

WILEY

Styles of Mental Representation
Author(s): Daniel C. Dennett
Source: *Proceedings of the Aristotelian Society*, New Series, Vol. 83 (1982 - 1983), pp. 213-226
Published by: Wiley on behalf of The Aristotelian Society
Stable URL: http://www.jstor.org/stable/4545000
Accessed: 30/06/2014 07:08

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The Aristotelian Society and *Wiley* are collaborating with JSTOR to digitize, preserve and extend access to *Proceedings of the Aristotelian Society.*

http://www.jstor.org

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

XIII*—STYLES OF MENTAL REPRESENTATION

by Daniel C. Dennett

More than thirty years ago, in *The Concept of Mind*,[1] Gilbert Ryle attacked a vision of the mind he called the intellectualist myth. This is the idea that minds are composed mainly of such episodes as the thinking of private thoughts, the consultation of rules and recipes, the application of general truths to particular circumstances, and the subsequent deduction of implications about those particulars. This vision of the mind was no less preposterous for being traditional, in Ryle's opinion, and his attack was a vigorous—if not rigorous—combination of ridicule and *reductio*. Of course he knew perfectly well that there really are such phenomena as the memorization of rules, the later consultation of those rules, the self-conscious deduction of conclusions from premises and the like, but he thought that viewing these schoolroom activities as the model for all intelligent behaviour was getting matters just backwards. In fact, he argued, the very existence of such public human practices as the deliberate self-conscious consideration of stated maxims and the deduction of conclusions from sets of premises on blackboards *depends* on the agents in question having full blown sets of mental talents. If one attempted to explain these prior and more fundamental competences as based in turn upon yet another intellectual process, this time an *inner* process of calculation, involving looking up propositions, deducing conclusions from them and the like, one would be taking the first step on an infinite regress.

It was a powerful and rhetorically winning attack, but not so many years later it was roundly and bluntly rejected by those philosophers and other theorists who saw the hope of a cognitive psychology, or more broadly, a 'cognitive science', a theory of the mind that was very close in spirit to the view Ryle was lampooning. In fact the reigning ideology of cognitive science

---

* Meeting of the Aristotelian Society held at 5/7 Tavistock Place, London WC1, on Tuesday, 31 May, 1983 at 6.00 p.m.

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

sets itself so defiantly against Ryle that it might with some justice be called *intellectualist science*. It seems to be just the sort of thing Ryle claimed was wrongheaded. Cognitive science speaks openly and unabashedly of inner mental representations, and calculations and other operations performed on these inner representations. The mind, it proclaims, is a computing device, and as Jerry Fodor, a leading ideologue of cognitive science has put it, 'no computation without representation'.[2]

How did this new movement shrug off Ryle's attack so blithely? Part of the answer is that what Ryle was attacking was not one view but a mishmash of several views. Many of his slings and arrows can be so easily dodged by sophisticated cognitive scientists that perhaps their respect for the rest of his arsenal has been unduly diminished. Ryle danced quite a jig on the corpse of Cartesian dualism, for instance, but cognitive science is openly materialistic or physicalistic, in a sophisticated way Ryle apparently underestimated and maybe never even considered. So cognitive science has no worries about 'ghosts in the machine'. Its hypotheses are frankly mechanical, not, as Ryle would have it, 'paramechanical'—mysteriously pseudo-mechanical. And cognitive science has no allegiance to 'privileged access', another of Ryle's bugbears. Indeed most of the mental representations that it talks of are presumed to be utterly inaccessible to the consciousness of the agent. It is a doctrine of *unconscious* mental representations for the most part. So this is not the intellectualism of the inner Cartesian theater, with everything happening on the stage of consciousness; this is 'backstage' intellectualism—and in the view of these new theorists, so much the better for it.

But there was more to Ryle's attack than that. He was deeply suspicious—and I think for good reason—of *any* claims made on behalf of inner representations, whether they were supposed to be materially embodied, mechanistically manipulated, outside of consciousness or not, because he thought that in all their forms such postulations of inner representations were incoherent.

This feature of Ryle's view has certainly not gone unnoticed. Why then has it been so widely disregarded? Surely the main contribution to the conviction that Ryle must be wrong on this point is the growing influence of the computer metaphor in the field. The mind is like a computer, runs the slogan, and a

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

computer is indeed a mindless manipulator of internal explicit representations—no infinite regresses there, surely, for there the computers sit. Whatever is actual is possible, so internal-representation-manipulators are not the perpetual motion machines Ryle would have us think. Cognitive scientists have thus felt comfortable speaking about information processing systems and subsystems that utilize sorts of internal representation. These representations are like words and sentences, and like maps and pictures; they're just enough like these familiar representations for it to be appropriate to call them representations, but they are also just enough unlike words and pictures and so forth to evade Ryle's infinite regress machine. That is at any rate the common understanding, not that it has yet received its proper defense.

I propose to look more closely at the issue, for while I think that the computer metaphor, properly used, *can* liberate the theorist from Ryle's worries, it is often abused by ideologues of cognitive science. It is obvious, let us grant, that computers *somehow* represent things by using internal representations of those things. So it is *possible* (somehow) that brains represent things by using internal representations of those things. But if we are to move beyond that modest bit of metaphysical elbow room and actually understand *how* the brain might represent by analogy with computer representation, we had best be clear about just how it is computers represent. There are several ways, or styles, of computer representation, and only some of them are plausible models for ways, or styles, of mental representation in the brain.

When cognitive scientists have talked about representations, they typically have committed themselves to some view about syntax—to use the shorthand term. That is, they have supposed themselves to be talking about representations about which the following sorts of distinctions can be made: there are structural elements that are symbols; there are multiple *tokens* of *types* of representation (where these types are individuated syntactically, not semantically); there are rules of formation or composition rules—something like a grammar—so one can form big representations out of little representations; and the meaning of the larger representations is a function of the meanings of their parts.

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

This heavy commitment to a syntactical picture is often made willingly enough, but if it is made, it is solely on the basis of aprioristic reasoning, for so far as I can see, while there has been plenty of interesting speculation, there is almost no empirical evidence yet that tends to confirm any substantive hypothesis about the nature of this supposed syntax of mental representation.[3] This is not at all to deny that excellent and telling evidence has been obtained by cognitivist research, but rather to claim that this evidence to date has been evidence at the semantic level only. That is, it has been evidence about what information is being somehow relied upon by various cognitive processes, not evidence about how this reliance is effected.[4]

We cannot yet say, for instance, whether various pieces of information that have been implicated one way or another in various cognitive activities and competences are represented 'explicitly' or 'implicitly' in the human cognitive system. One reason we cannot say this is the confusion about how to use the terms 'explicit' and 'implicit'. People working in the field have meant quite different things by these terms, and here I will try to clarify the situation somewhat by offering some definitions—not of the terms as they are used, but of some terms as they should be used.

Let us say that information is represented *explicitly* in a system if and only if there actually exists in the functionally relevant place in the system a physically structured object, a *formula* or *string* or *tokening* of some members of a system (or 'language') of elements for which there is a semantics or interpretation, and a provision (a mechanism of some sort) for reading or parsing the formula. This definition of explicit representation is exigent, but still leaves room for a wide variety of representation systems. They need not be linear, sequential, sentence-like systems, but might, for instance, be 'map-reading systems' or 'diagram interpreters'.

Then let us have it that for information to be represented *implicitly*, we shall mean that it is *implied* logically by something that is stored explicitly. Now 'implicitly' defined thus does not mean what one might take it to mean: 'potentially explicit'. All the theorems of Euclid are implicit in the axioms and definitions. And if you have a mechanical Euclid-machine—if you have the axioms and definitions explicitly stored in the

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

machine and it can churn out theorems—then the theorems it churns out and hence renders explicit were implicit in the system all along—they were implied by the axioms. But so were lots of other theorems the machine may *not* be able to churn out. They are all implicit in the system, given its explicit representation of the axioms, but only a proper subset of them are potentially explicit. (This is not a point about 'incompleteness'; I have in mind the more mundane limitations of middle-sized pieces of hardware obeying the Einsteinian speed limit during relatively brief lifespans.)

It is an interesting question whether the concept of *potentially explicit* representations is of more use to cognitive science than the concept of (merely) implicit representations. Put another way, can some item of information that is *merely* implicit in some system ever be cited (to any explanatory effect) in a cognitive or intentional explanation of any event? There is a strong but tacit undercurrent of conviction, I think, to the effect that only by being rendered explicit, only by being actually *generated* by something like the Euclid-machine, can an item of information *play a role*. The idea, apparently, is that in order to have an effect, in order to throw its weight around, as it were, an item of information must weigh something, must have a physical embodiment, and what could that be but an explicit representation or expression of it? I suspect, on the contrary, that this is almost backwards. *Explicit* representations, by themselves (considered in isolation from the systems that can use them), may be admirably salient bits of the universe, off which to bounce photons or neurotransmitter molecules or marbles, but they are by themselves quite inert as information-bearers in the sense we need. They *become* information-bearers only when given roles in larger systems, at which time those of their features in virtue of which we call them explicit play problematic roles at best.

One might well say that implicit representation isn't representation at all; only explicit representation is representation. But then one should go on to note that if this is what we are to understand by 'representation', there are ways of holding or even sending information in a system that do not involve representing it. After all, a spy can send a message from A to B indirectly by sending explicit premises from which the intended message, the information-to-be-sent follows.[5] Another important

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

point to remember about implicit storage of information is that it has no upper bound. It needn't take more space to store more implicit information.

So *implicit* depends on *explicit*. But in the sense of 'tacit' I will use, it is the other way round: *explicit* depends on *tacit*. This is what Ryle was getting at when he claimed that explicitly proving things (on blackboards and so forth) depended on an agent's having a lot of *knowhow*, which could not itself be explained in terms of the explicit representation in the agent of any rules or recipes, because to be able to manipulate *those* rules and recipes there would have to be an inner agent with the knowhow to handle those explicit items—and that would lead to an infinite regress. At the bottom, Ryle saw, there has to be a system that *merely* has the knowhow. If it can be said to *represent* its knowhow at all, it must represent it not explicitly, and not implicitly—in the sense just defined—but tacitly. The knowhow has to be built into the system in some fashion that does not require it to be represented (explicitly) in the system. People often use the word 'implicit' to describe such information-holding; what they mean is what I mean by 'tacit'.

Ryle thought that the regress of representers had to stop somewhere, with systems having merely tacit knowhow. He was right about that. But he also thought it was obvious that whole people weren't composed of smaller subsystems that themselves represented anything explicitly, and he was wrong about that. It isn't *obvious*—as several decades of cognitive psychology show. It might still be true—or at any rate much closer to the truth than the current ideology supposes.

All these terms—'explicit', 'potentially explicit', 'implicit' and 'tacit'—are to be distinguished from 'conscious' and 'unconscious'. Thus what you consciously represent to yourself is at best indirect evidence of what might be explicitly represented in you unconsciously. So far as cognitive science is concerned, the important phenomena are the explicit unconscious mental representations. Thus when Chomsky talks about the explicit representation of one's grammar in one's head, he certainly doesn't mean the conscious representation of that grammar. It is presumed to be unconscious and utterly inaccessible to the subject. But he also means that it is not merely tacit in the operation or competence of the system, and it is also

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

not merely implicit in something 'more basic' that is explicit in the head. He means to take the hard line: the grammar is itself unconsciously but explicitly represented in the head.[6]

We can understand the hard line by comparing it to a paradigm case of *conscious* explicit rule-following. Consider bridge players, and their relation to a familiar rule of thumb in bridge:

Third Hand High!

There, right on this page, is an explicit representation of the rule. It is often explicitly represented (in just these three English words) in books and articles on bridge; one can also often hear tokens of it yelled across bridge tables. The rule enjoins the third of the four players to any trick to play his highest card in the suit led. (This is a tactical rule—not a rule of the game. You can play bridge and not know the rule; you can play *good* bridge and not 'know' the rule.)

Consider the most extreme case. This is the person who consciously and even self-consciously and explicitly consults the rule, who when his turn comes to play a card thinks to himself (perhaps he even moves his lips!): 'Let's see, now, I think there's a rule here. Yes. "Third Hand High!" Am I third hand? One-two-three, yes, I'm third. I'm supposed to play the highest card in the suit led. That would be my jack.'—and then he plays his jack. That would be a case of explicitly following a rule: getting the rule out of memory, putting it up in the 'workspace' of consciousness, examining it, checking to see if its conditions are met, and on seeing that they are, 'firing off' the activity. Now Ryle's claim was that anyone who thinks that this is a good model for human mentation of all sorts, from tying one's shoes to understanding a sentence in one's native language, is shall we say—benighted. But that's just what the cognitive scientists—at least some of them—think. We can see what their point is by comparing our first bridge player with some other familiar types.

Consider next the 'intuitive' bridge player. Let us suppose that our intuitive bridge player never heard the rule in his life, and the words of the rule have never occurred to him in any language he knows. He's never 'reflected' on it, so when he thinks about which card to play he certainly does not

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

consciously, explicitly follow the rule. That leaves him only the position that most of us are in with respect to the rules of our native language. It doesn't follow that he's not *un*consciously following some explicit version of that rule. Let us suppose in any case that his dispositions to card-playing behaviour (card-*choosing* behaviour—we are to ignore the accompanying frowns, delays, and *sotto voce* mumblings) are indistinguishable from those of the first bridge player. The hard line hypothesis is that the backstage-processing account of this player's 'intuitive' card-choosing bears a very strong resemblance to the 'introspective' account our first player might give.

Finally, look at a third case: a player who combines the features of the first two—and hence is a much better bridge player than our first, and maybe better than our second as well. This person knows the rule, but is smart enough to realize that there are exceptions to it, that it shouldn't be slavishly followed. He can think about the rationale of the rule, and about whether this is a good opportunity to apply the rule.

This third player would be a much worse model for cognitive psychology to adopt (and I suspect Ryle had this sort of rule-consulter in mind when he dismissed the prospect), because he lacks a very important property the first player had: stupidity. A systematically important feature revealed by our first player's rule-following is the possibility of storing and 'acting on' something without really understanding it. It is the *worst sort* of classroom activity, the rote memorization, that supplies the best model for cognitive science, because it has the nice feature of decoupling memory from understanding. Memory of this sort is just brute storage (like a singer memorizing the lyrics of a Russian song without having the faintest idea what they mean).

This is just what we want it seems, if we're trying to explain understanding in terms of storage and manipulation. For we want our storers and manipulators to be stupider than our understander (of which they are proper parts); otherwise we'll get into a Rylean regress. The storers and manipulators must indeed have *some* knowhow—even our first bridge player knows enough to know how to apply the rule, and this is not nothing; just think of the bridge players who can't seem to get this simple rule into their thick heads. This knowhow in turn might be merely tacit, or based on some further internal rule-following

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

process of even narrower horizons and greater stupidity.[7] The possibility in principle of terminating this regress in a finite number of steps is a fundamental guiding insight of cognitive science. But will the regress actually be used? It must terminate in the end with merely tacit knowhow, but could Ryle turn out to be right after all about the 'size' of the largest merely tacit knowers: whole people? Could virtually all the *backstage* knowhow be merely tacit in the organization of the system? How powerful can a system of tacit 'representation' be?

This is a hard question to answer, in part because the critical term, 'tacit', still has been given only an impressionistic, ostensive definition. We haven't really pinned down what it should mean. Consider a benchmark question: does a pocket calculator represent the 'truths of arithmetic' explicitly, implicitly, or tacitly? A tiny hand calculator *gives one access to* a virtual infinity of arithmetical facts, but in what sense are *any* arithmetical facts 'stored' in it? If one looks closely at the hardware one finds no numerical propositions written in code in its interior. The only obvious explicit representation of numbers is either printed on the input buttons or, during output, displayed in liquid crystal letters in the little window.

But surely there is further explicit representation hidden from the user? Consider what happens when one gives the calculator the problem: $6 \times 7 = ?$ The calculator does the multiplication by swiftly adding (in binary notation) 7+7+7+7+7+7, and during the actual process it holds in its accumulator or buffer the interim totals of each successive sum. Thus we can clearly distinguish the process it goes through when multiplying 6×7 from the process it goes through in multiplying 7×6. In the former case the interim results are 14, 21, 28, 35 while in the latter case they are 12, 18, 24, 30, 36. Surely this is explicit and systematic representation of *numbers*, but where does the calculator represent any true arithmetical *propositions*? Its inner machinery is so arranged that it has the fancy dispositional property of *answering arithmetical questions correctly*. It does this without ever *looking up* any arithmetical facts or rules of operation stored in it. It was designed, of course, by engineers who knew the truths of arithmetic and the rules of arithmetical calculation, and who saw to it that the device would operate so as to 'honor' all those truths and rules. So the calculator is a

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

device with the dispositional competence to produce explicit answers to explicit questions (so these truths are *potentially explicit* in it), but it does this without relying on any explicit representations within it—except the representations of the questions and answers that occur at its input and output edges and a variety of interim results. The truths of arithmetic potentially explicit in it are thus not *implicit* in it, for there are no explicit truths in it of which these are the implications.

Outlandish as it may seem at first, it is worth comparing this view of the pocket calculator with Ryle's view of human beings. Ryle is the foe of internal representation, certainly, but he has the good sense to acknowledge what might be called *peripheral* explicit representation—at the input and output boundaries of people (!), as instanced by such familiar Rylean categories as *sotto voce* rehearsings, talking to oneself without moving one's lips, reminding oneself of interim results before getting on with the task at hand. Ryle's view, I take it, is that just as there is no deeper, covert-but-explicit representation of anything in the pocket calculator in virtue of which these conceded representation are manipulated, so there is no covert-but-explicit representation in us, in virtue of which we are enabled to say and think the things we do.

An interesting feature of the design process that yields such things as hand calculators as products is that the designers typically *begin* with a perfectly explicit specification of truths to be honored and rules to be followed. They eventually succeed in creating a device that 'obeys' the rules and 'honors' the truths without itself representing them explicitly at all. Does this process have an analogue in human mental development? Occasionally human beings learn skills that are *first* governed (or so 'introspection' strongly assures us) by quite explicit consultation of explicitly rehearsed rules—as with our first bridge player—but these skills, with practice, eventually become *somehow* 'automatized': the tennis player no longer mutters directions to herself as she prepares for a backhand stroke, the newly 'fluent' speaker of a second language no longer *consciously* checks to make sure the adjective agrees in gender with the noun it modifies.[8] What is going on in these probably related phenomena? One possibility, most vividly sketched by Fodor,[9], is that such automatization is a matter of merely *hiding*

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

the explicit-recipe-following beneath normal conscious access. (His example is tying one's shoes; he supposes that submerged beneath conscious access is an explicit recipe, called 'How to Tie Your Shoes'; it is retrieved by an equally hidden recipe-reading-and-following subsystem, which governs the actual shoe-tying process in the manner of our first bridge player.) Another possibility, suggested by the example of the calculator design process, is that practicing is somehow an analogous process of partial self-design: it yields, like its analog, a 'device' that 'obeys' the rules without consulting any expression of them.

This latter possibility may seem probable only for systems whose tasks—and hence whose rules of operation—are as static and unchanging as those 'hard-wired' into a calculator. But it is entirely possible to design systems that can change mode, switching from 'following' one set of tacit rules to 'following' another. An automatic elevator, for instance, can be made to follow one set of rules from nine to five on weekdays, and a different set of rules during off-peak and weekend hours. And of course it can be made to do this without either set of rules being explicitly represented in it; all it needs is a clock-controlled switch to take it back and forth between two different control systems each of which tacitly represents a set of rules of operation. This gives us a simple example of what we may call *transient tacit representation.* All the rules are tacitly represented all the time, but depending on the state of the system, only one set of rules is tacitly represented *as being followed* at any time. Or one might equally well say that the whole system tacitly and permanently represents the rule 'Follow rule system R on weekdays from nine to five and rule system R′ at other times', and the state of the system at any time transiently and tacitly represents the time of the week—a way of providing a *vehicle*—but not a vehicle of *explicit* representation—for such indexical propositions as 'Now it is weekend time'.

We can imagine similar systems in animals. We can imagine, for instance, an animal that is both aquatic and terrestrial, and when it is on the land it obeys one set of rules, and when it is in the water it obeys another. Simply *getting wet* could be the trigger for changing internal state from one set of rules to the other.

So far the systems I have described are switched from one state to another by a simple switch—an uncomplicated 'transducer'

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

keying on some feature of the environment (or the mere passage of time if the transducer is a clock), but we could have more elaborate switching machinery, so that which system of rules was transiently tacitly represented depended on complex distal features of the environment. If elaborate perceptual analysis machinery drives the system into its various states, then there is no apparent limit to the specificity or complexity of the state of the world, for instance, that could be tacitly represented by the current state of such a system. Not just 'Now it is weekend time' but 'Now I'm in grandmother's house' or 'Now there is a distinct danger of being attacked by a predator approaching from the north-north-east'. Such systems of tacit representation would need no *terms* to be 'translated by' the various terms in the theorists' attempts to capture the information tacitly represented by such states (such as the attempts appearing between quotation marks in the previous sentence). For the whole point of tacit representation is that it is tacit! States of such a system get their semantic properties directly and only from their globally defined functional roles.

But as the number of possible different states (each with its distinctive set of tacitly represented rules) grows, as a system becomes versatile enough to be driven into a great many significantly different control states, this profligacy demands that the design rely in one way or another on economies achieved via multiple use of resources. For instance, where the different states are variations on a theme (involving only minor changes in the tacitly followed rules) it becomes useful—virtually mandatory—to design the whole system to change states not by switching control from one physically distinct subsystem to another near-duplicate subsystem, but by changing some one or more features of the current subsystem, leaving the rest intact—changing states by *editing* and *revising*, one might say, instead of by *discarding* and *replacing*. Economies of this sort require systematicity; the loci into which substitutions can be made have to have fixed ways of changing their functions as a function of the identity of the substituends. The whole system thus does begin to look *somewhat* like a language, with counterparts for all the syntactical features mentioned at the outset.

Should we say that such a system finally emerges as a truly

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

explicit system of internal representation? There will certainly be uses for that view of such systems. But it is important to note that the 'syntactical' elements of such systems are to be viewed first as having an entirely internal semantics—'referring' to memory addresses, internal operations, other states of the system, and so forth, not to things and events in the outer world.[10] If these internal state-components are cunningly interanimated, the states achievable by them can bear delicate informational relations to events and things in the outer world, but then, insofar as the states of such systems can be interpreted as having external semantic properties, they obtain their semantic properties in just the same way—for just the same reasons—as the merely tacitly representing states. It is only the globally defined role of such a state (the role that is characterized in terms of the rules of operation the whole system 'follows' when it goes into that state) that fixes its informational or external semantic properties.

In an extended usage, then, it might be profitable to grant indirect external semantic properties to some of the elements of such a system of states. So it *might* be found that the state of someone responsible for his believing that snow is white has components identifiable as the 'snow'-component and the 'white'-component. But such 'sentences in the language of thought', if we decide it is wise to call them that, are in striking contrast to the sentences of natural languages. Given what the English sentences 'snow is white' and 'snow is cold' and 'milk is white' mean, we can say what 'milk is cold' means (whether or not some speaker or audience realizes it); but given the counterpart cases in Mentalese, we won't be able to *tell* what the 'milk is cold' state means—it may not mean anything at all—until we've determined its global role.

My goal in this paper has been simply to explore—and perhaps improve our view of—some open, empirically researchable territory. Ryle said, aprioristically, that we *couldn't* be mental-representation-manipulators; Fodor and others have said, aprioristically, that we *must* be. Some of the details of the computer metaphor suggest what we *might* be, and in so doing may eventually shed some light on what we are.[11]

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions

NOTES

[1] Gilbert Ryle, *The Concept of Mind*, Hutchinson, 1949.

[2] Jerry Fodor, *The Language of Thought*, Crowell and Harvester, 1975, p. 34.

[3] Edward Stabler, 'How are Grammars Represented?' (forthcoming in *Behavioral and Brain Sciences*) develops similar points in more technical detail.

[4] I illustrate my understanding of the deservedly controversial distinction between semantics and syntax via an analogy drawn from the world of espionage in a conversation with Jonathan Miller in his 'States of Mind' television series (BBC). An edited version appears in the book accompanying the series, Jonathan Miller, ed., *States of Mind*, BBC Publications, forthcoming.

[5] This possibility is ingeniously exploited by Jorge Luis Borges in his classic short story, 'The Garden of Forking Paths', in Borges' collection, *Labyrinths: Selected Stories and Other Writings*, edited by Donald A. Yates and James E. Irby, 1962, New Directions, New York.

[6] See, e.g., Noam Chomsky, *Rules and Representations*, Columbia Univ. Press, 1980. See also the excerpts, with commentary by many authors, in *Behavioral and Brain Sciences*, 1980, Vol. 3, p. 1-61. See also Stabler, *op. cit.*, for a careful analysis and rebuttal of Chomsky's position.

[7] For elaboration, see my 'Why the Law of Effect Will Not Go Away' and 'Artificial Intelligence as Philosophy and as Psychology', both in *Brainstorms*, 1978, Bradford and Harvester.

[8] One of the most fascinating cases of this move to automaticity is the self-training of a calculating prodigy reported in I. M. L. Hunter's classic article, 'An Exceptional Talent for Calculative Thinking', *British Journal of Psychology*, 1962.

[9] Jerry Fodor, 'The Appeal to Tacit Knowledge in Psychological Explanation', *Journal of Philosophy*, 1968. Note that Fodor does *not* mean by 'tacit' what I mean here.

[10] Cf. Jerry Fodor, 'Tom Swift and his Procedural Grandmother', in his *Representations*, 1981, Bradford/MIT. Fodor sees clearly that the internal semantics of programming languages and their kin do not *in themselves* solve the problem of mental reference or intentionality. My claim here is that we have not yet been given compelling reasons for supposing that any of the 'syntactical' elements of internal states that do have external semantic properties will themselves admit of any straightforward external semantic interpretation.

[11] Due to severe limitations of both time and space, this paper must be a truncated version of a larger project; hence its abrupt close with so many loose ends still to be tied off.

This content downloaded from 80.189.235.52 on Mon, 30 Jun 2014 07:08:33 AM
All use subject to JSTOR Terms and Conditions