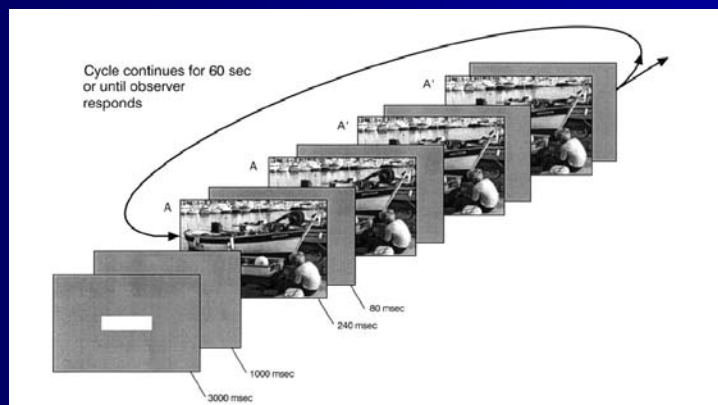
- Dretske tries to get around this (2006) by urging that perceiving is conscious only if subjects can cite what they perceive as a justifying reason for doing something.
- And, when subjects perceive subliminally (or in blindsight), they can't do that.
- This is meant to reflect the intuitive tie between consciousness and rationality.
- But one will be able to cite what one perceives as a reason for action only if one is conscious of the perceiving.
- So *being conscious of one's state is still the basic factor in that state's being conscious.*

- A first-order view also has difficulty in explaining how one's perceptual states, e.g., can diverge from what it's like for one to be in those states, as in change blindness (more about that specific difficulty in a moment).
- But Dretske's understanding on the "spot" kind of example is in any case tendentious.
- And without that tendentious understanding, the case isn't a counterexample to TP at all, and hence not to higher-order theories.

- It could be that one is conscious of one's experience of spot (and conscious of one's experience of the white area where spot was)—  
but not conscious of either experience as being different from the other experience.
- It could be, e.g., that one is conscious of the experience of spot just as part of the experience of the overall scene, and not as an experience of a black dot.
- Being conscious of something does not entail being conscious of it accurately or fully—or indeed in any particular way.

- So Dretske's case is not an uncontroversial counterexample to TP.
- Dretske's discussion of change blindness is tendentious in much the same way.
- First, a word about change blindness.
- Change blindness occurs when a change in a scene occurs, sometimes relatively salient, without being consciously noticed.
- There are at least three paradigms, which depend on attention (Simons), masking (Rensink), and saccading (Grimes). I'll say something about the last 2 of these.

- Rensink shows subjects scenes alike but with a reasonably salient change, which alternate, separated by a brief mask (solid gray, mudsplash, etc.):



- Subjects often fail to report any change.

- John Grimes used eye trackers so that changes occur during saccades, during which no retinal input reaches visual cortex ("On the Failure to Detect Changes in Scenes across Saccades," in Akins, 1996).
- Salient changes often went unobserved for a significant number of changes—e.g., a parrot that changes between red and green, or people's heads being switched.
- *It is arguable that change blindness sustains TP,* and hence higher-order theories of consciousness, contrary to Dretske's claims.

- When the Grimes parrot stimulus shifts between red and green, the actual visual state of subjects presumably shifts accordingly.
- But *what it's like for them often doesn't.*
- *So the way they are conscious of their visual states diverges from the actual states themselves.*
- And that divergence points to some kind of higher-order theory:  
*The need to distinguish the state itself from the way one is conscious of that state.*

- Dretske seeks to get around this by arguing that subjects are not, in such cases, actually blind to any change.
- That's because the changes (in the Grimes paradigm) occur during saccades, and so the changes are concealed from subjects.
- So Dretske argues that subjects are not blind to those changes, but only to differences that result from those changes ("Change Blindness," 2004).
- One might argue the same in Rensink's mudsplash paradigm, in which a mask intervenes in the presented change.

- But this ignores the way one's visual state may diverge from what it's like for one to be in that state.
- As Dretske concedes, one's seeing corresponds to the stimulus; one's visual state is of a green parrot when that's what one is presented with, and so forth.
- So, when subjects fail to report a change, what it's like for them to see the stimulus diverges from the seeing itself.
- It seems to subjects that they see a red parrot even though they're in a "green parrot" visual state.

- Since what it's like for subjects doesn't track changes in their actual visual states, subjects are, after all, blind to those changes.
- And changes in what it's like for one are independent of saccades (and masks).
- We'll return (Part IIa) to the claim that the state consciousness of qualitative states is inseparable from their qualitative character.
- But it seems in any case that TP and the higher-order theories based on that principle can withstand the challenge from Dretske's first-order account.

➤ Discussion? Questions?

➤ On, then, to Ib:

## Ib. Inner Sense, and Higher-Order Dispositions

- Higher-order theories all rely on TP:  
A mental state is conscious only if  
one is in some way aware of it.
- As mentioned earlier, this at best gives  
only a necessary condition for a state's  
being conscious.
- One might be conscious of some mental  
state one is in by taking somebody else's  
word for it or by applying a theory to  
oneself.

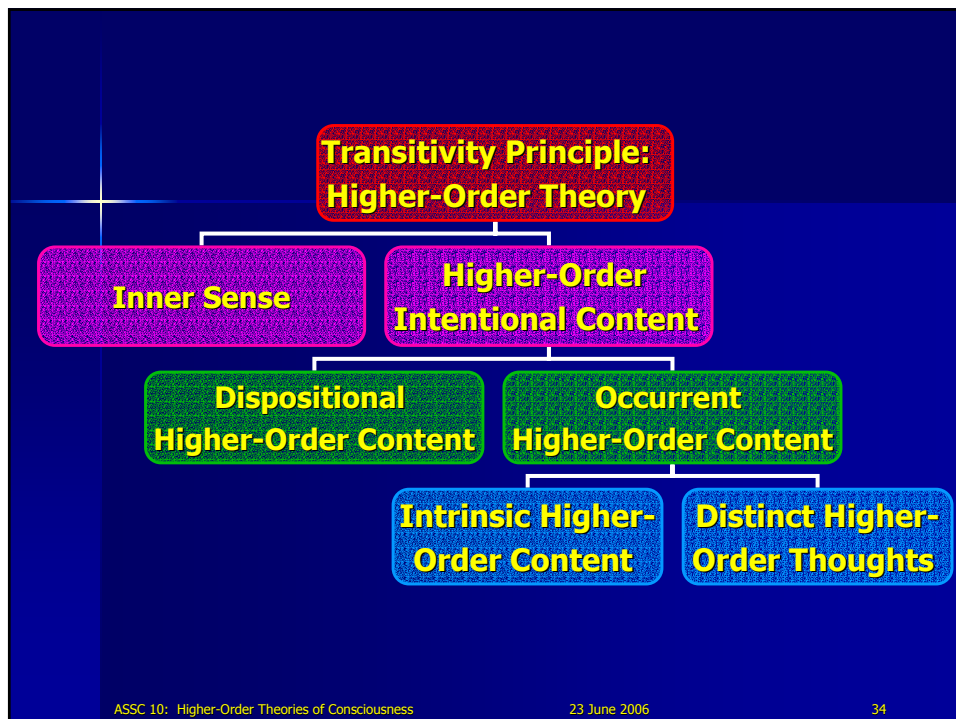
- In those cases, the state one is conscious  
of oneself as being in may well fail to be  
a conscious state.
- Traditionally these cases are ruled out by  
specifying that one be immediately or  
directly conscious of the target state.
- Thus, e.g., Descartes: "[T]he word  
'thought' applies to all that exists in us in  
such a way that we are immediately  
conscious of it" (AT VII, 160; see also AT VII 52).
- Indeed, it was rare before the late 19<sup>th</sup>  
century ever to describe mental states as  
conscious, as against as being states of  
which we are immediately conscious.



- But the traditional stipulation that we be immediately or directly conscious of our conscious states overshoots.
- A state will be conscious, on our intuitive understanding of that notion, if it simply seems to one that one is immediately or directly conscious of it.
- Compare perceiving: It typically seems that nothing intervenes or mediates between our perceiving and what we perceive.
- Similarly, a state is conscious if it seems that nothing mediates between that state and one's awareness of it.

- This weakened stipulation handles cases in which one is aware of one's own state simply by believing what somebody else tells one or simply by applying a theory to oneself.
- 'Simply' because one might be aware of the state both in a way that seems, subjectively, unmediated— and also from what somebody else says or applying a theory.
- It's enough for a state to be conscious if one would be conscious of it without any conscious observation or conscious appeal to theory.

- We can capture this by saying that a state is conscious if one is conscious of that state independently of any conscious inference.
- That's because conscious inference figures in all of the counterexamples to TP that we've been considering.
- This comes close to providing a condition for a state's being conscious that is not only necessary, but sufficient as well.
- But we need still to know in what way one is conscious of one's mental states when those states are conscious states— i.e., how TP is implemented:



- The standard and most frequent proposal for implementing TP is that *we sense or perceive our conscious states*.
- Thus John Locke: "Consciousness is the perception of what passes in a Man's own mind" (Essay 1975/1700 II, 1, 19);  
and our ideas of perceiving, thinking, and the like stem from an "internal Sense" (II, 1. 4) .
- Similarly, Kant speaks of an "inner sense" (der innere Sinn, K.d.R.V., 1787/1996, A22/B37).
- This goes back to Aristotle, who claimed that we perceive that we perceive (DA Γ2, 425b12-20; Met Λ9, 1074b35-6; EN IX 9, 1170a29-34).

- David Armstrong (1978/1980) and Bill Lycan (1996) are the best known contemporary exponents of such a view, though their view is combined with an intentionalist view of sensing and perceiving, which skews things.
- Inner sense is inviting as a way to implement TP because *sensing is what first comes to mind in connection with being conscious of things*.
- But sensing and perceiving are *not the only ways* of being conscious of things, and inner sense faces *serious difficulties as an implementation of TP*.

- *Sensing and perceiving invariably involve qualitative character.*

Perceiving has intentional content as well, whereas sensing has only qualitative character.

- But if a mental state occurs without qualitative character, it's not sensing or perceiving.
- And it seems clear that *whatever higher-order states implement TP, they do not exhibit qualitative character* (Rosenthal 1997, 2004).

- The principal difficulty has to do with *what mental qualities could occur in the higher-order sensing or perceiving.*
- Aristotle first raised this issue (DA Γ2), urging that the higher-order mental qualities *could be the same as the first-order mental qualities of the target.*
- But that doesn't explain how higher-order sensing would work with *nonqualitative, intentional states,* such a thinking.
- And *we're never aware of any mental qualities two times over*—once as qualities of first-order states, and a second time as qualities of higher-order states.

- One might explain why we're not aware of mental qualities in duplicate by noting that the higher-order states are seldom themselves conscious states.
- As noted in Part Ia, a state can make one conscious of something without itself being a conscious state—as in subliminal perceiving.
- But sometimes the higher-order state is conscious, as when we introspect. And even then we aren't conscious of any mental qualities that belong to our higher-order states.

- Also: Some experiences are cross-modal; what mental qualities could be exhibited by higher-order states in virtue of which crossmodal experiences are conscious?
- When we turn to qualitative consciousness (Part IIa), I'll propose an account of mental qualities that sustains the view that higher-order states have no mental qualities.
- Lycan has argued that the higher-order states resemble perceiving in various other, nonqualitative ways (2004, 2006).
- I find his suggestions unconvincing, but won't take that up now (Rosenthal 2004).

- Inner-sense theory often rests on a claim about monitoring: That the higher-order states serve to monitor their targets.
- It's thought that such monitoring enhances reasoning and planning.
- But it's not clear why that would be:
- Inference and reasoning hinge on causal ties that earlier intentional states have to subsequent states—simply in virtue of their content, not their being conscious.
- Higher-order states might tune up those causal ties, but they also sometimes might interfere and disrupt them, making the first-order causal ties less efficient.

- Indeed, the efficiency and accuracy of cognitive processing may be degraded by paying conscious attention to that processing, including verbal descriptions or responses (Dijksterhuis, et al 2006, Wilson and Schooler 1991, Schooler, Ohlsson, and Brooks 1993).
- The monitoring model is sometimes put forth to explain why state consciousness occurs at all—  
i.e., what role is played by a state's being conscious, and why should states ever be conscious.
- But we'll see in Part IIb that there are other ways to explain that.

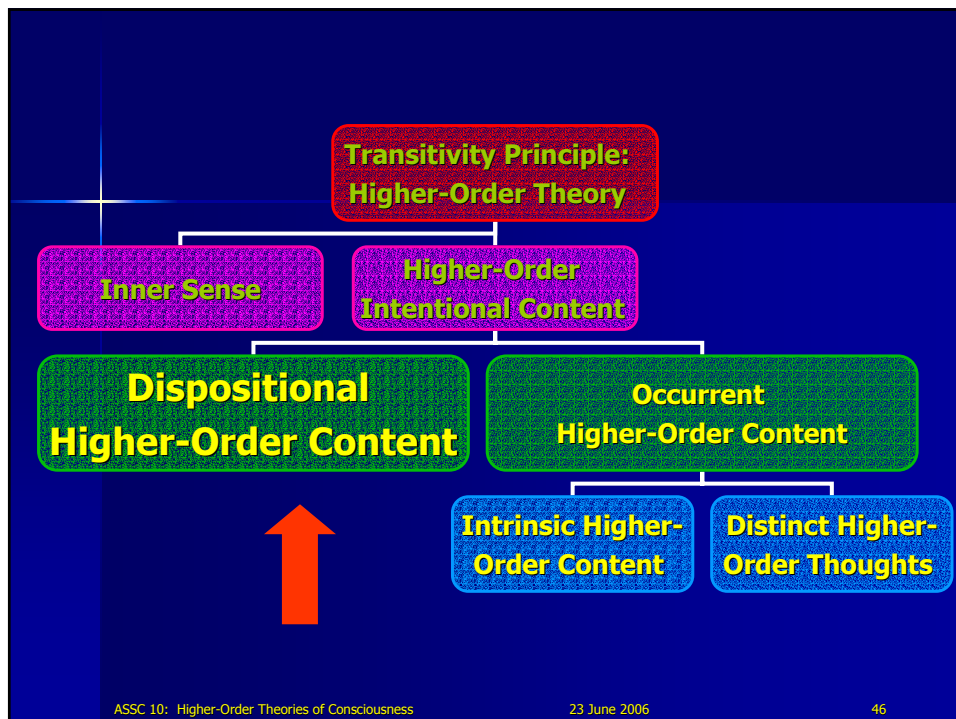
➤ Questions or discussion before going onto dispositional higher-order theory?

➤ Then onto that...

- If the higher-order states that implement TP lack qualitative character, they must be *purely intentional states*.
- Assertoric thoughts do make us conscious of things—at least *when they represent the things as being present*.
- A thought about Saturn or Caesar does not intuitively make one conscious of those things, since the thought represents them as being distant in time or space.
- But a thought about something in this room—independent of any perceiving—does make one conscious of that thing.



- So we can seek to implement TP by way of higher-order thoughts (HOTs).
- Peter Carruthers has objected to this, mainly because having HOTs for each conscious state, he urges, would overload our cognitive capacity (2000, 2001).
- Rather than overload that capacity with actual HOTs, he argues, we should instead simply posit dispositional HOTs—i.e., dispositions to have actual HOTs.
- Carruthers has developed a rather elaborate version of this dispositional HOT theory (2000, 2005) .



- It's unclear, however, why actual HOTs would result in any cognitive overload. We have cortical capacity to spare; so that's not the issue.
- Carruthers also argues that there would be no point to the actual occurrence of so much higher-order cognitive activity.
- This has to do with the function of consciousness; so let's postpone that second criticism until Part IIb.
- In any case, it's unclear why dispositions would occasion less overload than actual HOTs, since dispositions also require cognitive and cortical capacity.

- But these things aside, simply being disposed to have a thought about something cannot by itself make one conscious of that thing (Rosenthal 2004).
- So a dispositional higher-order theory cannot implement TP.
- Carruthers seeks to get around this point by invoking a particular theory of intentional content, on which a state's content is, in part, a matter of its propensity to cause behavior and other states.
- And he urges that this also relieves cognitive overload.

- Some states, Carruthers argues, have the potential to cause actual HOTs because of their connection to a dedicated mind-reading module.
- And that potential gives those first-order states themselves a higher-order content, in virtue of which they are conscious.
- This higher-order content imposes no demands on one's cognitive capacity beyond the first-order states' tie to the mind-reading system.

- This solution holds the theory hostage to a controversial view of intentional content, and it changes the theory to one that's first-order and intrinsicist.
- And there's also a pressing difficulty:
- States are often conscious at one time but not another.
- And, since Carruthers's solution demands that a state is conscious because, solely in virtue of its content, it has, itself, higher-order content that implements TP, no state that actually occurs consciously would be able to occur nonconsciously.

- A parallel difficulty about implementing TP also occurs in connection with Ned Block's notion of access consciousness, on which a state is access conscious if it's "poised to be used as a premise in reasoning, . . . [and] for [the] rational control of action and . . . speech" (1995; cf. 2001).
- Whatever the merits of that idea, it doesn't match any intuitive notion of a state's being conscious, since a state's simply being thus poised does not by itself result in one's being conscious of that state (Rosenthal 2002c).

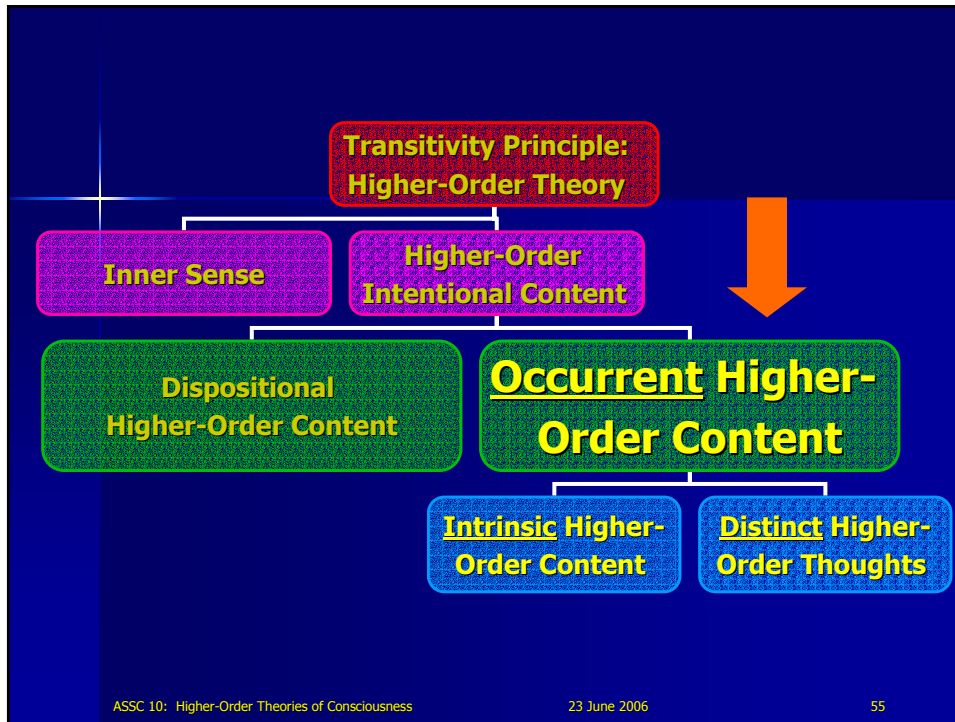
- Similarly for global-workspace theories (Baars 1988, 1997; Dehaene and Naccache 2001; Tononi 2004; van Gulick 2004), on which a state is conscious in virtue of its global ties to many cognitive states.
- But a state may have such ties without one's being at all conscious of that state, and so without implementing TP.
- Van Gulick adds that the "organization and intentional content" of the resulting "globally integrated complex . . . embodies a heightened degree of reflexive self-awareness" (2006).
- But this still needn't implement TP.

➤ Discussion? Questions?

➤ On, then, to Ic:

## Ic. Distinct Higher-Order States vs. Intrinsic Higher-Order Content

- Our goal to find how TP is implemented, i.e., how it is that we are aware of our conscious states.
- If it's not by way of higher-order sensing or perceiving, nor by way of dispositions to have higher-order thoughts (HOTs), it must be by way of some occurrent higher-order intentional content.



- Such higher-order intentional content would be to the effect that one is in the state in question, and that content would thereby result in one's being conscious of oneself as being in the relevant state.
- The occurrence of that content would implement TP.
- Moreover, the higher-order content would have to occur in connection with some assertoric mental attitude, since having a doubt about something or wondering about it, e.g., do not result in one's being conscious of that thing.



- But there are two ways occurrent higher-order intentional content might figure.
- It might be that a distinct intentional state occurs with the requisite content; but it might instead be that the higher-order intentional content is intrinsic to the conscious state itself.
- I'll call that second view intrinsicism. The classic articulation is Franz Brentano (1874), derived from Aristotle (e.g., Met Λ). Contemporary exponents include: Rocco Gennaro (1996), Uriah Kriegel (2005, 2006), Thomas Natsoulas (1999), and Amie Thomasson (2000); variant versions are put forth by Carruthers (2000, 2001), Van Gulick, and Levine (2006).

- As Kriegel observes (2005), intrinsicism is widely held in the phenomenological tradition.
- That's to be expected, since the main reason to hold intrinsicism stems from the phenomenological appearances: It doesn't seem, subjectively, as though conscious states involve two distinct states, one first-order and the other higher-order.
- But that's unsurprising: Since we're rarely aware of the higher-order content, we would seldom be aware of the distinct state that has that higher-order content.



- When we introspect, we are aware of both the introspected state and our being aware of that state.
- So it's reasonable to see introspection as that unusual case in which the higher-order state is itself a conscious state.
- Kriegel urges that we are always aware of the higher-order content, which he thinks is intrinsic to each conscious state.
- Introspection, he thinks, simply occurs when we shift attention to that higher-order content.

- But that's implausible.  
There are plenty of conscious states for which we plainly are in no way whatever conscious of any higher-order content—  
e.g., relatively peripheral perceptual sensations and passing thoughts.
- An Intrinsicist Argument:  
If we are aware of our conscious states by way of the higher-order content of a distinct state,  
that higher-order content might misrepresent the target first-order state.
- And some hold that consciousness cannot thus misrepresent one's mental life.

- Intrinsicists seek to preclude such misrepresentation by construing the higher-order content that implements TP as intrinsic to the target state.
- This need may seem especially pressing in connection with qualitative states: What would happen if one were in a red visual state, but had a distinct HOT that represented that visual state as green?
- So it may seem best just to rule such cases out, and intrinsicism may seem to be the best (or, indeed, only) way to do that.

- But (1) intrinsicism cannot help here; and (2) we have no sound reason even to rule such misrepresentation out.
- (1) Intrinsicism cannot help, because it gives us no guarantee that inaccurate higher-order intentional content cannot be intrinsic to a conscious state.
- For all intrinsicism tells us, higher-order content that one is in a green visual state could be intrinsic to a red visual state.
- If there is a way to rule to out that isn't ad hoc, intrinsicism doesn't provide it; and such a reason might in any case apply equally to distinct HOTs.

- But (2) a theory of consciousness need not rule out such misrepresentation.
- Such misrepresentation would be rare: First-order states have causal and functional ties with other mental states and with behavior.  
So, being inaccurately conscious of oneself as being a particular state would conflict with other aspects of one's mental life, both conscious and not.
- And these conflicts in causal connections will make such misrepresentation rare.
- That's altogether independent of whether the higher-order content is intrinsic.

- But these factors do not rule out occasional misrepresentation.
- Levine (2001, ch. 3) and Neander have urged that a red visual state with higher-order content of a green visual state cannot happen, since there would be no principled way to say what it would be like for one in such a case.
- But that's a mistake. On TP, what it's like for one is a matter of what states we're conscious of ourselves as being in.
- So what it would be like in such a case is a function of the higher-order content.

- This is a principled answer:  
*Consciousness is, after all, a matter of how one's mental life appears to one.*
- What it's like for one just a matter of *how one is conscious of oneself as being in various mental states*, qualitative states included.
- And that's again independent of whether the higher-order awareness is intrinsic or distinct: *What it's like for one* hinges simply and solely on *how one is aware of the state in question*—i.e., on how one's higher-order awareness, *whether distinct or intrinsic*, represents that state.

- The same points hold for a more radical kind of misrepresentation: *Higher-order content without any suitable target*—e.g., a higher-order content that one is in a state with the mental quality red, even though there simply isn't any such state.
- Again, there's no reason that isn't *ad hoc* to think that this can't happen on an intrinsicist theory.
- And again, *what it's like for one will be a matter simply of how the higher-order content represents things—what states one is conscious of oneself as being in.*

- The idea that we must rule out inaccurate higher-order awareness goes with the Cartesian view that, when it comes to the mind, there is no difference between appearance and reality.
- But that too is a mistake. The reality of consciousness is the way our mental lives appear to us; with consciousness, appearance and reality do coincide.
- And that, together with the Cartesian doctrine that all mental states are conscious, may seem to imply that mental reality is also exhausted by mental appearance.

- But mental appearance does not exhaust its reality:
- (1) Not all mental reality is conscious.
- And (2) mental reality involves the many ties each mental state has with others, and with stimuli and behavior.
- It's in any case arguably a good thing to hold open the possibility of inaccurate higher-order awareness.
- That possibility is testable:  
Some qualitative occurrence might prime for red qualitative states, even though the subject reports and is conscious of that qualitative occurrence as green.

- There is in any case an unavoidable difficulty in supposing that the higher-order content that implements TP is intrinsic to the relevant states.
- This difficulty has to do with the individuation of mental states.
- Intrinsicist claims that the higher-order awareness that implements TP is intrinsic to the state one is aware of; the anti-intrinsicist insists that this awareness inheres in a distinct state.
- But both claims are idle without some way to individuate mental states.
- Intrinsicists try to do this with content.

- Kriegel (2005), e.g., suggests that a conscious sensation of a red circle might have this intentional content:  
This particular is red and circular and represented (hereby) to be red and circular.
- The 'hereby' is to ensure that a single state represents something to be red and circular and represents itself as doing so.
- But Kriegel gives no reason to think that it is those contents which actually implement TP—that it's those higher-order contents in virtue of which we are aware of our conscious states.



- In any case, the individuating of mental states hinges not on their representational content, but on their mental attitude.
- A single state can have multiple pieces of representational content: qualitative character and intentional content, and indeed several of each.
- But no single state has two distinct mental attitudes.
- No single state can be a wondering and a doubting, or a believing and a wondering, or an expecting and a recalling, . . . .
- So mental attitudes individuate (intentional) mental states.

- Moreover, the higher-order awareness that implements TP always has an assertoric mental attitude:
- If I doubt that I'm in a particular state or wonder whether I am or expect myself to be, that won't make me conscious of myself as being in that state; so it won't implement TP.
- So when I consciously doubt, wonder, or expect something, there is the doubt, wonder, or expectation that I'm conscious of, and there is also my assertoric higher-order awareness of that state.



- Since the state I'm aware of is not assertoric and the higher-order awareness is, they must be distinct states.
- And that holds whether the higher-order awareness is purely intentional, which intrinsicalists usually say, or perceptual (as Brentano may have held).
- Kriegel has urged that a single state can have both "directions of fit" (2003; see Searle 1983). But mental attitudes are not just direction of fit, and even if a single state can have both directions of fit, it cannot have two distinct mental attitudes.

- We've ruled out several kinds of higher-order theory:
  - Inner sense—since there are no higher-order mental qualities;
  - Dispositional theories—since being disposed to be conscious of something doesn't implement TP; and
  - Intrinsic higher-order content—since mental states are individuated by their mental attitude.
- A theory based on distinct, occurrent HOTs is the only way left to implement TP (Rosenthal 2005, 1997/1990, 2002b).

➤ IIa and IIb look at two challenges to that theory:

IIa: Can distinct, occurrent HOTs do justice to qualitative consciousness?

IIb: Can such a theory explain the function of state consciousness, and, in particular, why distinct, occurrent HOTs arise at all and are usually so accurate?

➤ IIc, finally, considers two advantages of the theory: It explains the Libet-Haggard timing results and it explains the various ties consciousness has with speech.

➤ Discussion? Questions?

➤ On, then, to IIa:

## IIa. Qualitative Consciousness

- One apparent advantage of inner sense is that it seems that it could account for qualitative consciousness.
- Mental qualities differ in ways that we seem *unable to capture conceptually*.
- Indeed, the very nature of mental qualities may seem *ineffable*—may seem to *defy informative description*.

- These factors suggest that we're not conscious of our qualitative states by having purely conceptual states about them.
- Putting aside so-called nonconceptual content, the only alternative is that we're conscious of them by some inner sense.
- But inner sense would have an advantage over purely conceptual awareness only if inner sense itself uses *mental qualities* to represent its mental targets.
- And, since *there are no higher-order mental qualities*, inner sense fails.
- Can HOTs do the job?

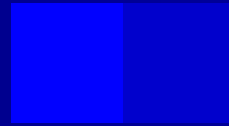
- The apparent ineffability of and mystery about qualitative character stem from denying that any informative account of conscious mental qualities is possible.
- We have to settle on that view for a kind of "Louis Armstrong" attitude (Block 1978) toward describing the nature of qualitative character: "If you gotta ask, you ain't never gonna get to know."
- This leads to criticism of higher-order theories, and indeed of TP:  
If qualitative character is ineffable and mysterious, *no way of being conscious of it could do justice to it.*

- So, for a higher-order theory to succeed with qualitative consciousness, *it must rely on a satisfactory theory of qualitative character itself.*
- This has not been widely recognized by higher-order theorists.
- Some higher-order theorists, such as Lycan (2004) and Carruthers (2000, 2001), *have tried to accommodate the apparent ineffability of qualitative character* by appealing to Brian Loar's (1990/1997) suggestive notion of *purely recognitional concepts.*

- A concept is purely recognitional if it applies not by way of any ties with other concepts, but solely in virtue of one's ability to recognize what type of thing it is.
- So perhaps we're conscious of our qualitative states by having higher-order thoughts (Carruthers) or perceptions (Lycan) that pick out qualitative states by way of purely recognitional concepts.
- We could thereby have the virtues of a higher-order theory and yet preserve the ineffable, mysterious nature of qualitative character.

- But it's unlikely that any concepts are purely recognitional in this way.
- We may sometimes recognize something, as Loar urges (e.g., a kind of cactus), without knowing what its kind is.
- But we can nonetheless always say something about it.
- Some recognitional abilities, such as face recognition, rely on dedicated cortical machinery that operates independently of the way we conceptualize the things we recognize.
- But we have no reason to think that such recognition involves concepts at all.

- Even if there are purely recognitional concepts that sometimes figure in our recognizing of things, it's *doubtful that they figure in the way we are aware of our qualitative states.*
- When two shades of blue, say, are very close, subjects are a lot *more reliable distinguishing them than they are at accurately identifying or recognizing* the two shades (e.g., Hurvich 1981; Halsey and Chapanis 1951, Burns and Ward 1982 [references all from Raffman 1995]).
- Similarly for pitch (Seashore, *The Psychology of Music*) and other perceptible qualities.



- *A word on this finding:* Diana Raffman (1995) has invoked this finding to undermine representational (or intentionalist) theories of mental qualities (Dennett 1991, Harman 1990, Armstrong, Lycan, and in modified form Shoemaker 1994 and Tye 1997) — on which all mental properties are intentional and there are no mental qualities, properly so called.
- *On this view, the only mental properties perceptual states have are intentional properties,* which represent properties of the things being (or seeming to be) perceived.
- I won't talk about representationalism.

- And Thomas Metzinger has argued that this finding—which he calls *the memory constraint*—shows that differences in qualitative experiences are “so subtle and fine-grained” that they aren’t “available for memory nor for cognitive access in general” (2003: 70).
- If so, no higher-order theory that invokes cognitive (intentional) higher-order states would work, since *the subtle, fine-grained differences among qualitative states could not be captured by any differences in cognitive (intentional) content.*

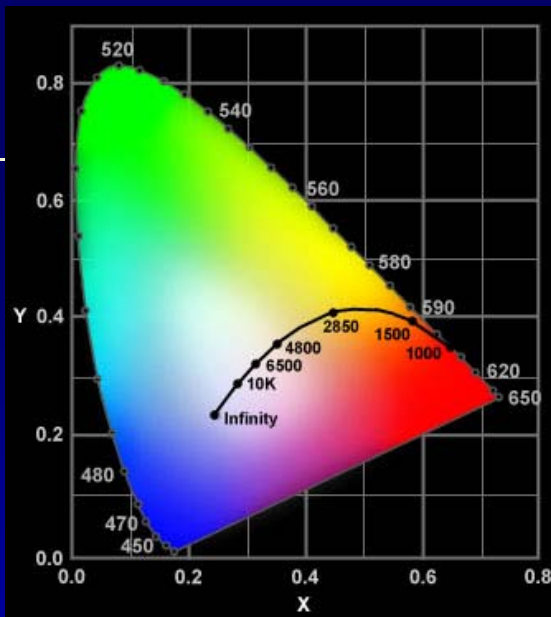
- I’ll argue in a moment that Metzinger’s conclusion is rash, and that *this finding does not conflict with a cognitive higher-order theory, such as the HOT theory.*
- (But my argument will not also protect representationalism against that finding.)
- But it’s first worth noting that these findings *cast doubt on our being aware of mental qualities by way of purely recognitional concepts.*
- A concept, again, is purely recognitional if it *applies solely by way of recognitional abilities, and wholly independently of ties to other concepts.*





- But if we're aware of occurrent qualitative states by way of concepts that rely solely on recognitional abilities, why would we be better at distinguishing presented qualities than we are at recognizing them?
- Whatever the impact of this finding on representationism or on higher-order theories generally, it undermines the appeal to recognitional concepts in explaining how we're aware of concurrent qualitative properties.

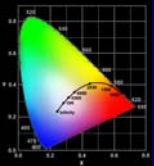
- What would explain this finding? And can we explain it in a way that fits with a higher-order theory of qualitative consciousness? In particular, can we explain it in a way that undermines the seeming ineffability of qualitative character—i.e., in a way that allows us to give an informative description of mental qualities?
- It's best in addressing these questions to start with an informative account of mental qualities that will fit with higher-order theories, and see whether it fits with the finding we've been focusing on.

- We can begin by reflecting on how we respond to a request to describe mental qualities that differ only slightly—say, two very similar shades of yellow.
- There are two standard things we say:
  - (1) We might say that one is the yellow of bananas and another is that of lemons; we describe shades by reference to the color of standard physical objects.
  - (2) We also rank shades in terms of brightness, saturation, and hue (e.g., one is closer to green than another); we in effect locate shades in a quality space of colors:



The CIE (Commision Internationale d'Éclairage) Chart represents hue and saturation in 2 dimensions, with fully saturated hues on the outside.

- We can combine the two kinds of informative description we give of slightly differing shades in a way that points to a theory of mental qualities.
- Humans can discriminate shades of color exhibited by physical objects in very fine-grained ways. Using just-noticeable differences (or, perhaps better, matching [Goodman 1966]), we can construct a quality space that captures the physical shades humans can discriminate.
- These are colors of physical objects—such as bananas:  and lemons: 

- There must be distinct visual states for each shade of physical color that a subject can discriminate, states that resemble and differ in ways that correspond to the discriminable differences in physical color.
- Mental qualities are the properties visual states have in virtue of which one can discriminate physical colors.
- So, if the space of physical colors is:  
 the very same quality space will also map the similarities and differences that define the mental color qualities.
- Similarly with other families of qualities.

- The mental qualities of each modality are defined by their location within a quality space that is fixed, in turn, by the ability to discriminate the physical perceptible properties accessible by that modality.
- I.e., the mental qualities are defined by a pattern of similarities and differences that reflects the similarities and differences found in those perceptual abilities.
- This reflects our tendency to describe mental qualities both in terms of such similarities and differences and by reference to the perceptible properties of various physical objects.

- This way of describing mental qualities makes no reference to what it's like for the creature in question.
- Nonconscious discrimination would suffice for constructing the quality space in terms of which the foregoing account defines mental qualities.
- Typically as discrimination becomes harder, confidence reduces faster than performance.
- So forced-choice tasks should distinguish shades well above chance even when subjects say they look the same.
- And that's nonconscious discrimination.

- So we can define mental qualities by a quality space built up from discriminative abilities independent of what it's like for the creature.
- That allows for mental qualities to occur both consciously and not— which squares with results in priming and other forms of subliminal perceiving.
- It also allows for informative description of mental qualities—belying the idea that they are ineffable and mysterious—and in ways that fit with commonsense talk.
- And, important for our purposes, it fits well with HOTs.

- Qualitative states, on the foregoing account (Rosenthal 2005, ch. 5-7, 2001, 1999), are not in themselves conscious.
- According to TP, a qualitative state, like any other mental state, is conscious just in case one is conscious of oneself as being in that state.
- The foregoing account accommodates the HOT theory because it specifies what the content is of the HOT in virtue of which each particular qualitative state would be conscious.

- *A visual state has a mental quality of red, e.g., in virtue of its place in the quality space of mental color qualities—*  
i.e., in virtue of the similarities and differences that states has to others that exhibit mental color qualities.
- So *the HOT in virtue of which that state would be conscious will describe the state as having a mental quality with the relevant location in that quality space—*  
i.e., a mental quality with the relevant pattern of similarities and differences from other mental color qualities.

- That helps explain why our sensations of color are *usually conscious only in respect of relatively coarse-grained mental qualities.*
- The HOTs in virtue of which those states are conscious usually locate the relevant mental qualities relatively coarsely within the relevant quality space.
- But focusing on a color experience can lead to one's being conscious of it in respect of far more fine-grained similarities and differences—  
i.e., a more fine-grained location in the relevant quality space.



- Having a variety of concurrent color experiences, moreover, helps refine the location of any particular color experience within the space of mental color qualities.
- That's why *we're conscious of shades in a far more fine-grained way when we actually see them than when we're called on to recognize or remember them.*
- Pace Metzinger, that result not only fails to undermine the HOT hypothesis; *the hypothesis actually helps explain it.*
- (Since the explanation trades on how we're conscious of *mental qualities*, it may not help with representationalism.)

- Some find the HOT theory unconvincing when it comes to qualitative consciousness.
- On that theory, qualitative states unaccompanied by HOTs are not conscious; there is nothing it's like for one to be in those states.
- But how can *purely intentional states, such as HOTs, result in there being something it's like for one to be in a particular qualitative state?*



- As Van Gulick puts the worry, the theory doesn't locate the "what it's like" in the qualitative state itself, and it seems that the "what it's like" can't be an aspect of a purely intentional state.
- So the "what it's like"—the qualia, as he refers to them—are "stranded" (2004).
- But there is reason to think that *intentional states can, after all, make the difference between there being something it's like for one to be in a particular qualitative state and there not being anything it's like for one.*

- Consider somebody for whom the conscious experience of tasting any red wine is the same as that of tasting any other, or somebody for whom the conscious experience of hearing a clarinet is the same as that of hearing an oboe—they're all indeterminate woodwinds.
- *Sometimes one learns consciously to distinguish these experiences by learning new words on which to hang the previously indistinguishable conscious experiences.*

- How could learning new words help?
- Learning a new word is learning the concept that the word expresses—  
i.e., learning to have thoughts that involve that concept.
- In this case, the new concept is that of the experience in question—  
in the more fine-grained way in which one comes to be conscious of it.
- The new concept is the very concept needed in a HOT that describes the experience in the more refined way in which one comes to be conscious of it.

- In this kind of case at least, the very ability to have a HOT with the more refined content does make the difference between a more and a less fine-grained "what it's like."
- So it should be no surprise if having a HOT to the effect that one is in a particular type of qualitative state actually results in there being something it's like for one to be in that state.
- The HOT makes one conscious of oneself as being in that state, and what it's like for one is a matter of what states one is conscious of oneself as being in.

➤ Discussion? Questions?

➤ On, then, to IIb:

## IIb. The Function of Consciousness, and Why Higher-Order States Occur

- As noted in Part Ib, Carruthers (2004) challenges the positing of occurrent HOTs on adaptive grounds: What adaptive value or function would occurrent HOTs have? What role would they play?
- Some things evolve without any particular adaptive value: Spandrels.
- But in any case, HOTs don't reduplicate first-order content.

- The content of HOTs has to do with what states one takes oneself to be in. So that content does not replicate the content of the state itself.
- Saying or thinking, e.g., that it's raining is different from saying or thinking that one thinks it is.
- Note Moore's paradox: It's contradictory to say 'It's raining and it's not raining' or 'I think it's raining and I don't think it is', but it's not contradictory (though it is absurd in some other way) to say 'It's raining but I don't think it is'.

- Is there adaptive value to one's having thoughts about what states one is in?
- As noted in Part Ib, it's sometimes argued that there is:  
that such HOTs enhance one's ability to reason and plan rationally—presumably because such HOTs allow us to survey our thinking and correct ourselves.
- Jeffrey Gray (2004, chs. 7-8) has argued that conscious qualitative states allow for late error detection, though it's unclear why nonconscious qualitative states wouldn't work as well.

- Edmund Rolls has developed an especially sophisticated argument for this view, which relies on the need for neural nets to have higher-order monitoring in correcting multi-step plans (Rolls 2004; Rolls 2005, chs. 10, 3).
- Higher-order states, on Rolls's view, arise because of the need for such correcting, and their occurrence results in our being conscious of ourselves as being in the target first-order states.
- But there is reason to doubt that picture.

- Typically (always?) when we correct our reasoning or planning, it's because of the (causal) interaction of first-order intentional states, not some 2nd-order surveying of those first-order states.
- Multi-step plans then likely reduce to single-steps, and corrected by eliminating conflicts solely among first-order states.
- We're conscious of our mental states as enhancing rationality or error detection and correction; but our being conscious of states as playing that role doesn't show that they actually do so. (Cp. Libet's results, Part IIc.)

- Each intentional state has causal ties to many others; indeed, on some theories (functionalist theories), it's in virtue of such connections that each state has the intentional content it has.
- When we think two things that conflict, each tends to cause a thought that contradicts the other and that, by itself, causes us to question each. Typically the stronger thought—that which has the more robust causal ties to most other thoughts—wins out.
- *The functional roles of the various first-order states suffice for rationality.*

- Higher-order monitoring seldom figures in this process.
- And when it does, it's typically *awkward and slow*; note how "writing it down" often helps if we want to induce rationality on our thinking by an actual process of monitoring our thinking.
- So it's unlikely that this rationalist picture reflects selection pressures that could have led—  
evolutionary accident aside—to our coming to have HOTs about many of our ordinary mental states.

- But there are, in any case, other ways to explain why HOTs arise— and also why they typically are fairly accurate.
- Two different stories are needed: one for HOTs that describe our first-order states in intentional terms and another for HOTs that describe those states qualitatively.
- First the qualitative.
- Normal perceiving involves intentional content that's directed solely at the thing one perceives; there's no need for any thought directed at the perceptual state.

- But perceptual errors, e.g., about whether something is red, may make one register cognitively the disparity between the thing misperceived and the state in virtue of which one misperceives it.
- Use 'red\*' for the mental quality of red. Misperceiving things as being red will lead to one's having the concept of a red\* state—a state one typically is in when something red is in front of one.
- Perceiving something as red involves not only the mental quality, red\*, but in addition the intentional content that, roughly, something red is in front of one.



- One's concept of a red\* state is the concept of a state one is typically in when something red is in front of one.
- So, in a creature with the concept of a red\* state,  
*the perceptual thought that something red is in front of one will to some degree dispose that creature also to have the thought that it is in a red\* state.*
- *Nonperceptual thoughts* that something red is in front of one won't have that effect,  
since they don't occur in connection with red\* qualitative character.

- Qualitative states are states that respond to corresponding perceptible properties.
- So, in a creatures with (minimal) concepts of various qualitative states,  
*simply perceiving things will come to facilitate the occurrence of HOTs about the qualitative character of its perceptual states—HOTs that are relatively accurate.*
- Note: One virtue of a HOT theory is its *flexibility in capturing the different ways one can be conscious of a state.*
- E.g., the mental qualities in virtue of which a qualitative state is conscious can be more or less finely taxonomized.

- But also one could be conscious of a perceptual state—which involves both qualitative character and intentional content—solely in respect of its qualitative character, and not at all in respect of its intentional content.
- That's what might happen with many creatures whose perceptual errors had led to their having the concept of various qualitative states.
- Its HOTs would represent their targets solely in respect of their qualitative properties, and not in respect of their intentional content.

- Things are somewhat more complicated with HOTs that describe their targets in respect of their intentional content.
- Perceptual errors suffice for one to come to have a minimal concept of a qualitative state—a state one is in when the relevant kind of perceptible object is there.
- But it takes more for a creature to come to have the concept of a state with intentional content.
- Very likely, a creature will come to have that concept only after it has come to have the ability to speak, and to describe its speech behavior (Wilfrid Sellars 1956).

- We can think of thoughts—states with intentional content—as states that often issue in corresponding speech acts, as well as associated nonverbal behavior.
- So consider a creature with the ability to speak and to describe its speech behavior—and who comes thereby to ascribe thoughts to itself and others in a strictly third-person way.
- When this creature (our ancestors? [Sellars]) ascribes thoughts to itself in a strictly third-person way, it has an inferential (higher-order) thought that it has the relevant (first-order) thought.

- The greater that creature's fluency in self-ascribing thoughts on a third-person basis, the more disposed it will be to have the relevant (usually accurate) HOT whenever it has a first-order thought.
- And the more disposed to have such HOTs that creature becomes, the more likely those HOTs will occur independently of any conscious inference.
- Its first-order intentional states will then facilitate the occurrence of seemingly noninferential HOTs; it will be conscious of itself, in a seemingly direct way, as being in states with intentional content.

- The facilitation of HOTs about states in respect of their intentional content rests on an automatic performance-conditional equivalence of saying and thinking, e.g., 'It's raining' and 'I think that it's raining'.
- The performance-conditional equivalence one level up, between 'I think that it's raining' and 'I think that I think that it's raining', is awkward and far from automatic.
- So the 3rd-order thoughts that occur in introspective consciousness are themselves relatively rare, and involve deliberate mental effort.

- The two facilitation stories suggest that nonlinguistic creatures, though they may often be conscious of themselves as being in states with qualitative character, are very likely never conscious of themselves as being in states with intentional content.
- That accords with pretheoretic intuition: We feel pretty confident that the qualitative states of nonlinguistic creatures are conscious, but it seems a lot less clear that the same is true about their intentional states.

- A HOT theory offers, then, a way of explaining why, when, and in what way the mental states of various creatures are conscious,  
and why they are conscious in reasonably accurate ways.
- And that explanation does not require any implausible assumptions about the function of mental states' being conscious, e.g., assumptions about the role of consciousness in rationality and self-correction.
- (On this, see Rosenthal 2005, ch 7, §6 and ch 10, §5.)

➤ Discussion? Questions?

➤ On, then, to IIC:

## IIC. Speech and Consciousness, and Timing

- Consciousness has two connections with speech, which it's crucial to distinguish.
- If I consciously think that it's raining, I can report my thought by saying 'I think it's raining'.
- But I can also verbally express that same thought, by saying simply 'It's raining'.
- First a word about the tie consciousness has with reporting; then the more complex tie with expressing.

- In arguing for TP in Part Ia, the tie with reportability came up.
- Conscious states are those we can report being in.
- Assume the HOT hypothesis:  
That a state is conscious just in case one has a distinct, occurrent, assertoric HOT that one is in the state in question.
- We can express our assertoric thoughts in speech—by saying something that has the very same content and a corresponding illocutionary force.
- A report that one is in some state is also an expression of a HOT about that state.

- Let's turn to the tie with expressing. Consider the remarks, 'It's raining' and 'I think it's raining'.
- As noted in Part IIb, these have the same conditions of assertibility—the same performance conditions.
- Whenever it's appropriate to say one, it's appropriate to say the other.
- Moreover, this performance-conditional equivalence is wholly automatic and second-nature for us.
- When we say one of these things, we may well not recall, even a moment later, which of the two we said.

- But the two mean different things—they have different truth conditions.
- The truth conditions of 'It's raining' have to do with the weather, those of 'I think it's raining' have to do with the contents of one's mind.
- These things figure in explaining a striking, otherwise unexplained thing about human thought and speech:  
**Whenever one expresses a thought in speech, that thought is conscious.**
- Thoughts expressed nonverbally, by contrast, need not be conscious.



- This correlation may well underlie the tendency of many thinkers since Descartes to assume a special tie between consciousness and speech.
- Only creatures with the capacity to talk, Descartes thought, have conscious thoughts (AT VI 58-9, IV 573-576, V 275-279).
- And that may well be so, given the considerations in Part IIb about what's required for the facilitation of HOTs about intentional states.
- But Descartes also held that all mental states are thoughts and also that they're all conscious states.

- And that led him to conclude that nonlinguistic creatures have no thoughts –indeed, no mental states at all.
- Still, the more limited conclusion may well be right:  
that the mental states of nonlinguistic creatures are never conscious in respect of their intentional properties, but only in respect of their qualitative character.
- The foregoing considerations about the facilitation of HOTs allow an informative, satisfactory explanation of why the thoughts of nonlinguistic creatures may well fail to be conscious as such.

- HOTs also allow an explanation of the regularity that verbally expressed thoughts are always conscious.
- Whenever one says 'It's raining', one expresses one's thought that it's raining.
- But recall the automatic, second-nature performance-conditional equivalence of saying 'It's raining' with saying 'I think it's raining'.
- Given that, whenever one says 'It's raining', one might as easily have said 'I think it's raining'.
- So, whenever one says 'It's raining', one is disposed to say 'I think it's raining'.

- But one can be disposed to say 'I think it's raining' only if one has the thought that that speech act would express.
- And the thought that speech act would express is a HOT:  
The HOT in virtue of which one's thought that it's raining is conscious.
- So: Because of the automatic character of the performance-conditional equivalence between the two speech acts, simply saying 'It's raining' actually disposes one to have a HOT that one has the thought that it's raining.

- When one expresses one's thoughts nonverbally, there's no performance-conditional equivalence between that nonverbal expression and a speech act that would express a suitable HOT.
- So nonverbal expressing of thoughts can occur without the thoughts' being conscious.
- But how about speech acts that express HOTs, themselves? When one says 'I think it's raining', the only thought one is normally conscious of is the first-order thought that it's raining, not any HOT.
- Why this exception?

- HOT theory again provides an informative explanation.
- When one says 'I think it's raining', that's performance conditionally equivalent to saying:  
'I think that I think that it's raining'.
- But here that performance-conditional equivalence is not at all automatic. So making a second remark does not dispose one to make the 3rd-order remark,  
and so not to have a 3rd-order thought that would make one's 2nd-order thought conscious. (All this in Rosenthal 2005, ch. 10.)

- Does all verbal expressing of one's thoughts result in their being conscious?
- If not, must some additional condition be met—perhaps that *the expressing involves some intention to communicate?*
- That's doubtful: Thoughts expressed in *soliloquy* are conscious, though they're expressed with no communicative intent.
- Verbal expressing, by itself, arguably results in the expressed thought's being conscious *whenever the expression is performance conditionally equivalent to a higher-order remark—and that equivalence is second nature.*

- Also, communicative intent can, like other intentions, occur nonconsciously. And when it does, *the state one nonconsciously intends to communicate need not itself be conscious.*
- It might seem that if one does intend to communicate one's thought, that thought must be conscious.
- *Perhaps intending to communicate one's thought must involve some higher-order awareness of that thought.*
- Maybe. But even if so, there is a crucial *ambiguity* in speaking of communicative intent, which makes a difference here.

- If one intends to communicate that one has a particular thought, perhaps one must be conscious that one has that thought—though even that's not obvious.
- But when one intends to communicate one's thought, one typically intends only to communicate the proposition that one thinks:  
One intends simply to communicate, say, that it's raining—  
not that one has an intentional state that has that content.
- And then one needn't be in any way conscious of the intentional state itself.

- Striking experimental work by Benjamin Libet (1985; Libet, Greason, and Wright 1983), replicated and refined by Patrick Haggard (1999; Haggard and Eimer 1999; Haggard, Newman, and Magno 1999), has occasioned much interest among consciousness researchers.
- This work seems to show that, when we consciously decide to do something, the neural event that initiates the action occurs prior to that conscious volition.
- And this result seems to conflict with our commonsense idea that volitions cause voluntary actions.

- Libet and others have argued that a second finding mitigates that conflict.
- Although the actions subjects consciously decide on are neurally initiated prior to their conscious decisions, they retain some ability consciously to "veto" an action—after the neural initiation but before the action.
- But since the action is initiated before the conscious decision, that doesn't help if the question is about what role conscious decisions, and hence consciousness generally, have in initiating action.

- I won't say much about Libet's other result, that even though subjects' conscious experiences of somatosensory stimulation can occur as much as 500 ms later than the actual stimulation, subjects experience them as occurring earlier—within 10-20 ms of the actual stimulus.
- And I won't say anything about Haggard's argument (Haggard and Eimer 1999) that the readiness potential Libet isolated corresponds only to a generalized preparation to act, and that a subsequent Lateralized Readiness Potential (LRP) initiates the specific action, since that LRP also occurs well before any conscious decision.



- What matters for higher-order theories is that the phrase 'conscious decision' conceals an ambiguity:  
It could be a subject's decision together with the consciousness of that decision.  
But it could instead be the decision itself, which might well first occur without being conscious, and only subsequently come to be conscious.
- And, whatever the readiness potential corresponds to, subjects' reports don't reveal their decisions,  
but only their consciousness of their decisions.

- So the experimental findings show that action in these cases is initiated not by conscious decisions, but rather by decisions that aren't (yet) conscious.
- But a conflict with common sense remains. We tend to think of decisions—and other mental states—as having consciousness built into them (hence the apparent appeal of intrinsicism).
- So, if consciousness of a decision occurs later, we don't consider a possible earlier decision that isn't conscious.
- But the earlier decision may be the same decision—just not yet conscious.



- Higher-order theories—on which a state's being conscious is a distinct matter from that state's simply occurring—invite this distinction between a decision of which one may at first fail to be conscious and the later consciousness of that decision.
- (Intrinsicalism and dispositional theories obscure that distinction somewhat, and also face independent difficulties.)
- So Libet's and Haggard's timing results fit neatly with (some) higher-order theories, and so provide some confirmation of those theories. (Rosenthal 2002a.)

- The Libet-Haggard experiments don't apply to the multi-step reasoning that Rolls focuses on in connection with the function of states' being conscious.
- So it would be good to design experiments that would do so.
- Of special interest would be whether the consciousness of the states that figure in inferential processes also occur later than those states and the subsequent states that follow inferentially.
- If so, consciousness would have little effect on rational thinking.

- The more pressing conflict with common sense has to do with our sense of free agency.
- This sense of freedom has to do with our conscious volitions; nonconscious volition carries no intuitive sense of freedom.
- So, if decisions initiate actions before becoming conscious, what happens to that sense of free agency?
- Our sense of freedom in connection with conscious volitions arguably stems simply from our failure to be conscious of the mental antecedents of those conscious volitions.

- Still, it's not credible that there are no mental antecedents of our conscious decisions and volitions; they're caused by previous mental occurrences just as much as are our nonconscious volitions.
- But we're struck by our failure to find in consciousness any mental occurrences that cause our conscious volitions, and so we see those conscious volitions as free.
- That doesn't happen with nonconscious volitions, since we don't find them in consciousness at all, and hence aren't struck by our not being conscious of their mental causes.

- Distinguishing between mental occurrences and our consciousness of those occurrences helps in this way to explain why our sense of free agency seems bound up with volitions only insofar as they are conscious. (Rosenthal 2005, ch. 13, §7, 2002a, §4.)
- This is another benefit of higher-order theories—at least those which draw a clean distinction between mental states and our consciousness of them.

**THE END**