

Consciousness and Intrinsic Higher-Order Content

© **David Rosenthal**
City University of New York
Graduate Center
<http://web.gc.cuny.edu/cogsci>
Tucson VII, April 7, 2006



OVERVIEW

(Some slides were skipped in Tucson)

- I. Higher-Order Content and the Transitivity Principle**
- II. Two Kinds of Intrinsicism**
- III. Arguments from the Phenomenological Appearances**
- IV. Troubles for Intrinsicism**

I. Higher-Order Content and the Transitivity Principle

- Any theory of consciousness must explain how conscious mental states differ from mental states that aren't conscious (or else it must argue that all mental states are conscious).
- I'll assume that the second option is out: that mental states of every type can occur both consciously and not.

- An obvious, and widely accepted, proposal for explaining how conscious states differ from mental states that aren't conscious is that a mental state is conscious only if the subject is conscious (or, equivalently, aware) of that state.
- I'll call this the Transitivity Principle (TP).
- I'll say something about the main arguments against TP, and I'll rehearse one compelling argument in support of TP.

- A widely adopted argument against TP (and so against higher-order theories generally):
If, as on TP, a mental state's being conscious consists in one's being aware of it,
why isn't a coffee cup conscious when I'm conscious of the cup? (Due first to Dretske.)
- Kriegel has urged that intrinsicism is needed to defeat this objection, since then the higher-order awareness is intrinsic to the conscious state, whereas the awareness of a coffee cup is not intrinsic to the cup.

- But it's being intrinsic wouldn't make the relevant difference here.
- Why should an awareness of a state make it conscious if it's intrinsic to that state, but not make a state (or a coffee cup) conscious if it isn't intrinsic to it?
- What's wrong with this challenge to TP is rather that it conflates two distinct kinds of consciousness.
- The consciousness that some mental states exhibit is not the same property as the consciousness that some objects (such as people) exhibit.

- A mental state's being conscious consists in its being a state that one is conscious of oneself as being in, since a state's being conscious is a matter of how one's mental life appears to one.
- But this has no application to coffee cups.
- A coffee cup's being conscious would be, if anything, something like a person's being conscious, not like a mental state's being conscious.
- And people are conscious (and coffee cups in cartoons) if they're awake and responsive to sensory stimulation.

- Some see a related difficulty, that a non-conscious higher-order state couldn't be a source of consciousness—that it couldn't confer consciousness on another state.
- Intrinsicism seems to avoid this problem, since any state with higher-order content is itself automatically conscious.
- But there's no problem about conferring, since a state's being conscious is not a property passed from one state to another, but a matter of how one's mental life appears to one: a matter of one's being conscious of oneself as being in that state.

- And even if the conferring problem did hold up, intrinsicism by itself couldn't help deal with it.
- A state that included its own higher-order awareness need not be conscious *in respect of* that higher-order awareness (pace Kriegel; more on this in §III); it could be *in respect only of those other mental properties that the higher-order awareness makes one conscious of*.
- So the problem would remain of how an intrinsic, but *nonconscious*, higher-order awareness can confer consciousness *on the rest of the state*.

- Two other reasons to reject TP:
- One (due to Searle): We can't *observe* our conscious states.
- True; but observing isn't the only way for one to be conscious of something.
- Another reason (Dretske): We're not always aware of the *differences among* our conscious states.
- But one needn't be aware of conscious states *in respect of* all their aspects: Each mental state is conscious *in respect of those mental features one is conscious of*.

- An argument *in support of TP*: Conscious states are those we can report being in—*in some intuitively immediate way* (more on that in a moment).
- Moreover, reports express thoughts—in this case, *thoughts that one is in the state in question*—which make one aware of the target.
- 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 thought.

- There are issues about the *experimental implementation* of TP (see, e.g., Lau, Merikle), which I won't address here.
- My focus in what follows will instead be on how to implement TP *theoretically*—and, indeed, only on *one aspect* of that question.
- Differences about how a theory of consciousness can best implement TP have led to *the variety of "higher-order" theories of consciousness* developed in the recent literature.

- Some theorists, e.g., have proposed that we are aware of our conscious states by perceiving them (or something like that).
- Others, myself included, have posited higher-order thoughts (HOTs) as the best way to implement TP.
- These are the thoughts that we express when we report our conscious states, and that we could express whenever a state is conscious.

- But the issue that matters for us this morning is which of these is true:
 - (1) The higher-order awareness of a conscious state that implements TP is distinct from that state.
 - (2) That higher-order awareness is intrinsic to the state that's conscious—it is a part or aspect of the state.
- On (1), one is aware of a conscious state by being in a 2nd, distinct state whose content makes one aware of the 1st state.
- We can call that content "higher-order," since it's about another mental state.

- (2), by contrast, holds that this higher-order content is actually a *part* or *aspect* of the state one is conscious of.
- On that view, every conscious state not only has *the various mental properties one is conscious of it as having*, but also has a *higher-order content* in virtue of which one is conscious of that state.
- I'll call this second view *intrinsicism*.
- This family of theories traces through the Austrian philosopher Franz Brentano (1874) all the way back to Aristotle's *de Anima*.

- I'll argue for *the first theoretical option*—that the higher-order content that implements TP *belongs to states that are distinct from the states one is aware of*.
- In my next section (§II) I'll describe two versions of intrinsicism—and I'll then argue (§III) against various considerations advanced in support of it, many of which appeal, in a certain way, to the *phenomenological appearances*.
- Finally (§IV) I'll raise some *independent difficulties for intrinsicism*, which point toward implementing TP by appeal instead to *distinct* higher-order states.

II. Two Kinds of Intrinsicism

- A conscious state is a state one is aware of by way of the occurrence of some higher-order content about that state.
- Why think that this higher-order content is intrinsic to the state one is conscious of?
- One reason is to underwrite, and explain, the close connection that higher-order content seems to have to the target state.

- Van Gulick writes that a state becomes conscious by being "recruit[ed] . . . into a globally integrated complex whose organization and intentional content embodies a heightened degree of reflexive self-awareness" (his emphasis).
- In being thus recruited, the state becomes a part of a larger state that "embodies . . . [such] reflexive self-awareness."
- And Kriegel holds that the awareness of one's conscious states must be immediate, and he explains this immediacy on the hypothesis that the awareness is a part or aspect of the state that one is aware of.

- Kriegel's version of intrinsicism holds that each conscious state has, as a part or an aspect, a higher-order content that refers to that state.
- Van Gulick's version proposes that states are conscious in virtue of belonging to a global process that in some way embodies reflexive self-consciousness.
- More generally, intrinsicism holds that a state comes to be conscious only if the subject's awareness of that state bears some suitably intimate tie to that state—a tie that requires that the state and the awareness of it be a single occurrence.

- Thus Kriegel urges that, since consciousness seems to be a natural kind, it exhibits "an underlying unity."
- But, the unified phenomenological appearance of consciousness may not be a reliable guide to its underlying reality. (More on such appearances in a moment.)
- And even our pretheoretic sense that consciousness is a natural kind may not be reliable:
We tend pretheoretically to see all commonsense, macroscopic phenomena as natural kinds until theory tells us otherwise.

- Van Gulick's intrinsicism is of a piece with his appealing view, which he calls HOGS (higher-order global states).
- But that theory—like Baars's related global-workspace theory, Tononi's information-integration theory, and Dehaene and Naccache's global-neuronal-workspace framework—doesn't implement TP: What is it in the global process into which a state is recruited in virtue of which one is aware of that state?
- Also, many states have global connections without being conscious—e.g., nonconscious thoughts, feelings, and desires.

- And many conscious states presumably lack any significant global connections—global connections that are distinctively mental—
e.g., relatively peripheral perceptions that are nonetheless conscious.
- Van Gulick tells us that the global state "embodies . . . reflexive self-awareness," partly because the complex or many of its parts refer to oneself.
- But globally connected mental states may well do that without being conscious.
- TP requires that one be conscious of the state, not simply conscious of oneself.

- For a state to be conscious, on Kriegel's view, it is not enough that a higher-order state occur with it; the two must be fused into a single state, perhaps by occurring in neural synchrony.
- But an explanation of consciousness must be cast in the first instance not in neural terms, but in mental terms; only after we understand consciousness in mental terms can we find its neural correlate.
- Moreover, since the fusion of the two states is unnecessary for implementing TP, it's unclear that there is any reason to insist on it.

III. The Argument from Phenomenological Appearances

- Arguments for intrinsicism often appeal to the phenomenological appearances.
- Such an appeal might argue either
 - ➔ (1) that the theory is needed to save the phenomenological appearances; or
 - ➔ (2) that the phenomenology reveals something about the nature of the underlying mental reality.

- But these are very different.
- Any theory must explain, and do justice to, the phenomenological appearances, and so must rely on an accurate description of those appearances.
- But that explanation need not take those appearances to reflect or match the actual mental properties of the relevant states.
- The consciousness of mental states is simply how those states appear to one—i.e., the way one is conscious of oneself as being in various mental states.
- A state's being conscious is a matter of how one represents that state to oneself.

- Since the consciousness of a state is the way the state appears to one, the actual mental properties a state has may diverge in some respects from the way one represents that state to oneself—i.e., from its phenomenological appearances.
- Phenomenology is (first-person) mental appearance, and crucial to get right; but there is more than that to mental reality.
- That a state's phenomenological appearance may diverge from its actual mental properties is evident from cases of mental states that aren't conscious.

- That's because mental states that aren't conscious have no phenomenological appearance.
- And, as with other cases of being conscious of things, when a state does become conscious, there's no guarantee that the way one is conscious of it will match all of that state's actual mental properties.
- When we explain the phenomenological appearances, moreover, we must do so by appeal to the actual mental properties of whatever mental states are involved in those appearances.

- Indeed, explaining the phenomenological appearances by appeal simply to those appearances themselves would at best be uninformatively circular.
- The phenomenological appearances are the explanandum—what is to be explained—
not the explanans—the factors that do the explaining.
- Some theorists urge that this distinction between phenomenology and underlying mental reality is problematic; I'll address that alleged problem in my concluding section.

- Because phenomenology and underlying mental properties may diverge, phenomenological appearances are not always a reliable guide to underlying mental reality.
- One example is especially crucial to how we implement TP: We are aware of our conscious states in a way that seems phenomenologically direct, unmediated.
- But it may be that much actually mediates between those states and our awareness of them; it's just that nothing seems, phenomenologically, to mediate.

- Compare perceiving: Nothing seems, phenomenologically, to mediate between physical objects and our perceptions of them, though much actually does.
- Similarly, it might be that a lot actually mediates between a conscious state and our higher-order awareness of it, even though it seems, subjectively, i.e., phenomenologically, as though nothing does.
- The phenomenological appearances alone tell us only that consciousness seems unmediated, not that it actually is.

- We can explain the apparent immediacy as due to HOTs that rely on no conscious inference ("noninferential HOTs").
- That condition is not ad hoc, as Van Gulick and Kriegel independently maintain.
- Rather, it's invoked specifically to implement TP: It's to capture the way our awareness of our conscious states seems subjectively direct—i.e., unmediated.
- A state is not conscious if one is aware of it just by consciously observing oneself or taking somebody's word about the state.

- Kriegel has urged that the appeal to nonconscious inference can't work.
- That's because conscious inferences always rest on conscious premises; but the state HOTs refer to isn't conscious until a HOT makes one conscious of it.
- But the premises one uses in inferring that one is in some mental state have to do only with one's behavior, the application of a theory to oneself, and what others tell one.
The state itself does not figure as a premise in one's inference.

- By itself, moreover, intrinsicism does not explain the phenomenological immediacy of our higher-order awareness:
That awareness might *be intrinsic* and yet *be mediated—and even seem to be.*
- Not only might there be factors internal to the state that mediate one's higher-order awareness; internal factors might actually make it *seem to be mediated.*
- (Intrinsicism is irrelevant to *referential immediacy*, should that obtain, and *causal immediacy* is unlikely, since something must mediate states' switching between being conscious and not.)

- It doesn't *seem phenomenologically* as though the higher-order awareness of conscious states consists in *distinct states.*
- But that doesn't support intrinsicism:
Distinct higher-order states (e.g., HOTs) would themselves *seldom be conscious.*
- A HOT is conscious only when one has a *third-order thought* about it—
as in *introspective* consciousness.
- And, when the higher-order state is *not* conscious, one is aware only of the first-order target;
so it will then seem *subjectively* as though there is only a single state.

- Since we're seldom aware of the higher-order awareness of conscious states, it seldom even *seems* phenomenologically as though our higher-order awareness is intrinsic to the state it's about.
- It's even unclear that intrinsicism can explain the *difference between introspective and nonintrospective consciousness*.
- Intrinsicists typically urge that attention shifts in introspecting to the higher-order awareness, so that the higher-order content is focally, and not merely peripherally conscious.

- But there are problems with that.
- For one thing, that requires that the higher-order content already be conscious, albeit only peripherally, and it's unclear that it generally is.
- One inviting model for such shift in attention, endorsed by Van Gulick, is Chris Hill's volume-control model of introspection:
Attention to a state's mental properties enhances its intensity and vividness, and even its character and existence.

- But we have no reason to think that these changes are in the target state's mental properties, as against in the way we're conscious of those properties.
- I.e., we have no independent evidence—say, from priming or other experimental work—that changes in introspected states are real, rather than merely apparent.
- And we need such independent support to infer from phenomenological appearances to the actual mental properties of the state one is conscious of.

- Phenomenology does tell us about our higher-order awareness when we introspect our mental states.
- But what it tells us is very likely only that our higher-order awareness doesn't seem wholly independent from the states we're thereby conscious of, not that our awareness seems actually to be intrinsic to that state.
- And these are different: States can be distinct and yet not be wholly independent—e.g., by having various causal connections.

- Kriegel urges that the higher-order content that implements TP is always phenomenologically manifest—i.e., that it is always conscious.
- If so, introspective consciousness might after all result from a shift in attention to that higher-order content, since the content would already be conscious.
- But it's unlikely that we are in general conscious of any higher-order content.
- This is clear from the case of conscious, but relatively peripheral perceptions; plainly we are conscious of no higher-order in connection with them.

- Kriegel argues that the higher-order content must be conscious because "unconscious representation does not qualify as awareness (in that it is not phenomenologically manifest)."
- He continues: "To be aware of a table, one must harbor a conscious representation of it" (his emphasis).
- And he concludes that the subjectivity of a "conscious experience cannot consist in an unconscious mental state."

- But a state can result in one's being conscious of something even when that state is not itself a conscious state.
- When we subliminally perceive, we are aware of something by way of a state of seeing that is not itself conscious.
- The subjectivity of a conscious experience consists in one's being conscious of oneself as having that experience.
- And the higher-order state in virtue of which one is thus conscious of oneself need not be a conscious state.

- So even though it seems subjectively that only one state is involved in a state's being conscious, there is no theoretical reason why there may not actually be two distinct states: The state one is conscious of and the distinct consciousness of that state.
- Again: Explaining the phenomenological appearances means explaining how those appearances arise.
It doesn't require taking the appearances themselves to reveal or match the actual mental properties of the relevant mental states.

- The phenomenological appearances tell us only how conscious states seem to be in our stream of consciousness, not how those states really are, independent of the way they appear.
- Running those together is the Cartesian error of assuming that, for mental occurrences, there is no distinction between appearance and reality: that mental appearance is mental reality.
- But it's important to distinguish the mental from the consciously mental—i.e., from consciousness.

- The appearances of consciousness are, of course, the reality of consciousness: The consciousness of mental states is the way those states appear to one.
- But not all mental states are conscious. And mental states that aren't conscious have an underlying mental reality—but no phenomenological appearance at all.
- So, even though the appearance of consciousness is its reality, the appearance of mental states may well diverge from the actual mental properties of those states.

- Intrinsicism seems simply to build the phenomenological appearances into the theory, relying on the Cartesian thesis that mental appearance simply is mental reality.
- But I've argued in any case that the phenomenological appearances will be the same whether or not the higher-order awareness of conscious states is intrinsic to those states.
So there is no theoretical advantage to supposing that it is.

IV. Troubles for Intrinsicism

- The claimed advantages of intrinsic higher-order content are arguably illusory.
- But there are also difficulties in supposing that the higher-order content that implements TP is intrinsic to the relevant states.
- A major difficulty has to do with the individuation of mental states.

- Intrinsicism 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.
- Kriegel has proposed various conditions under which there would be only one state;
but it's not obvious that they work.

- If one is *glad* that *p*, one must *believe* that *p*. So perhaps believing that *p* is actually a *part* of being glad that *p*.
And this might then model the way one state can be part of another.
- No. We can equally say that one can't be glad unless a *distinct state of believing* also occurs. Indeed, that's the usual line.
- *Binding*: Pieces of information about red and about circularity are *integrated* when a single state represents a red circle.
So perhaps that kind of integration also occurs *when one represents oneself as being in that first-order state*.

- Kriegel is clear that such binding by itself doesn't yield consciousness.
- Even though the bound representations of red and circularity often has conscious subjective character, it needn't: *Nonconscious sensations of red circles* occur, as priming shows.
- It's also not clear that *separate* informational states about redness and circularity are *mental* states at all—as against *subpersonal* states.
- So binding by itself also fails as a model for an intrinsic higher-order awareness of a mental state.

- Kriegel's most promising suggestion is 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 itself* as representing something to be red and circular.
- But, by itself, the mere availability of such contents gives us *no reason to think that any states actually have such content*, as against there simply being two states, one representing the other.

- Kriegel holds that a state will have that content whenever the relevant neural representations occur in synchrony.
- That's presumably why he insists, as noted earlier, that such neural synchrony is needed for consciousness.
- But, as noted earlier, we need an account of consciousness in strictly mental terms before we can find a neural correlate.
- And how could neural synchrony transform the content of a representation of representations of red and circularity into a single representation that "hereby" represents all three representations?

- Kriegel holds that the higher-order content could refer only to a state's first-order content, and yet make one conscious of the state's higher-order content as well.
- That, he argues, is because one can refer to a whole by referring one of its parts.
- But the relevant whole needn't include the higher-order content, since that content could refer to a state with only first-order content.
- Even if neural synchrony fused the first- and higher-order contents into a single state, the higher-order representational content would be unchanged.

- 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,
- 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, and 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 (whether it's a thought or a perception), they must be distinct states.
- (Note that mental attitudes are more than just "direction of fit"; even if a single state can have both directions of fit, it can't have two distinct mental attitudes.)
- (And note that, since being glad involves a distinct mental attitude from belief, the two must involve distinct states.)

- Since purely qualitative states have no mental attitude, they too are distinct from states of higher-order awareness, which must have an assertoric mental attitude.
- Still, it may seem that the consciousness of qualitative states calls for special treatment.
- In particular, there are issues that have suggested to some that—in the qualitative case—one's higher-order awareness cannot be distinct from the state one is conscious of.
- I'll close by considering an argument to that effect.

- Suppose that one's higher-order awareness is *distinct* from the state that it's an awareness of.
- Then perhaps one could have, e.g., a *red* sensation (which in itself isn't conscious) and a higher-order awareness of that sensation *as a green sensation*.
- One might even have that higher-order awareness *without any sensation at all*.
- *It may seem unclear what these situations would be like for one;* so it may seem best just to rule them out.
- And intrinsicism may seem appealing as a way to do that.

- But *intrinsicism doesn't rule out such absent or inverted states*.
- Even if each conscious state did represent itself, it might nonetheless *misrepresent* its other mental properties—i.e., the ones we're conscious of.
- And even if each higher-order awareness is ordinarily intrinsic to the state it's about, that doesn't show that it couldn't still *occur on its own*, without the other mental properties it purports to refer to.
- So—without some *ad hoc* way to rule out these cases—intrinsicism also allows them to occur.

- But what it's like for one in these cases is not actually problematic.
- Since consciousness is simply how one's states appear to one, a state's being conscious consists in one's being conscious of oneself as being in that state.
- That's independent of whether the higher-order awareness is intrinsic or distinct.
- So what it's like for one will in either case hinge simply on how one is aware of the state in question—
i.e., how one's higher-order awareness—
whether distinct or intrinsic—represents that state.

- Some regard these cases as impossible, since they think we can have access to mental qualities only through those qualities' being conscious.
- That necessary tie to consciousness forces a contingent tie with perception, and hence the possibility of undetectable absent and inverted qualities—
while ruling out the inversion or absence of the states we're conscious of.
- But the occurrence of nonconscious qualitative states (as revealed, e.g., by priming) undermines the necessity of the tie mental qualities have to consciousness.

- Van Gulick sees a problem of "stranded qualia," that neither the nonconscious first-order state nor a purely intentional higher-order state captures conscious qualitative character.
- But the nonconscious first-order state does capture the functional role of qualitative character.
- And the higher-order state makes one conscious of oneself as being in such a state.
- The two features of so-called qualia are carried by distinct states in a useful explanatory division of labor.

- And it's in any case theoretically a good thing to *hold open* the possibility of inaccurate higher-order awareness.
- That possibility is testable—again, e.g., by priming experiments:
Some qualitative occurrence might prime a subject for red qualitative states, even though the subject *reports* and is *conscious of* that qualitative occurrence as being green.
- Positing distinct higher-order states not only ensures theoretical flexibility; it reflects the difference between mental appearance and mental reality.

Summary

- Intrinsicism is a kind of higher-order theory whose apparent advantages over theories that posit distinct higher-order states rest either on conflating phenomenology with explanation, or on not noticing that the same issues arise whether the higher-order awareness is intrinsic to the states one is conscious of or distinct from those states.

THE END