

## CONTENT INSIDE OUT

MARK RICHARD

*Philosophy Department, Harvard University*

Joseph Mendola's *Anti-Externalism*<sup>1</sup> defends internalism,

... the view that all the conditions that constitute a person's thoughts and sensations are internal to their skin and contemporaneous, inside and now ... [certain] mental states—for instance my beliefs, desires, and sensations—would exist and retain their characteristic contents even if there were nothing outside me and no past, as long as what is currently inside my skin were unchanged. (1)

Its first section sketches an internalist response to the arguments of Putnam, Kripke, and Burge that motivate externalism. The second section sketches an account of the content of qualitative states, what Mendola calls 'qualia empiricism'. And the last section responds to Wittgensteinian and interpretationist objections that content must be public in ways that internalism denies. I will concentrate on the arguments in section I.

Most of Mendola's discussion is directed toward characterizing the *content* of our beliefs and other psychological states. But what is content? Most accounts presuppose that the contents of propositional attitudes determine their truth conditions and that their ascription explains and rationalizes behavior. Mendola certainly thinks that content explains and rationalizes; he would add that belief content must in some sense capture "what it is like" to have a belief.

Does internalist belief content determine truth? An internalist *could* be revisionary about this. He could say that most of our beliefs aren't true, or that the notion of truth is the wrong notion to employ in discussing our beliefs. Mendola flirts with such revisionism. But he allows that many externalist intuitions about truth conditions are sound, and claims that on the whole internalism can account for those intuitions. He suggests, for example, that to a first approximation, the content of a name N is a rigidification of a description like 'the person named "N"' (37). The content of 'water', to a first approximation, is a rigidification of 'the stuff that best satisfies the cluster C of descriptions I associate with "water"', the cluster C containing descriptions of properties I take to be analytically connected to 'water', or to be a priori true of it (41).<sup>2</sup>

Such proposals are familiar, though Mendola gives them some new twists. Objections to such proposals are equally familiar: they mischaracterize

1. Oxford University Press 2008. All references indicated parenthetically.

2. The content of a name will include nonlinguistic properties; the content of a predicate will not refer to the cluster, but to the properties therein. Nothing I will say turns on this.

content, making it metalinguistic; they are at variance with our intuitions about when people share beliefs; they multiply contents in an ad hoc way in order to say that states that are same ‘from the inside’ have different content. What is new is the way Mendola responds to such objections: He says there are perfectly parallel objections to externalist proposals about content. Thus, the objections carry little or no dialectical weight against internalism. Furthermore, Mendola argues, the considerations that Burge *et al.* invoke in many of their arguments rest, at the end of the day, on

relatively arbitrary and quirky features of our notions of truth and reference, . . . [that are] unimportant to any proper account of what we think (88).

Burt and his twin’s thoughts may differ in truth conditions or even truth. But such rococo social embellishments of sentential content are no more relevant to characterizing what someone believes than is his silverware pattern or the month he was born.

The three sections that follow discuss Mendola’s proposal about the content of names and kind terms, his dialectical argument, and the alleged irrelevance of truth to mental content.

(1.) According to Mendola, your average adult has an ‘internally determinate understanding of what it is to be called by a name’, one that is ‘fixed by our internally constituted dispositions to respond in various ways to various situations’ (34). When Kripke outlines the “causal theory of reference” in *Naming and Necessity*, Mendola says, he ‘depends on our shared understanding’ of naming ‘to deliver [the] intuitions [that validate his account]’. This understanding informs and provides the content of our uses of verbs like ‘names’, ‘calls’, and ‘refers’.<sup>3</sup> Thus my use of ‘calls’ has an internally determined content that links my mental name tokens with their referents. This means I correctly describe whatever I refer to with a token mental name *n* with ‘the thing I call “n”’. I can rigidify this description as ‘the thing it is actually the case that I call “n”’.<sup>4</sup> This is the content of my use the name ‘n’.

What are these dispositions and how do they determine when it’s correct to say that a name *n* names an object *o*? We are pretty much left on our own

3. We gain (imperfect) conscious access to this understanding by thinking about the sorts of cases Kripke and others discuss.
4. This proposal has to navigate around the objection that it doesn’t allow a name to refer to an object at a world at which it doesn’t exist. Mendola invokes Kaplan’s *dthat* operator to help us understand his proposal. Kaplan, in the *Afterthoughts* to *Demonstratives*, allows that there are two ways of understanding ‘*dthat*’. On one understanding—not the one Mendola wants—it is a ‘demonstrative surrogate’ which, when combined with a description *d* yields a term whose content at a context *c* is *the object* that *d* denotes relative to *c* and *c<sub>w</sub>*. On the other understanding it is a rigidifying operator: “The complete *dthat* term [on this understanding] would then be a rigid description which induces a complex “representation” of the referent into the content.” On this understanding, the term can only pick out at *w* objects which are *in the domain of w*. Perhaps Mendola thinks quantifying over merely possible objects is consistent with the physicalism he repeatedly invokes. If so, he can say that *necessarily*, *for any x*, *necessarily x exists*, and he has no problem with names rigidly designating. Otherwise, he has to deny that such things as *Feynman = Feynman* are necessary truths, or tell complicated stories about what necessary truth is.

to answer this question. Clearly included are dispositions to respond to hypothetical cases. Perhaps included are dispositions to affirm on reflection various theories. Also included are dispositions to correct ourselves and accept correction (298). My guess is that Mendola thinks that these dispositions determine a reference relation roughly so: We are to idealize, abstracting from limits of time, fatigue, and so forth. Name  $n$  refers to object  $o$  if, were we to consider more and more cases and proposals, we would eventually settle on an account of the naming relation from which it follows (when the account is coupled with nonsemantic facts) that  $n$  names  $o$ .

It is debatable that everyone's dispositions would eventually converge on cases like Gödel/Schmidt or Madagascar. People are, after all, driven by theoretical commitments, passion, and stupidity to false conclusions. There are very bright people—many, alas, in philosophy classes—who just don't share the relevant intuitions, for whom nothing short of brain cleansing seems likely to change matters. But it is hard to believe that these people speak a language with a very different semantics from that of our own, since it is so plausible to ascribe them the intention to mean by their words what those around them mean. They have, that is, an intention that what certain of their mental states mean is to be determined by what's going on outside of them.

I don't see that such an intention is impossible, incoherent, or particularly unusual. It is perfectly capable of carrying the day, in the sense of determining the meaning of some of one's vocabulary. But if people have such intentions and those intentions are responsible for some of the meaning of some of their terms, then internalism is false. For then it is simply false that if things "outside" of these people were changed, that would not effect what their words mean. And I take it that whatever content is supposed to be, word meaning is a species of it, one that determines some belief content.

To my mind, that point is decisive. But suppose it is denied. There are *still* rather serious problems with Mendola's proposal.

Mendola's idea is that (i) our 'internally determinate dispositions' to apply a term and to respond to examples involving the term constitute, or are even identical with the content of that term; (ii) we have all accepted Saul Kripke as our Semantic Lord and Savior after meditating on chapter and verse of *Naming and Necessity*; so (iii) our internally determinate dispositions already determined that 'refers' has an extension that verifies the causal theory of reference. Grant (i) and (ii). (iii) doesn't follow. After all, dispositions, even if intrinsic to an object, need not be *essential* to it. Dispositions change, and one way that they change is through meditating on examples. If you go back and look at the literature from the 70s and 80s, it's pretty clear that people were very strongly disposed to *reject* Kripke's account. People *changed their minds*. **Suppose** that people's dispositions to react to examples determine what they mean by 'refers'. Then surely, before Kripke's work was circulated people typically *meant something different by 'refer' than they do now*. Indeed, they meant something so different that what they meant was better captured by something like the cluster theory commonly associated with Wittgenstein.

Myself, I think this result is absurd. Kripke was not converting the Wittgensteinians; he was correcting them. I suspect Mendola would agree with this last

comment, but would protest that the wayward Wittgensteineans always had an underlying disposition to accept the sort of criticism that Kripke gave of their first order dispositions to deploy an unrigidified cluster theory. I'm not sure why we are supposed to think this is so. More importantly, even if it's true, how does this show that the content of 'refers' is constant across the decades? Why think—if we accept that underlying dispositions by themselves determine what we mean—that the extension of 'refers' is the same today as it was 40 odd years ago? Forty years ago, the disposition to accept a nonrigidified, descriptive account of reference had the upper hand; today, it does not. Forty years ago, Dummett and others refused to accept criticisms of this view; today, they do. If dispositions determine content, and thus determine what we are talking about, why doesn't this imply—absurdly—that Kripke changed the semantics of our language?

(2.) I turn to Mendola's dialectical argument: The most common objections to internalism are just as much objections to externalism as to internalism. So they can't provide a basis for preferring externalism to internalism.

How so? The standard objections are ones to the way in which internalists try to account for the examples that motivate externalism. According to Mendola, every externalist response to Frege's puzzle and to puzzles about empty names "strains our intuition in ways reflected in" internalist responses to Kripke *et al.*

Mendola focuses on three objections. The first, **belief ascription** (BA), has it that when speakers of different languages have incomplete or faulty understandings of their words, an internalist like Mendola's must say that those speakers express different beliefs with sentences that are conventional translations of one another; but that's not how we ascribe attitudes, as attitude ascription "involves principles of attribution that run through . . . public language" (47).<sup>5</sup>

**Subject matter** (SM) is the objection that Mendola identifies the belief that Tom has arthritis with something along the lines of the belief that the person who is actually named 'Tom' has the disease that affects one's joints, is had by many elderly people and is what I in fact call 'arthritis'. But my belief is surely *not* about words.

Finally, Mendola postulates three kinds of narrow content. Very narrow contents are unshareable, being individuated in terms of token mental names or experiences. These "explain" how internal twins who refer to different things may have beliefs with different contents.<sup>6</sup> Somewhat narrow contents are 'just like' very narrow contents but are individuated by types, not tokens. These give a sense in which twins who refer to different things share content. And there is "content, which subtracts metalinguistic elements," supposedly

5. Mendola agrees that he must deny that monolingual Italian speakers who have 'different very incomplete or erroneous understandings' of the Italian 'artrite' from our own understanding of 'arthritis' share our beliefs about arthritis.

6. I must protest. If this "explanation" really explained the fact, it would commit us to saying that everyone, at every moment, has thoughts that no one could possibly share with them, thoughts which apparently cannot persist over time. What earthly explanatory purpose would such thoughts serve?

shared by members of different linguistic communities. To raise an eyebrow at the idea that we have three kinds of content, here is to express the **multiple content** (MC) objection.

All in all, I think these objections are pretty devastating. Imagine my surprise, then, to discover that my own views are open to the same objections, as are, Mendola claims, the views of Salmon, Soames, Crimmins, Perry, Evans, McDowell, Wiggins, Campbell, Millikan, Thau, Donnellan, Almog, Kripke, and Braun. Here, I address only what Mendola says about me. I think what he says of me is representative of the sort of thing he says about the others, and that they can give similar responses.<sup>7</sup>

The view of mine to which Mendola objects to is this.<sup>8</sup> (1) Both public language sentence uses and mental states have content. Content bearing mental states, at a high level of abstraction, are like sentences insofar as they have “parts”—*representations*, call them—that have contents which they contribute to the content of the whole. (2) Content is broadly Russellian—it is “made up” of individuals, their properties, and relations. (3) We can distinguish the bearer of content—a sentence use, a token mental state—from its content. Content bearers have properties that are of interest to us—syntactic and phonological ones for sentences, computational and causal ones in the case of belief states. These properties vary quite independently of content. (4) When I ascribe an attitude to you, I often convey facts about the properties of the content bearer of your attitude, as well as its content. I do this by presupposing a sort of “translation manual” that maps my words (in the sentences I use to ascribe your attitudes) onto your representations (in the mental state that realizes your attitude).

(5) Compositional semantics demands that we systematically assign truth conditions to sentences. If we take the semantic value of a complement clause to be the result of fusing its Russellian content with its syntax, we can give an elegant statement of the truth conditions of belief ascriptions. Think of the Russellian content of:

(a) Twain is happy  
as being something like

(b) <Twain, being happy>.

Take the semantic value of

(c) that Twain is happy  
to be something like

(d) <<Twain, ‘Twain’>, <being happy, ‘is happy’> >.

Think of the believer’s token beliefs in like fashion. Call these pairings of words or representations with Russellian content *articulated thoughts*. Spot me the idea of one articulated thought translating another. Spot me the idea of such a

7. Because of space limitations, I ignore empty names. But any externalist account of the content of beliefs that draws the distinction drawn in this section can give a perfectly adequate account of how we individuate beliefs with empty names.

8. It’s developed in detail in *Propositional Attitudes* (Cambridge University Press, 1990). *Meaning in Context*, Volume I (Oxford University Press, forthcoming) collects a number of papers in which the view is developed and defended.

translation obeying the rules for translation a belief ascriber presupposes—other than the fact that translation preserves Russellian content, the details don't matter today.<sup>9</sup> With these ideas in hand, we can say that

(B) *Odile believes that S* is true in a context provided that we can, while conforming to the context's translation manual, translate an articulated thought of Odile's with the semantic value of the phrase *that S*.

How is any of this liable to the trio of objections BA, SM, and SC? Mendola says that on this view,

The belief that Hesperus has risen and that Phosphorus has risen count as distinct, since they are mediated by different words . . . This view implies that people who don't share languages and hence words cannot strictly speaking share beliefs (59).

This exasperates me. I say that beliefs are contentful mental states, ones that involve representations with content that they contribute to the content of the belief state. **No one** would say that this implies that "strictly speaking" people cannot share beliefs. I say that the content of these states is Russellian, so that people who speak different languages have beliefs with the same content. I say that when we talk about other people's beliefs, we often convey information not only about the content of their beliefs but about properties of the underlying state, the content bearer. To say this is not to say, imply, implicate, or otherwise suggest that "strictly speaking" people with different languages cannot share beliefs. I flesh these ideas out by saying that belief ascription involves a kind of translation of the mental state of the believer with the words of the belief ascriber. This does not imply that we cannot share beliefs in the strictest sense possible. To show that these ideas can be given a compositional semantics I invoke machinery that makes explicit the sense in which attitude ascription involves translation. Why would anybody think that using this machinery so much as **suggests** that different language users can't think the same thing? People speak different languages. When we interpret another's language, there is *somewhere* in the background translation from his idiom to ours. Does anyone think that *this* shows that people cannot strictly speaking **say** the same thing?

Mendola continues:

. . . because belief reports are not intuitively about the words that mediate . . . beliefs, it [the objection just rehearsed] suggests the subject matter objection (59).

This exasperates me. First, SM is the objection that certain *beliefs* are not about words. How does the above account even *suggest* that beliefs **or** their ascriptions

9. Here is an example of the sort of translation rules I assume are presupposed in a context:

(R) in ascribing attitudes to Odile, 'Twain' as a name of Twain (i.e., <Twain, 'Twain'>) is only to represent Odile's uses in thought of 'Twain' as a name of 'Twain' (so, it can represent <Twain, 'Twain'>, but not, for example, <Twain, 'Clemens'>).

are about words? Belief ascriptions in English, as used in many contexts, say things true at possible worlds in which no one speaks English. The formal semantics makes words parts of the **semantic values** of complement clauses, simply to make explicit the mechanism I say underlies attitude ascription. That doesn't make the ascriptions about words in any intuitive sense. You can't just read what a sentence is about, in an intuitive sense, off the formal machinery the semanticist uses; you need to listen to the explanation the semanticist gives of the machinery.

Mendola continues:

Richard suggests in reply to worries like these that belief ascriptions . . . may report that individuals [have articulated thoughts] that are close enough matches to the [articulated thoughts] used in the ascriptions. So he accepts some violations of the principles of disquotation that govern common belief ascriptions . . . And because proper ascriptions are in this way relativized to contexts he in fact deploys multiple notions of content. From within different contexts, different belief ascriptions characterize the same belief. So it would seem that the belief ascription objection and the multiple-contents objection [are as telling against Richard as they are against internalism] (59).<sup>10</sup>

Aimed at Mendola, the BA objection is that if we take belief ascription to ascribe beliefs with internalist content, we assign truth condition *wildly* at odds with intuitions.<sup>11</sup> I am flabbergasted at the suggestion that I am vulnerable to this sort of objection. A good part of the *point* of the above account of attitude ascription is to show that one can take content to be Russellian while assigning truth conditions to attitude ascriptions that **conform** to pretheoretical intuitions about their truth. It would be one thing if Mendola pointed to cases where I don't validate intuitions. But he doesn't. BA is his problem, not mine.

Sometimes when Mendola speaks of 'the belief ascription objection', he seems to have in mind an objection like

- (6) Everyday attitude ascription is governed by principles along the lines of
  - (D) If someone understands and accepts a sentence, their use of it expresses something they believe.
  - (T) If S (when x uses it) expresses something x believes, and T translates S (in an everyday sense) into English, then *x believes that T* is true.<sup>12</sup>
- (7) Thus, if a view implies that *x believes that T* is false when x understands and accepts S and T translates S into English, that view is at odds with our ordinary practice of attitude ascription.

10. No citation of anything I have said accompanies this passage, so I can only guess what Mendola has in mind; furthermore, 'close enough match' does not map onto any notion I make use of.

11. To repeat, this is because (i) Odile and I won't share narrow content, but (ii) it's very often the case that all that is necessary for the truth of *Odile believes that S* is that she have a belief that shares wide content with *S*.

12. Of course, these principles need to be qualified in various ways—for example, principle (T) needs to be qualified because *S* or *T* may contain indexicals.

He thinks, since I allow that there are some arcane cases (like Pierre's) in which (D) or (T) fails, I am as vulnerable to such objections as he.

It is true that attitude ascription is governed by principles along the line of (D) and (T) **in the sense that** it is often the case that if Odile accepts S and T translates S, then we can truly say *Odile thinks that T*. This is a good rule of thumb, for often our interests are confined to getting across the broad content of another's thought. But the idea that we are slaves to (D) and (T) is *meshuggena*. Suppose Mary, who accepts 'Twain is an author', walks away from Twain because she looks at him and thinks 'he's no author'. If you ask 'Why is Mary walking away from Twain?' I'll say 'she doesn't realize that Twain's an author.'

What is close to the truth is something like:

- (8) When we are only interested in conveying the truth conditional content of a belief, D and T are pretty good guides to how to truly ascribe beliefs.

The sort of internalism that Mendola endorses seems to be pretty deeply inconsistent with this, at least given that it is committed to attitude ascription being ascription of narrow content. Narrow content on Mendola's story is just very, very idiosyncratic. But externalist accounts of content are not threatened by any of this.

I have gone on too long about this. So, as far as the idea that I, like Mendola, invoke multiple contents, I will just say this. It strikes me as a confusion to say that the words with which someone frames his belief are part of the *content* of that belief. I do not say that. It seems to me a confusion to say, with Mendola, that the *token* sensation associated with a perceptual belief is part of its *content*. These claims rob content of the ability to be something that can be shared by different language speakers, or which a person can believe at different times. I do not say these things. I do think that we sometimes manage to get across facts about the way that a belief is realized, when we ascribe it. But to say that is not to posit multiple types of content.

(3.) I'll close with some remarks about what Mendola calls the 'deep deference' objection to internalism, and about his claim that sometimes the truth conditions of our thoughts are "psychologically irrelevant."

The 'deep deference' objection is the observation that Mendola's response to Kripke requires that the notion of reference is part of many narrow contents; but isn't reference itself a 'deferential notion'? The important issue is not so much whether the notion of reference is deferential, but whether the semantics of the concept expressed by 'refers' is determined externally, whether through deference, or other social, or environmental relations. Mendola's response is that 'sometimes the identification of references and the sorting of beliefs into true and false is not in fact relevant to what people think ... Sometimes, facts about reference are semantically and psychologically unimportant' (79–81).

What does this mean, and how is it responsive to the objection? One way Mendola elaborates it is as follows:

... sometimes people's thoughts are vague or confused in such a way that it is perfectly clear what someone thinks ... but indeterminate whether it is true or false ... In such cases ... truth and reference don't matter to psychological semantics (80).

An example Mendola gives is due to Brian Loar:

[Loar writes] 'Suppose Bert is a full member of two English-speaking communities that differ linguistically in small ways of which he is unaware. The first is ours, where 'arthritis' means arthritis; but in the second 'arthritis' means tharthritis ... in explaining Bert the psychologically natural thing to say is that he has just one belief, one way of conceiving what is wrong with his ankles.' ... [Here] reference and truth [are] indeterminate while psychologically relevant content is not (82).

I take it the idea is that since Bert 'has just one belief' but its truth conditions can't be those of the arth- or tharthritis beliefs of others, we can just disregard the question of what the belief's truth conditions might be; in this case, the truth conditions of a belief do not help individuate its content.

I agree with what Loar says in the citation. But how is this supposed to show that reference and truth are sometimes irrelevant to belief content? I take it that to say that they are irrelevant is to say something like this: sometimes reference and truth are not to be employed in individuating content—in deciding whether two mental states realize the same belief. To show that, it would be necessary to provide a case in which two states realize the same belief, but have different referential properties. But pointing to a case like the one Loar points to does not seem to do this. If we go along with Loar, we say that Bert has a single belief expressed by 'John has arthritis' whose content is semantically indeterminate. Is this belief the same as that expressed with the sentence by those who are members of only one of the two communities? It seems obvious that it's not, as the other beliefs have determinate truth conditions, while Bert's does not. To say this is of course **not** to deny that there is a single, determinate belief state that Bert is in when he thinks 'John has arthritis'; in that sense, he has a single belief.

I don't see how the case shows that truth conditions are sometimes 'psychologically irrelevant'. It surely doesn't show that they are irrelevant to individuating beliefs. The fact that we may not be able to render a vague belief in our perfectly precise language doesn't show that the belief doesn't have semantic properties, or that these properties are not relevant to individuating it.

Here is what Mendola says is his "most important" point about the relation between truth and thought. According to Mendola, our notions of truth and reference involve

relatively arbitrary and quirky features ... [F]acts about reference ... for that reason ... [are] unimportant to any proper account of what we think ... Two individuals may share narrow contents, and be in circumstances where their thoughts have the same semantics and the same proper semantic

values, but in which the thoughts don't have the same references or truth-values, simply because of arbitrary quirks of convention (88–9).

Here is a representative sample of what Mendola has in mind:

There are little grounds to include Pluto as a planet and exclude [the outlier ice ball] UB313, but some question as to whether to include both or neither. That there are merely 8 solar planets . . . was recently decided by a vote . . . If someone who was unaware of this situation believed in May 2006 [before the vote] that there were 8 planets, then I claim that the eventual outcome of this arbitrary vote isn't relevant to the psychologically relevant content of their thought, even though it decided the truth of that thought (92).

If I understand Mendola, he thinks that an X who thought in May 2006 that there were eight planets had a true belief. He also thinks that if there were a colony of earthlings on Mars whose Astronomical Union had voted differently, and someone Y in the colony was a “narrow duplicate” of X, Y would have the same belief about Pluto as X, but, apparently, Y's belief would be false, not true. What belief is it, that the two share, but which is ‘true for X’, and ‘false for Y’? Apparently, the belief that Pluto is a planet. This sounds distressingly like a violation of Leibniz's Law.

What is motivating Mendola is the idea that the fine details of the extension of a predicate or a concept don't have much explanatory oomph when it comes to explaining things like the behavior of a believer. That's of course correct. But our notions of contentful states—of belief, knowledge, seeing, saying, and so forth—mix the semantic and psychological. There might be a useful notion *somewhat* like the notion of belief that abstracted from some or much of the semantic properties that our actual notion of belief has in virtue of its tie to the truth conditions of the sentences we use to ascribe it. If there is, there might be a notion of content that likewise abstracts therefrom. But this can't be what Mendola has in mind, since he so clearly means to be talking about the contents of beliefs and perceptual states.

The notion of the belief that S is the notion of a state that is true iff S. Since for certain S, Mendola seems willing to concede that the truth conditions of ‘S’ are determined externally, I just don't see how he can deny that the truth conditions of the belief that S are determined externally, too. Indeed, he does concede that, but he says that it is ‘psychologically irrelevant’. But if he concedes this much, he must concede that some beliefs would **not** be just the same if what was going on outside of us were to change in relevant ways. We would no longer have *those* beliefs, but ones that had different truth conditions. This is what the externalist says.