Why were you initially drawn to philosophy of mind?

In the summer of 1956 a half-year correspondence began between Roderick Chisholm and Wilfrid Sellars, prompted by Sellars' having sent to Chisholm page proofs of his forthcoming “Empiricism and the Philosophy of Mind.” In their letters, Chisholm and Sellars debated various issues about how to understand the intentionality of mental states. Their principal dispute was about whether, as Sellars argued in his article, one can understand and explain the intentionality of thoughts by appeal to the independently understood intentionality of speech.

Sellars' still seminal, monograph-length “Empiricism and the Philosophy of Mind” appeared that year in the first volume of *Minnesota Studies in the Philosophy of Science.* A couple of years later the Chisholm-Sellars correspondence itself appeared in the second Minnesota Studies volume, along with a reprinting of Sellars’ still seminal, monograph-length “Empiricism and the Philosophy of Mind.”

---


of Chisholm’s important “Sentences about Believing.”3 The correspondence immediately provoked an excited, inconclusive, and often perplexed reaction among many in the philosophical community.

A few years later, I was a graduate student at Princeton, where Richard Rorty had come to be deeply influenced by Sellars’ thinking on these and many other issues. I became intrigued by the questions posed in the correspondence, and undertook a dissertation under Rorty’s direction to try to sort those out various issues about intentionality, and see where the truth lay. In addition to having Rorty’s penetrating guidance, I was very fortunate that Sellars visited Princeton in the spring of 1965. I met with him virtually every Tuesday afternoon for what turned into an informal tutorial, which led the following semester to my own, subsequently published correspondence with Sellars about those questions.4

Sellars argues in “Empiricism and the Philosophy of Mind” that we can understand intentional mental states as folk-theoretical posits, originally invoked to explain the occurrence of speech acts as well as rational behavior unaccompanied by speech. Such states are posited as having properties analogous to the semantic and illocutionary properties of speech acts, and those posited properties are the intentional content and mental attitudes of intentional states. After a time, individuals come to have subjectively unmediated first-person access to many of these posited intentional states.

Chisholm had argued that we cannot understand the relevant properties of speech acts except by seeing the speech acts as caused by intentional states. I argued that Chisholm’s argument seemed compelling only because the folk theory Sellars described had come to be deeply entrenched in our commonsense ways of thinking about speech and thought.

Following Sellars, I argued that since we can in fact understand others’ speech acts and we plainly have no access to their thoughts

---


I had originally imagined that I might specialize in some area of the history of philosophy, ancient or early modern, or perhaps in social and legal philosophy. But I became so intrigued by these issues about intentionality, mind, and language to let them go. And it came to seem to me, as it does still, that these issues are among here most fundamental in philosophy, and that getting straight about them would have a big payoff in understanding many other issues.

What do you consider your most important contribution to the field?

The work I’ve done in philosophy that’s most widely discussed has to do with the higher-order-thought theory of the consciousness of mental states that I articulated first in the early ’80s and have expanded on and developed since.5

After graduate school, I wanted a change from the issues of my dissertation, and turned to questions about mind-body materialism, which were discussed very heatedly in the ’60s and ’70s. The main problem for mind-body materialism was widely thought to be consciousness, by which people had in mind the consciousness of qualitative mental states. How could a materialist view do justice to what it’s like for one to be in a qualitative mental state?

It seemed that, to be convincing, an answer to this question would need to rely on more than just good arguments in defense of materialism. One would need in addition to provide a satisfactory explanation of consciousness, and in particular of qualitative consciousness. I did not think that the eliminativist program of Rorty, Paul Feyerabend, and Paul Churchland6 was acceptable, and at the end of an article arguing as much I put forth in outline a theory of consciousness.7 On that theory, mental states, includ-

7 Keeping Matter in Mind," Midwest Studies in Philosophy, V (1980): 295-
ing qualitative states, are never intrinsically conscious; so mental
states, again including qualitative states, often occur without be-
ing conscious. The conscious states differ from those that are not
conscious because of the occurrence in the conscious cases of some
factor extrinsic to the state itself.

It’s clear that when a mental state is conscious, the individual
who is in that state is in some way aware of the state. If one is in
some mental state but wholly unaware of being in it, that state
does not count as being conscious, on any intuitive, commonsense
understanding of consciousness. So the extrinsic factor that differ-
entiates conscious from nonconscious mental states must be some
kind of awareness of those states that are conscious. So the im-
portant thing would be to see what kind of awareness is operative
here.

David Armstrong had argued, following Locke and many others,
that the awareness is perceptual in nature. But it seemed to me
that no such higher-order perceptual awareness actually occurs.
For one thing, there are no higher-order mental qualities, and
no awareness will count as being perceptual if it altogether lacks
such qualities. The only alternative is that we are aware of our
conscious states by having higher-order thoughts (HOTs) about
those states. A HOT is simply a thought to the effect to one is in
a particular mental state.

For the target state to be conscious, one’s awareness of it must
seem to arise spontaneously and without any mediation. This will
be so if it seems subjectively that there is no inference or self-
observation that leads to one’s awareness of the target, that is,
if one is aware of no such inference or self-observation. The HOT
theory can easily provide for this. For one thing, the HOTs are
seldom themselves conscious, since there will seldom be third-
order thoughts about them. Because HOTs are seldom subjectiv-
ely available, it won’t seem subjectively as though the aware-
ness of mental states that results from such HOTs is mediated.
Positing HOTs can thus explain how we are aware of those of our
mental states that are conscious.

8 D. M. Armstrong, *A Materialist Theory of the Mind*, New York: Hu-
Paul, 1993; "Is Introspective Knowledge Incorrigible?*, *The Philosophical Re-
view* LXXII, 4 (October 1963): 417-432. See John Locke, *An Essay Concern-
ing Human Understanding*, ed. Peter H. Nidditch, Oxford: Clarendon Press,
1975/1700, e.g., p. 105 (II, i, 4).
Positing such HOTs to explain how conscious states differ from mental states that are not conscious has turned out to have many theoretical benefits, allowing informative explanations of things that it’s unlikely any other theory can handle. For example, HOTs allow informative explanations of why a state’s being conscious coincides with its being noninferentially reportable, why thoughts are always conscious when they’re expressed verbally though not when they’re expressed by nonverbal behavior, and why our conscious mental lives appear subjectively to be unified.9

The HOT theory fits nicely with Sellars’ account of the connection sketched earlier between the intentionality of thought and that of speech. Sellars argued that we could see mythical ancestors as having posited inner states with intentional properties to explain verbal and nonverbal behavior. The ancestors that deployed that theoretical structure would in time have come, he urged, to have first-person access to some of their intentional states. But as students of Sellars have noted, he glides rather quickly past the question of how it is that this first-person access might come about.10

The HOT hypothesis provides an informative answer. On Sellars’ account, our mythical ancestors posit intentional states to explain their own speech acts as well as those of others. Sellars usually presents this positing as an early stage for human thinking about mind in general. But he also occasionally notes that we can understand such positing also as an early stage in the development of each individual.11 With practice, each person will become adept and habituated to applying to oneself the folk theory that posits intentional states.

As with the application of any folk theory, in time this positing of intentional states will become relatively automatic. At that point, one’s simply being disposed to say something will by itself dispose one to ascribe to oneself a internal state that has intentional properties that correspond to the speech act’s semantic properties. This disposition to ascribe an intentional state to

---

9See Consciousness and Mind, chs. 2, 10-11, and 13, respectively.
10All Sellars says is: "And it now turns out—need it have?—that [one] can be trained to give reasonably reliable self-descriptions, using the language of the theory, without having to observe his overt behavior" ("Empiricism and the Philosophy of Mind." 59).
oneself will result in a thought to the effect that one is in that very state. Because one comes to have a HOT that one is in that state, the state itself comes to be conscious, and one thereby has first-order access to it. The HOT hypothesis helps sustain Sellars’ theory of the nature and origin of our folk understanding of intentionality.

It’s widely acknowledged that the most difficult aspect of consciousness is the consciousness of qualitative mental states, which has been the focus of much discussion in connection with mind-body materialism. And many have thought that, however well the HOT theory may handle the consciousness of intentional states, it cannot handle qualitative consciousness, that is, there being something it’s like for one to be in qualitative states.

Qualitative consciousness may indeed be intractable to explanation if one regards qualitative states as essentially conscious, so that such states cannot occur without being conscious. But qualitative states do often occur without being conscious, as in subliminal perceiving.

So being conscious cannot be essential to a state’s being qualitative. And even when qualitative states are conscious, their being conscious, which they have in common, is distinct from the qualitative character, which differs from one state to the next. So in addition to a theory of what it is for mental states to be conscious, we also need an independent theory of what it is for mental states to have qualitative character.

Unlike Sellars’ view of intentionality, which fits well with an informative theory of consciousness, his view about qualitative mental states makes it more difficult to explain consciousness. For one thing, Sellars held that qualitative states are necessarily conscious, which makes for a mystery about how such necessarily conscious states could fit within the natural order. In the face of that apparent mystery, therefore, Sellars seeks to sustain mind-body materialism by holding that the commonsense conscious qualities we think of our mental states as having can somehow be relocated in a new guise within an ultimate materialist framework.

But Sellars’ actual account of qualitative mental properties can be detached from his insistence that mental qualities cannot occur without being conscious. And once we drop the view that mental qualities always occur consciously, Sellars’ account of mental qualities yields, with some adjustment, an informative theory of mental qualities independent of consciousness.

As Sellars observed, our folk descriptions of the mental quali-
ties of perceptual states appeal to the similarities and differences that hold among those qualities. And we can get a handle on the similarities and differences that hold among mental qualities that occur in perceiving by noting that they are parallel to the similarities and difference that hold among the physical properties we perceive. Just as we can understand the intentional content of thinking in terms of the semantic properties of speech, so we can understand the mental qualities of perceiving in terms of the perceptible properties of physical objects.

And that insight about mental qualities can, moreover, be readily detached from claims about the consciousness of qualitative states. Mental qualities are the properties of perceptual states in virtue of which we discriminate perceptible properties, whether such discriminating occurs occurs consciously or not. So we have in Sellars’ insight about mental qualities the making of an informative theory of qualitative character, a theory that can fruitfully combine with an independent theory of consciousness.

On this theory of mental qualities, we individuate each mental quality by reference to its position in a quality space distinctive of the perceptual modality in question. And that quality space is homomorphic to the quality space that defines the physical properties discriminable by that modality. Since perception can discriminate perceptible properties even when the relevant perceptual states fail to be conscious, the resulting theory of mental qualities is independent of consciousness.12

This homomorphism theory of the mental qualities fits well with the HOT theory. On the HOT theory, a state is conscious if one has a thought to the effect that one is in the state in question. So when a qualitative state is conscious, the HOT in virtue of which it’s conscious describes the mental quality of its target in terms of that quality’s position in the relevant quality space. This enables the combined theories to deal effectively with a number of traditional issues, such as the alleged conceivability of inverted or absent qualities, and also to fit comfortably with many empirical findings, such as masked priming, confabulatory consciousness, and blindsight (see §IV).

---

12 See Consciousness and Mind, chs. 5-7.
What is the proper role of philosophy in relation to psychology, artificial intelligence, and the neurosciences?

Every discipline seeks to establish boundaries that are reasonably well recognized, based on features that help fix those boundaries. But the accumulation of knowledge and development of theory in the last century or so has made such boundaries ever more difficult to draw. Mathematics now gets input from physics, and there are increasing disciplines that build bridges between adjacent sciences, such as physics, chemistry, and biology.

A traditional way to delimit philosophy has been to urge that philosophy, like logic and mathematics, deals with truths that hold independently of any empirical input. But as W. V. Quine, Sellars\textsuperscript{13}, and others have forcefully driven home, this tradition is indefensible. It is very likely impossible to isolate either a purely conceptual or a purely empirical aspect of any truths whatever.

Work in philosophy of mind does somehow seem different from research in experimental psychology, cognitive neuropsychology, and related areas of scientific investigation. But if we follow Quine and Sellars in rejecting the idea that some sentences are true come what may, we can’t account for that apparent difference by saying that philosophy of mind deals with pure conceptual analysis or metaphysically necessary truths.\textsuperscript{14} And philosophy of mind must in any case respect findings in experimental psychology and cognitive neuropsychology, and is at least to that extent continuous with those fields.

Many follow Quine’s suggestion that philosophical work involves the search for truths that are located near the center of our web of belief.\textsuperscript{15} Such truths seem to resist revision, but only because rejecting them would prompt us to reject such a large number of other beliefs as well. In contrast, disciplines such as experimental psychology and cognitive neuroscience work to establish truths.


\textsuperscript{14}For useful doubts about the metaphysical necessity, see Jerry A. Fodor’s Eastern Division APA Presidential Address, "What is Universally Quantified and Necessary and A Posteriori and It Flies South in the Winter?", \textit{Proceedings and Addresses of the American Philosophical Association}, 80, 2 (November 2006): 11- 24.

located nearer the periphery of that web of belief. Because those truths have sparser connections with the rest of the web, their ties to empirical discovery are more salient.

This model does far better at capturing the difference between philosophical investigation and work in adjacent areas than a model that construes philosophy in terms of conceptual analysis. Little in Aristotle’s *Metaphysics* or Descartes’s *Meditations* or Kant’s *Critique of Pure Reason* seems to be intended as conceptual analysis, proper. It’s more credible to see these works as theorizing about various topics at a very high level, further removed from immediate empirical input than the sciences. Like philosophy, physics tells us about substance and physical reality and psychology about the nature of mind and knowledge. But theory building in the sciences is relatively tightly connected to empirical input. Philosophy, by contrast, is theorizing at a relatively abstruse, sometimes speculative level. The difference between distinctively philosophical work and work in the sciences is a matter of degree, not of kind.

This way of seeing the difference explains the many fruitful interactions between philosophical and more strictly empirical work on the mind. Philosophy cannot and should not try to dictate to the sciences, about the mind or anything else. But philosophy can suggest and evaluate theoretical contexts within which we can seek to understand the empirical findings of experimental psychology and cognitive neuroscience. Philosophy of mind can help guide the theorizing that occurs in the sciences proper.

Indeed, some psychologists and cognitive neuropsychologists have welcomed theorizing from the philosophy of mind in what seems to be just this spirit. Thus, neuropsychological theorizing by Lawrence Weiskrantz and Edmund T. Rolls16 and psychological theorizing by Zoltán Dienes and Josef Perner17 have taken on aspects of the higher-order-thought model. In the opposite direction, Austen Clark has invoked work in psychophysics to develop a powerful

---


It is doubtful that philosophy of mind will disappear or merge into those scientific disciplines; relatively speculative, high-level theorizing will continue to be fruitful and interesting. But there is every reason to expect that such productive cross-fertilization will continue to grow.

Is a science of consciousness possible?

The scientific study of consciousness is not only possible; it’s actual and under active development. Doubts about the scientific treatment of consciousness stem mainly from seeing qualitative consciousness as accessible only from a first-person point of view. Qualitative consciousness is a matter of what it’s like for one to be in a state with qualitative character. And it’s tempting to hold that what it’s like for an individual can only be known by that individual. If so, my conscious experience of red physical objects might, for all anybody can know, be what it’s like for you to experience green objects, and conversely.

There are two compelling reasons to doubt that our knowledge about conscious qualitative character is limited to such first-person access. One reason stems from considerations that pertain to the so-called other-minds problem. In many circumstances, you will know whether you’re in pain and I can at most guess. But sometimes there can be no serious doubt about whether you’re in pain, even from a third-person point of view. If you suffer great bodily damage and you’re writhing and crying out in an agonized way, nobody who sees you can have any serious doubt about it.
about whether you’re in pain. Successful pretence may occur with middling pain, but not with excruciating, agonizing pain. I may not know exactly how your pain feels to you, but it’s simply silly to suppose you might not be in pain.

So in certain cases at least we have third-person access to others’ mental states. The trick in philosophy of mind is to make the occurrence of such third-person access fit theoretically with the acknowledged first-person access we have to many of our own mental states. But just as we should not seek to solve that problem by jettisoning or even minimizing our first-person access, so we mustn’t jettison or minimize our third-person access.

We need, then, a theory of qualitative character on which we can explain both first- and third-person access. The combined homomorphism and HOT theories described above are one good way to go. Together they show how we have first-person access, by way of HOTs about our own qualitative states, and third-person access, by the tie mental qualities have to the perceptible properties that the various mental qualities enable us to discriminate. And once we see how we have third- as well as first- person access to conscious qualitative states, reasoned resistance to the scientific treatment of consciousness dissolves.

There is another reason to reject the idea that qualitative consciousness is accessible only from a first-person point of view. Qualitative character occurs when there is something it’s like for one to be in a qualitative state. But qualitative character also occurs without being conscious. Indeed, it’s here that some of the most striking empirical findings help. There are experimental paradigms in which subjects are unable to report perceptible stimuli, either because the stimulus is so brief or because it is followed so quickly by another, masking stimulus. Since the stimulus is not reportable, it’s not registered consciously. But despite this, the stimulus will in many cases have a robust, testable effect on subjects’ subsequent mental processing.21

Similarly, blindsight is a condition in which damage to primary

visual cortex removes the possibility of conscious visual sensation for a particular area of the visual field. But despite their inability to register consciously any relevant mental qualities, blindsight subjects guess with astonishingly accuracy about perceptible properties, such as color, shape, orientation, and motion. In these cases, individuals with no relevant conscious mental qualities nonetheless evidently sense these distinctively perceptible properties. Experimental work shows that qualitative character sometimes occurs without consciousness, and this in turn constrains how we should think and theorize about qualitative consciousness.

Those who hold that traditional philosophical views must instead constrain the interpretation of empirical findings would dispute this conclusion. They would argue that these priming effects and accurate guessing need not be due to the occurrence of mental qualities. Qualitative character, they would urge, is necessarily conscious; so it must be some nonmental, nonqualitative occurrence that is responsible for these psychological results and abilities in the absence of conscious awareness.

But this line of thought is unconvincing. The effects on subsequent mental processing pertain, in the visual case, to perceptible color, shape, and location. And these are exactly the properties that mental qualities enable us to access in the conscious case. The qualitative similarities and differences in perceptible color, shape, and location are, moreover, registered psychologically, since they play a distinctive, dramatic role in our psychological lives. The best explanation of such registration of qualitative differences is that it is due to differences among the mental qualities that occur without being conscious. To deny that the registration of such qualitative differences involves mental qualities is arbitrary and ad hoc, the denial of an obvious conclusion simply to rescue an outmoded traditional theory.

What are the most important open problems in contemporary philosophy of mind? What are the most promising prospects?

Perhaps the most widely discussed topic in philosophy of mind over the last roughly 30 years is the nature of intentional content. But there has been little agreement among those who write about intentional content as to how to explain its occurrence. Some theorists espouse a holist approach, along lines of a functionalist or conceptual-role theory, whereas others adopt an atomistic theory
of such content. Some explain content by appeal to something like biological function, whereas others see content as hinging on a causal or counterfactual connection with the things the content represents. Some see content as in part, at least, a matter of the external physical or social environment, whereas others insist that content supervenes solely on physical states of the organism. Given the lack of convergence on these and related issues, it is natural to see the nature of intentional content as one of the most pressing open problems in contemporary philosophy of mind.

But there is another issue that perhaps cuts more deeply. It is common in philosophy of mind to adapt Kant’s well-known distinction between concepts and intuitions to draw a firm distinction between intentional content and qualitative character. Intentional content is characteristic of thoughts, desires, and other purely intentional states, whereas qualitative character distinguishes bodily and perceptual sensations.

This is arguably a good move. It is likely that whatever theoretical treatment succeeds with intentional content, a different account altogether will be needed to explain the nature of mental qualities. For one thing, intentional content occurs in sentence-sized units. Even if sentence-sized content is built up from sub-sentential pieces, such as concepts, qualitative character never occurs in such sentence-sized units. For another, it’s likely that an explanation of mental qualities will have to appeal in part to the differences among sensory modalities, whereas no parallel appeal seems needed, or even possible, in explaining intentional content.

But whatever differences separate intentional content from qualitative character, these mental properties sometimes occur together, as in perceiving and, arguably, the emotions. Perceiving something involves both sensing and conceptualizing that thing, and the conceptual content and mental qualities go together in an especially intimate way. Similarly, if I’m angry, I’m angry that something is so, but my anger also typically has a characteristic qualitative feel, involving bodily sensations.

So if we follow the modern philosophical tradition in drawing a firm divide between intentional content and qualitative character, we must explain how the two fit together in perceiving and the emotions. This is arguably a somewhat neglected question. Some theorists have hoped that so-called nonconceptual content may

\[ ^{22}\text{Gareth Evans, } \textit{Varieties of Reference}, \text{ ed. John McDowell, Oxford: Clarendon Press, 1982; Christopher Peacocke, } \textit{Sense and Content}, \text{ Oxford:} \]
help to build something of a bridge between the two as they occur in the perceptual context. But it’s unlikely that nonconceptual content will do the job. When we see a rabbit, as such, the qualitative character that pertains to the rabbit must somehow operate in tandem with with the concept of a rabbit. And it’s unclear how invoking nonconceptual content can help us understand how this happen.

Representationalist views, such as that of Fred Dretske and Gilbert Harman, promise an informational continuity between the conceptual and qualitative in perception. But it is unlikely that representationalist approaches can do justice to the distinctively qualitative character of perceiving. Champions of phenomenal intentionality, who urge that there is something it’s like for one to be in intentional states, sometimes regard intentionality as altogether inseparable from phenomenal character. But it’s doubtful that intentional content always occurs in connection with conscious qualitative character. And if not, the problem of linking qualitative character with intentional content remains.

Philosophical thinking about the mind has sometimes denied any theoretical discontinuity between the qualitative and the conceptual. Aristotle saw thinking as in the first instance an abstract form of perceiving, and therefore saw the content characteristic of thinking as continuous with the qualitative character of perceiving. And it’s arguable that folk attitudes about thinking and sensing more closely resemble Aristotle’s in this respect than Kant’s. Moreover, the distinction between the conceptual and the sensory is often elided or drawn rather casually in cognitive psychology, suggesting that cognitive psychologists may, like the folk, see continuity instead of a firm divide.

But few today would champion a theoretical account of thinking that is parallel to that of sensing, such as Aristotle’s theory of sens-

---

sible and intelligible forms. And if different theoretical treatments are needed for intentional content and qualitative character, we also need an account of how the two figure together in perceiving and the emotions. Like the very nature of intentional content, itself, its connection to the qualitative may well be a major open problem in philosophy of mind.