**Metacognition and Consciousness**

David Rosenthal
CUNY Graduate Center
Philosophy and Cognitive Science
http://tinyurl.com/dr-cuny/

**OVERVIEW**

I. Metacognition and Conscious Awareness
II. Misrepresentation
III. Utility
1. Metacognition and Conscious Awareness

Metacognition is the cognitive registration of one's cognitive processes and states. Classic examples include the sense we sometimes have that we know or that we will recall something, and the sense that we have a word on the tip of our tongue—though many other cases of higher-order cognition are also called metacognition.

A family of theories of consciousness hold that a state's being conscious consists in one's being aware of that state. Should that too count as metacognition?

Not everybody endorses such higher-order (HO) theories of consciousness; so-called first-order (FO) theorists (e.g., Ned Block) deny that a state's being conscious involves any such higher-order awareness (HOA).

HO theorists rely on the commonsense observation that a state of which one is wholly unaware is not conscious—a test also widely used in experimental work. So a state is conscious only if one is in some way aware of it—only if one has a HOA.

This provides a natural explanation of how conscious states differ psychologically from mental states that are not conscious. A FO theory, which denies appeal to a HOA, is unlikely to be able to do that.
Some FO theories rely on a neural marker of a state's being conscious (e.g., Lamme 2010). But that won't help unless we can already distinguish conscious from nonconscious mental states, and so show it's that neural marker that subserves the conscious cases.

Other theorists say simply that we know from the inside—from a first-person point of view—what conscious states are. But that either is uninformative or reduces to having a HOA of the conscious cases.

Some FO theorists say that conscious states are those that involve attention, but much recent work shows a double dissociation between consciousness and attention (Koch & Tsuchiya 2006; Bokstel 2011; Norman et al, submitted).

Dretske has urged that a state’s being conscious consists not in one’s being aware of that state, but in the state’s making one aware of something. But subliminal states do make one aware of stimuli—though not consciously aware of them.

A popular type of FO theory appeals to a global workspace (GW): Conscious states are available for processing by many cortical subsystems (Dehaene; Baars; cf. Block’s access consciousness and Dennett’s cerebral celebrity).

But many desires and beliefs likely have relatively global effects without being conscious. And peripheral, unattended visual or auditory states can be conscious without having any global effect at all.
Indeed, unlike being aware of a state, it’s unclear why global accessibility has any bearing on whether a state is conscious.

The intuitive idea is that a state’s being conscious enhances its rationality in thought, decision making, and executive control. So global accessibility would accompany consciousness.

I’ll take up the utility of a state’s being conscious in §III. For now I’ll just note that, as with a neural marker, showing that consciousness does enhance a state’s utility requires already knowing how conscious and nonconscious mental states differ. Being aware of utility when a state is conscious isn’t enough.

If the default for mental states were that they’re conscious (Descartes 1641, Dennett 1991, Block 2011a), a psychological account might not be needed of how conscious and nonconscious mental states differ. It might simply be that something blocks some states from being conscious.

And FO theories would fare better if they didn’t need to explain that difference, since that was their main difficulty. And then the appeal to HOAs would seem at best redundant.

But we have ample reason—from masked priming (Marcel 1983a) and blindsight (e.g., Weiskrantz 1997)—to hold that mental states do occur often without being conscious.
So we can assume that some HOA figures in a state’s being conscious. I won’t today go into what kind of HOA, though I’ve argued elsewhere that the HOA must be a thought that one is in the relevant state.

Is this HOA metacognitive? The classic cases of metacognition I mentioned earlier suggest important differences.

Most crucially, such cases of metacognition may not result in one’s being aware of relevant lower-level cognitive contents.

Suppose one has Orwell’s real name on the tip of one’s tongue (TOT). One has a sense that one knows the name, but the name escapes one—at least at that time. Does this metacognition result in consciousness?

One is aware of having the knowledge in question, but unaware of the knowledge itself. One is at that time unaware that Orwell’s real name is ‘Eric Arthur Blair’. The knowledge itself is not conscious. One only has a conscious sense of having it.

Not every HOA results in consciousness—in consciousness of the state one’s HOA is about. In this case, one is aware of some knowledge—aware of having it—but the knowledge one is aware of isn’t conscious.

Similarly with many other metacognitive cases of sensing that we know things.

How can we explain that—and explain it compatibly with a HO theory of what it is for a mental state to be conscious?
As a first pass, note that in TOT one has a sense that one will be able to recall the information, though one can't right now. Similarly with so-called feeling-of-knowing judgments (Nelson & Narens 1994). That isn't a HOÂ that one is in a mental state.

But TOT is more complex. One is, after all, aware of being in a particular mental state—a state that carries the information that Orwell’s real name is ‘Blair’. So why isn’t that state a conscious state?

One is aware of being in a state that says what Orwell’s real name is—but one is not aware of the state in respect of its content. Hence the strange TOT feeling: The state is conscious in one way but not in another.

One has a thought that Orwell’s real name is ‘Blair’. That thought is not conscious as described in that way: One is unaware of having the thought that his name is ‘Blair’.

But one is aware of having a thought that provides his real name; described in that way, the thought is conscious.

This is an important phenomenon, which I’ll come back to. States can be conscious in respect of some mental properties but not others. Often this goes unnoticed; we may be aware of sensing red, but not at all aware of sensing a specific shade of red. TOT is a dramatic example—in which this kind of disparity is not only noticed, but leads to a dissonant, uncomfortable feeling.
Some (Levin et al 2000) see change detection as metacognitive: One detects changes in one’s perceptions. But metacognition need not even here result in the cognition’s being conscious; change detection occurs even when the perceptions aren’t conscious (Fernandez-Duque et al 2000; Laloyaux et al 2003).

HOAs are seldom themselves conscious; a HOA would be conscious only if one were aware of it—only if it were accompanied by a third-order awareness. Metacognitive states, by contrast, are often conscious—though not always (Kentridge & Heywood 2000).

It matters that metacognition differs from HOAs, since some (e.g., Seth 2008) just dismiss HO theories as implausibly metacognitive.

Metacognition does in some ways resemble the HOAs that result in mental states’ being conscious. Both HOAs and metacognitive states, e.g., appear subjectively not to depend on inference, observation, or other mediation.

But a salient difference lies in their content: Metacognitive judgments are typically about what one can or will know or recall. HOAs, by contrast, are about what mental states one is currently in, described in terms of those states’ mental properties.

But there’s no harm in extending the notion of metacognition to cover HOAs so long as one keeps differences from standard cases in mind, and there’s some reason to do so.
If one has little confidence in the accuracy of a perceptual decision about a stimulus, it's natural to see that as indicating that one was just guessing about the stimulus.

Guessing indicates an absence of conscious awareness, and high confidence indicates one isn't just guessing.

A subject's confidence rating of accuracy should thus reflect whether a perceptual state is conscious (Peirce & Jastrow 1884; Cheeseman & Merikle 1984, 1986; Merikle 1992).

Appeal to explicit confidence ratings has been refined by having subjects place bets on a perceptual decision (Persaud et al 2007), though this may be no more accurate than explicit confidence ratings (Dienes & Seth 2010).

High confidence and willingness to wager might be due to many factors. Still, they typically coincide with subjective reports (Persaud et al 2007; Dienes & Seth 2010) — presumably because lack of conscious awareness results in low confidence and hence a reluctance to wager.

Confidence and wagering likely vary even with the different respects in which one’s perceptions can be conscious.

E.g., in Sperling’s partial-report paradigm, subjects are briefly presented with a 3 x 4 matrix of letters. Without subsequent cuing, subjects are consciously aware of all 12 as letters — but can identify only about 3-4 of them.
But subjects can also identify 3-4 letters in any of the three rows—if cued for that row even after the display has vanished. In both cases they identify only 3-4 letters. But since they can do so for any cued row, the FO information must be there for all 12. So the limit is in subjects’ HOAs—which represent their perception as being of 12 letters, but represent specific identities only for 3-4 of the characters.

My hunch is that subjects would also be confident and would bet that they’re all letters—but about specific identities only for 3-4. Confidence and betting would coincide with the specific respects in which one is aware of perceiving the characters.

Subjects’ awareness of all the letters just as letters might generate confidence that they could identify all the letters; generic identifying often goes with being able to identify it specifically. So they might feel sure pre-cue that they could identify all. This might explain why some Sperling subjects say that they consciously saw more than they can remember (Sperling 1983).

But confidence about identifying them all the letters might quickly vanish after several trials in which subjects learned that they successfully identified only 3-4.

It’s worth testing whether confidence and wagering does vary in these ways.
The HOAs that figure here are distinct in nature from metacognitive judgments. But HOAs do result in perceptions’ being conscious in ways that correspond to one’s willingness to wager and one’s not having a sense that one is guessing.

And willingness to wager and awareness of guessing resemble judgments of learning and feeling of knowing—both paradigmatic metacognitive judgments, though they focus on future cognitive performance.

So the appeal to conscious guessing or wagering to gauge conscious awareness does tie such awareness to some standard types of metacognitive judgment. We can think of HOA as a *type* of metacognition.

---

**II. Misrepresentation**

Metacognitive judgments plainly sometimes misrepresent one’s cognitive states. One sometimes judges erroneously that one will recall something or that one has information one cannot now retrieve.

And metacognition may seem in this way to differ from our subjective awareness of our conscious states. Consciousness, one may think, cannot misrepresent mental states. If it seems subjectively that I feel or see something, perhaps these subjective appearances simply cannot be mistaken.
The dominant view, traditionally and even today, is that consciousness cannot fail to represent our mental lives accurately.

Indeed, this is often advanced (Neander 1998, Levine 2001, Block 2011) as a decisive reason to reject HO theories. If a state's being conscious consisted in a HOA of that state, that HOA can go wrong in ways assumed impossible for consciousness. This makes it tempting to see HO theories as explaining consciousness as a type of metacognition.

Still, none of this can be much of a worry for HO theories. Such theories don't say that consciousness does misrepresent, only that it could. So if it turns out it can't, HO theories can simply stipulate as much.

But we should also look more closely at the claim that consciousness can't subjectively misrepresent the mental states we're in.

On reflection, this view is quite surprising. Consciousness is the way our mental lives appear to us, and appearances don't always match the way things actually are. Why, then, shouldn't our subjective appearances about our mental lives sometimes do so? And if they do, consciousness won't after all differ in this way from metacognition.

No reason is ever given for the claim that consciousness cannot misrepresent our mental lives; it's supposed just to be obvious. So it's worth asking why this traditional view is so deeply entrenched.
We can readily check whether appearances are accurate when non-mental reality is at issue. But it may seem that we can’t test whether subjective awareness represents our mental lives accurately. Subjective access is often seen as the gold standard—the only standard—about mental reality, with no way to correct itself. But subjective access isn’t always accurate if some mental states occur nonconsciously, since subjectivity fails to reveal them. So we must balance first- and third-person considerations about mental functioning, taking both into consideration, permitting neither to be the last word. Third-person access often trumps subjectivity.

In TOT we’re subjectively aware of a state as saying what Orwell’s real name is, but unaware of it as saying it’s ‘Blair’. Similarly, we’re often aware of seeing red but not any particular shade, though the visual state must represent an exact shade. The HOAs in these cases represent the states incompletely—a kind of inaccuracy. Are HOAs ever inaccurate more robustly? John Grimes (1996), working in George McConkie’s lab, used eye trackers to switch displays during saccades. 18% of subjects were unaware of a salient change from red to green. So when the stimulus was green, they were aware of a perception of green as if it were a perception of red.
Consciousness here dramatically misrepresents our perceptual states:

A less exotic case: We can discriminate similar colors more finely when presented together than in succession. ‘This affects the subjective appearances; sensations that seem indistinguishable in succession seem distinct when together:

What it’s like to have each sensation differs in the two cases—though the sensations in themselves are distinct, since the shades are distinct. Consciousness misrepresents the states when they occur in succession.

Consciousness can also misrepresent our thoughts, desires, and intentions. Nisbett & Wilson (1977) found that though subjects preferred identical fabrics placed to the right, they denied that position effect and confabulated reasons for their preference. In these cases, subjects were subjectively aware of themselves as actually having the confabulated reasons. Consciousness misrepresented them as being in states that reflected those confabulated reasons.

Consciousness, like metacognition, can and often does, misrepresent what mental states we’re in. And on a HO theory of consciousness, this is not only to be expected, but is readily explained.
It tends not to be noted that GW theories also allow for misrepresentation by consciousness, and readily explain it. On a GW theory, states are conscious in virtue of the availability of their content for various cortical systems.

But it's possible that a state's content is made available to those systems in a way that's distorted or otherwise inaccurate. The properties available to those systems would diverge from those of the state; so the state would presumably be conscious in respect of mental properties not its own.

Misrepresentation is not, as usually held, possible only on a HO theory, but on the most prominent FO theory as well.

Though it's no surprise that metacognition can—and sometimes does—misrepresent, a strong intuitive sense may persist that consciousness simply cannot do so.

That goes with an equally robust subjective sense that the property of a state's being conscious is inseparable from the state and its other mental properties. These appearances seem to favor a FO theory—though not a GW theory, on which being conscious is distinct from the state.

How can these compelling intuitions square with a HO or GW theory, on which misrepresentation can and presumably does occur, and a state’s being conscious is distinct from that state?
Any acceptable theory must do \textit{justice to the subjective appearances}. But it need not do so by taking the appearances to be veridical. It can do so instead by \textit{explaining why it is that we have the subjective appearances in question}.

We \textit{must} in any case explain that. And the appearances’ being veridical wouldn’t help at all—except on the dubious assumption that the mind is transparent to itself.

And we can explain why subjectivity seems always accurate. We’re rarely aware of the HOA; so it seems subjectively that there’s only one state. So being conscious seems inseparable from the state, leaving no room for misrepresentation.

Another criticism of HO theories is that they might allow one to be subjectively aware of any mental state one might wish (Balog 2000). That seems especially so for my view, on which the HOA is a thought—a state with intentional content and an assertoric mental attitude. Can’t one, after all, think whatever one wishes?

No. One can’t produce subjectively unmediated HOAs at will; we can’t simply think assertorically whatever one pleases.

As misrepresentation, consciousness again, and again perhaps surprisingly, resembles metacognition. Metacognitive access to one’s cognitive states is also not in standard cases within one’s control.
III. Utility

I’ve contested the prevailing view that, whereas metacognition can misrepresent our cognitions, subjective consciousness cannot misrepresent our mental lives.

Let me turn now to another comparison of consciousness with metacognition, again taking issue with the dominant view. Metacognition plainly has substantial utility for psychological functioning, and it’s typically held that consciousness does as well. I’ll argue that consciousness actually has little utility and that—surprisingly—it’s in this way unlike metacognition.

It’s useful to have a sense of how much one knows or will recall; so metacognition can have very substantial utility. But such metacognitive judgments don’t result in states’ being conscious; so metacognitive utility needn’t carry over to consciousness.

Still, it’s tempting to think consciousness has utility because we often understand phenomena by appeal to their utility. But not always; we also often understand them by appeal to what causes them.

We’re subjectively aware of a state’s utility when the state is conscious, but that utility needn’t depend on its being conscious. Indeed, consciousness is blind to whatever utility nonconscious states might have.
We should dispel some sources of possible equivocation. Cohen & Dennett (2011) urge that states’ being conscious must have some function. But they mean just that their being conscious must have some discernible effect, e.g., reportability. That doesn’t automatically confer utility. We must also distinguish whether some particular utility depends on or simply accompanies a state’s being conscious. Perhaps neural signal strength is needed for a state to be conscious—as many hold. Signal strength might also result in utility of the state. But if a single factor yields both utility and consciousness, utility wouldn’t be due to the states’ being conscious.

Finally it’s crucial to distinguish a state’s being conscious from a creature’s being conscious or its being conscious of things. The obvious utility of the last two doesn’t imply utility for the first, since each can occur independently of the others. All that aside, there are two main reasons states think a state’s being conscious has utility: rationality and social interaction. Rationality is a matter of whether the states’ representational contents fit together rationally. So the challenge for thinking that utility derives from rationality is to show that the rational ties don’t result simply from causal ties among the states, independent of the states’ being conscious.
If causal ties among the states result in their contents’ fitting together rationally, that utility is independent of consciousness. And there are theoretical and experimental reasons to think that this is the case.

Representational content involves causal ties among mental states; otherwise the connections among states with similar and different content would be mysterious. That's a theoretical reason to see rationality as due to content, not consciousness.

Contents among states can conflict, and that can lead to revising one’s thinking and planning, correction of errors, and executive control that inhibits or enhances particular course of action and thought.

And experimental results show states have a rational effect before becoming conscious (Libet 1985, Haggard 1999) and that nonconscious planning can be better than conscious planning (Dijksterhuis et al 2006; Usher et al 2011).

Those considerations pertain not only to thoughts and desires, but to perceptions as well. Perceptual role provides ties among perceptual states that subserve rationality and related types of utility.

Nonconscious perception in blindsight, prosopagnosia and amnesia impair flexible thinking about relevant matters (Weiskrantz 1997). But it may be hard to tell that this is due to the states’ not being conscious, as against other deficits in these conditions.
Larry Jacoby’s (1991; Debner and Jacoby 1994; Jacoby et al 1993, 1994) exclusion task is sometimes held to support a tie between consciousness and intentional action. This would constitute a kind of utility for consciousness that’s related to rationality.

Subjects are visually presented with a word, say, ‘reason’, and asked to complete a word stem, say ‘rea—’, with any word other than the presented word.

When a word is presented for 500 ms, subjects see it consciously and mainly succeed in following the instruction; when it’s presented for only 50 ms, they report seeing no word—but tend to complete the stem with the word that was presented.

So subjects intentionally exclude a word only when they consciously see it.

But that’s not because intentional action requires consciousness:

Subjects are instructed not to complete the stem with a word they see. But when they’re aware only of seeing a blank screen (which has greater neural signal strength than the nonconscious sensation of the word), they’re not aware of seeing the word—so they think they don’t see it.

It’s not that consciousness is needed for intentional action. Rather, subjects following the instructions in the 50-ms case think they don’t see the word (though it still primes their response).
What about utility due to social interaction? It’s useful to know what others are thinking and feeling, and we can report these states only when they’re conscious.

But reporting isn’t needed for others to know about these states; we can instead verbally express them. E.g., I can say not that I think it’s raining, but simply that it’s raining, and not that I’m angry, but that what you did was bad.

Indeed, it’s unclear that reporting has any cognitive or other advantage over verbally expressing our mental states. Reporting is a key way to tell that a state is conscious, but its being conscious may not be needed to express it verbally.

One might contest that; aren’t all verbally expressed thoughts conscious?

Yes; but not because their being conscious is needed to express them verbally. When one says it’s raining, one is equally disposed to say that one thinks it’s raining. Indeed, one may not recall which one said, even a moment earlier.

So when one says it’s raining, one’s being also disposed to say that one thinks it’s raining reflects one’s having a HOT that one thinks it’s raining. Verbally expressed thoughts are always conscious not because consciousness is needed, but because of the automatic use equivalence of saying something and saying that one thinks it.
A state’s being conscious is not needed for the social utility of others’ learning what we’re thinking, perceiving, or feeling.

Frith has developed a challenging view, on which the conscious awareness of one’s own and others’ agency yields a distinctive type of social utility. Frith notes that such awareness does not enhance control of our actions. But he urges that it does help generate social cooperation, which plainly does have considerable social utility.

In subliminal perception, one is aware of the perceived stimulus, though one is not consciously aware. So the question arises whether the awareness of agency that’s useful here must be conscious awareness.

There is reason to think not. Infants at 1 year can distinguish others’ voluntary and involuntary actions (Tomasello 2000). Perhaps that requires conscious awareness of the movements, but it’s by no means obvious.

And newborn chicks distinguish animate from inanimate movements (Mascalzoni et al 2010), an ability presumably related to the ability to distinguish voluntary from involuntary. It’s even less likely that this requires relevant conscious awareness.

One can report one’s mental states only if they are conscious. And reporting is a social act. But one can also communicate one’s states by expressing them verbally, and for that they needn’t be conscious.
Some see the HOAs that figure in states’ being conscious as related to the ability to tell what mental states other are in (e.g., Carruthers 2000), which has great social utility. And if the two abilities are related, that may point to a utility for the HOAs. But the two abilities evidently rely on very different factors. Our mind-reading ability depends on inferring from observations of others. HOAs might rely on inferring as long as it isn’t conscious, but it does not rely on self-observation. Since mind reading and HOAs must rely on different factors, it’s unlikely that they are related, and so unlikely that the utility of mind reading carries over to consciousness.

GW theories build utility into a state’s being conscious, since availability to cortical systems plainly has utility. Similarly for Block’s notion of access consciousness. Do my arguments against utility, then, beg questions, by relying on HO theories as against GW or other theories? No. The arguments against utility do not rely on adopting a HO theory, though many considerations that tell against utility also support HO theories. But they do the two things independently. Surprising though it is, we have reason to doubt that a state’s being conscious has any significant utility. Consciousness does differ in this way from metacognition.
Summary

➢ Though there is little harm in extending the notion of metacognition to cover the HOA we have of conscious states, most forms of metacognition occur without the relevant cognitions’ being conscious.

➢ Consciousness resembles metacognition in sometimes misrepresenting the states we are conscious of being in. And it differs from metacognition in lacking significant utility. The temptation to think otherwise is due in both cases to taking at face value the deliverances of consciousness itself.

Thank you for your attention