Forthcoming in *La Philosophie d'Oxford au 20ème Siècle : Approches du Sens Commun*, edited by B. Ambroise and S. Laugier, Hildesheim, Olms.

Ryle's Regress and the Philosophy of Cognitive Science

Julia Tanney

> In *Final Fantasy* [the first film to have an entirely computer-generated female lead] not only were the mouths of the real actors filmed and digitised, so that the virtual characters' lips could be synchronised with the sound, but every bit of walking, clambering and jumping was first done in a studio by stuntmen, so that the movements could be pasted wholesale on to digital skeletons…The directors have made people do all this, and then thrown the people away. As such some viewers might be forgiven for considering *Final Fantasy* to be a giant con trick…the cast of *Final Fantasy* are merely the pixellated ghosts of human movement. It is a melancholy species of shadow puppet theatre.
> (Steven Poole's review of *Final Fantasy: The Spirits Within*, *New Statesman*, 6 August 2001; 29)

Ryle's regress objection to the 'Intellectualist Legend' – that intelligent activity requires prior theoretical operations – was recognized by Fodor to present a powerful conceptual obstacle to the premise that underlies cognitivist approaches in the sciences. He attempts to thwart Ryle's argument in *The Language of Thought* by accusing him of confusing causal and conceptual explanations and claiming that, by analogy with computers, we can see how the appeal to explicit rules is halted at the first level since second-order rules are reducible to built-in causal processes.

In this paper I argue that Fodor's arguments against Ryle fail. In the first section, I suggest that Fodor's appeal to the 'empirical necessity' of theoretical operations misfires because he is the one who has misunderstood the difference between causal and conceptual questions. In section two, I argue that the fact that second-order rules are reducible to causal processes shows, not that the regress is halted, but that we cannot consider intelligent activity by analogy with computers. In section three, I discuss the philosophical motivation for introducing rules into an account of intelligent activity in the first place.

I.

In *The Concept of Mind*, Ryle spells out a vicious regress that confronts the rationalist or 'Intellectualist': performing some activity intelligently, rationally, or with reason cannot require prior theoretical operations such as deliberating, calculating, or

following rules since these are activities that are themselves performed intelligently, rationally, or with reason. So it must be possible to act intelligently, rationally, or with reason without prior deliberation, calculation, or rule-following. Otherwise one would need to suppose the existence of prior theoretical operations *ad infinitum*.

In the introduction to *The Language of Thought* Fodor remarks that it is difficult to think of an area of cognitive psychology in which the array of arguments in *The Concept of Mind* would not apply or in which Ryle does not apply them.[1]

> Indeed, it is perhaps Ryle's *central* point that 'Cartesian' (i.e., mentalistic) psychological theories treat what is really a *logical* relation between aspects of a single event as though it were a causal relation between pairs of distinct events. It is this tendency to give mechanistic answers to conceptual questions which, according to Ryle, leads the mentalist to orgies of regrettable hypostasis: i.e., to attempting to explain behaviour by reference to underlying psychological mechanisms. (5)

Fodor goes on to say:

> If this *is* a mistake, I'm in trouble. For it will be the pervasive assumption of my discussion that such explanations, however often they may prove to be empirically unsound are, in principle, methodologically impeccable. (5)

Having said this, Fodor spends a few pages addressing Ryle's arguments before developing his own particular version of the representational theory of mind: one that construes cognition as rule-governed computations over syntactically structured symbols or representations.[2] I will be considering these arguments as we go along.

Language-learning, perception, and rational choice are paradigmatic of abilities, according to Fodor, that admit of a cognitive/computational explanation. Fodor's argument for this begins with his presenting the following model as 'an overwhelmingly plausible' account of how at least some behaviour is decided upon.

> [a.] The agent finds himself in a certain situation (*S*).

[1] Fodor's recent *LOT 2* takes most of his earlier arguments for granted; it assumes that he has met Ryle's challenge. Although he does revisit the 'pragmatist' criticism of the appeal to explicit rules, his arguments tend to reiterate, but do not much expand upon, the ones he addresses in more detail in the earlier work I am considering here.

[2] This discussion of Ryle follows a more extensive treatment in his earlier work *Psychological Explanation*. See note 5, below.

> [b.] The agent believes that a certain set of behavioural options (*B1, B2,… Bn*) is available to him in *S*; i.e., given *S*, *B1* through *Bn* are the things the agent believes that he can do.
>
> [c.] The probable consequence of performing each of *B1* through *Bn* are predicted; i.e., the agent computes a set of hypotheticals of roughly the form if *B1* is performed in *S*, then, with a certain probability, *Ci*. Which such hypotheticals are computed and which probabilities are assigned will, of course, depend on what the organism (*sic*[3]) knows or believes about situations like *S*. (It will also depend upon other variables which are, from the point of view of the present model, merely noisy: time pressure, the amount of computation space available to the organism, etc.)
>
> [d.] A preference order is assigned to the consequences.
>
> [e.] The organism's choice of behaviour is determined as a function of the preferences and the probability assigned. (28)

In a footnote, Fodor admits that the model does not provide a 'logically necessary' condition for rational behaviour:

> …the conceptual story about what makes behaviour rational presumably requires a certain kind of correspondence between behaviour and belief but doesn't care about the character of the processes whereby that correspondence is effected. (29)

But a few sentences earlier in the text, conceding that the model is highly idealised, he argues that the most this concession shows is that 'the behaviours we produce aren't always in rational correspondence with the beliefs we hold.' Here he does seem to suggest that the deliberation model he describes provides a necessary condition for bringing about a rational correspondence between beliefs and action—at least for human beings: when the model is not adhered to, the behaviour is not rational. For he goes on to suggest that though angels may be rational by reflex, the model, 'or something like it', may be 'empirically necessary' for bringing about a rational correspondence between the beliefs and the behaviours of human and other 'sublunary' creatures.

> ..some agents are rational to some extent some of the time, and …when they are, and to the extent that they are, processes [like these] mediate the relation between what the agent believes and what he does. (29)

What started out as an intuitively plausible model of how at least some behaviour is decided upon has turned into a theory about what is necessary for rational action in

[3] Note that agents are the possessors of the abilities to begin with and then, without argument or remark, the abilities are attributed to organisms.

human beings. What occurs in some passages as an empirical hypothesis about underlying mechanisms occurs in others as a covert conceptual claim: it suggests what we correctly count as rational action in human beings. Fodor can no longer appeal to intuitions for this claim, whatever its status, so where is the argument? If the overt conceptual story, as he concedes, merely requires a correspondence between attitudes and behaviour and is silent about any processes involved, then angels may indeed be rational without deliberating, but so may human beings. Curiously, Fodor also admits (in the same footnote) that the production of behaviour that follows the deliberation model is not *sufficient* for the rationality of the behaviour, since the beliefs involved may be superstitious or the preferences perverse, or the computation grossly unsound.

Here we hit upon the matter that bothered Ryle in his arguments against the Intellectualist (and, indeed, Wittgenstein in his discussion of rules). When we consider the (relatively rare) circumstances in which we *do* deliberate as the model suggests, a multitude of things can go wrong. In predicting the probable consequences of a range of behaviours, assigning a preference order to them, determining his choice as a function of the preferences and the probability assigned, the agent is, we assume, acting rationally. Since the ability to deliberate about one's options is itself a higher-order rational ability, this higher-order ability cannot be required for rationality. It cannot be considered necessary because the alleged explanation (the higher-order ability) presupposes the very thing (rational action) it is supposed to explain. It is not that it is merely insufficient and that something else is needed: it is insufficient because the explanandum is presupposed in the explanans. That is why it cannot be considered necessary either. That is the gist of Ryle's infinite regress objection to the Intellectualist Legend.

Let us look again at Fodor's claim. If it is conceded that an agent need not deliberate as the model suggests in order for his performance to count as rational (which, for argument's sake, let us accept, lines up with his beliefs) then what does it mean to say that it is empirically necessary? What could it mean to say that human agents must deliberate as the model suggests when it is conceded that even if they were to, doing so is neither necessary nor sufficient for what we would justifiably count as rational action? In order to clarify this, Fodor refers us back to his earlier treatment of the difference between the kind of conceptual story that typically interests analytic

philosophers and the type of causal story that interests psychologists. The suggestion is that Ryle, in his impatience with those who attempt to offer causal answers to conceptual questions, is guilty of confusing conceptual with empirical claims, for the fact that there can be a conceptual answer to the question what makes [an] *x* [an] *F* does not rule out the possibility of a causal story as well.

> In general, suppose that *C* is a conceptually sufficient condition for having the property *P*, and suppose that some individual *a* does, in brute fact, satisfy *C*, so that '*Pa*' is a contingent statement true of *a*. Then: (a) it is normally pertinent to ask for a causal /mechanistic explanation of the fact that '*Pa*' is true; (b) such an explanation will normally constitute a (candidate) answer to the question: 'What makes *a* exhibit the property *P*?'; (c) referring to the fact that *a* satisfies *C* will normally *not* constitute a causal/mechanistic explanation of the fact that *a* exhibits the property P, although, (d) reference to the fact that *a* satisfies *C* may constitute a certain (different) kind of answer to 'What makes '*Pa*' true?'…(8)
>
> To put this point as generally as I know how, even if the behaviourists were right in supposing that logically necessary and sufficient conditions for behaviour being of a certain kind can be given (just) in terms of stimulus and response variables, that fact would not in the least prejudice the mentalist's claim that the *causation* of behaviour is determined by, and explicable in terms of, the organism's internal states. (8)

Of course it is true that conceptual claims involving causal concepts may invite the search for underlying mechanisms. If the concept of a heart is the concept of that which pumps blood throughout the organism, a search for that which functions as the pump in different creatures would be in order; the concept of poison invites an investigation as to whether a particular chemical, for example, causes illness or death; and the concept of disease allows, and then may change it contours to accommodate the results of, a search for viruses or bacteria responsible for its symptoms.

Suppose Fodor is right that the concept of rationality requires agents' beliefs and actions to line up in rational correspondence. An explanation of this could be as follows: We ascribe beliefs, etc. as part of the enterprise of making sense of an agent and we do this by attributing to him beliefs and other attitudes that line up with his actions, as described. Such is the goal of the enterprise; that is why actions and

beliefs (etc.) tend to line up.[4] So far there is nothing to suggest that a causal story explaining the rational line-up is to be found inside the individual or his brain.

In order to underline the point, consider Fodor's own discussion of the conceptual and causal explanations that may both be given for General Mill's claim that 'Wheaties is the breakfast of champions'. The conceptual explanation is that a non-negligible number of champions eat the wheat-flake cereal; a causal story perfectly consistent with this, Fodor suggests, may advert to the vitamins and special springiness of the flake's molecules.

The problem with this analogy is that the underlying causal story is only consistent with the conceptual one if the 'because' in 'because a non-negligible number of champions eat it' is a causal one. One of the main aims of Ryle's work is to suggest that mental expressions, among a large number of others, discharge their explanatory role other than by attributing a property (*a fortiori* a property with causal powers) to an object. Thus, as with the cases for which the Wheaties example is intended to be a model, here too, we have been given no reason for looking inside the composition of the cereal in our search for a mechanism that underlies the conceptual claim. For a non-negligible number of champions may eat Wheaties because they have been paid a lot of money to endorse the product; or because the champions they most admire do; or because they are given the breakfast free during sporting events, and so forth and so on. In these cases, even if the molecules in Wheaties flakes *are* especially springy, this will be completely irrelevant to an explanation why a large number of champions eat them.

It may be instructive to take a quick look at other powers – 'thick' powers we might say – that may be found on a high rung of what Ryle calls 'the ladder of sophistication' in order to see how silly it would be to search for so-called 'natural' properties or relations to explain them. A penny has a certain purchasing power; a football has score-enabling power; and a bishop in chess has the power to move diagonally across the board. If an anthropologist from Mars were to wonder what

[4] Compare: the explanation why so many games of chess end up in check-mate is that such is the goal of the game.

explains these respective powers he would be making a mistake to take the penny, football, or bishop back to his laboratory to search for their physical realizations. But this is obviously not to suggest that the respective powers are supernatural: he is just looking in the wrong place for the explanation he seeks. Nor is this to deny that the penny, the football, or the bishop have interesting physical properties, perhaps even properties that enable them to perform the roles they do. If a penny were not a certain size and weight it could not be used in transactions for gum balls; if the football were not shaped exactly so, a Beckham could not bend it and if a chess piece were made of ice it would not survive competition on a very hot day. Nor is this to deny that the penny is ripe for a mineralogical examination, a historical-numismatic examination, an investigation for fingerprints, for counterfeiting, and so on indefinitely. But none of these investigations are going to shed much light on its purchasing power. A credit card would be a more suitable candidate, for its magnetic strip and chip encode information that plays an even more complex role in the sort of economic transactions in which it trades. But however complex the credit card, we will not find an answer to how it gets its purchasing power without adverting to the banking institutions and economic environment in which it plays its role. The human brain is, by all accounts, the most complex and wonderful object in the world. But so far we are left with no reason to suppose that an answer to the question about why our beliefs line up with our actions is one that can be given by looking at second-order properties that supervene on matter that is to be found inside the agent's skull.

Like the alien anthropologist who takes the penny home to examine it, the cognitive psychologist may well be looking in the wrong place for an explanation of the agent's rational powers. To insist that the purposive elements of the mental phenomena under examination – and thus the normative dimension along which they can be assessed – may be partly explained by special content-bearing states supervenient upon the brain is as misguided as insisting that the normative dimension upon which a game of football can be assessed may be partly explained by special score-enabling properties (which, in a 'long-arm' version, reach out into the world) that somehow emerge from the molecules of the football.

I suggest there is no reason to believe that deliberation—or something like it—is necessary for rational action in human beings. Although we deliberate as the model

suggests some of the time Ryle's argument shows that the supposition that we must do so leads to a vicious regress. To anticipate what is to come: to show that a hypothesized mechanism which is alleged to play the role of, or go proxy for, deliberation does *not* lead to regress fails to meet the point of Ryle's objection, which is that deliberation of the kind suggested by the model is not, because it cannot, be conceptually required for rational action. It seems that Fodor, and not Ryle, is guilty of confusing conceptual with empirical claims.[5]

II.

The cognitive-psychological models Fodor describes (that of deliberation, perception, and hypothesis-testing for language-learning) presuppose that the agent has access to a representational system of considerable richness. This system, he argues, must share a number of characteristic features of real languages and it requires an ontological commitment to the processes and states ascribed by that model.

> …[D]eciding is a computational process; the act the agent performs is the consequence of computations defined over representations of possible actions. No representations, no computations. No computations, no model. (31)

When he considers our understanding of the language of representations over which the alleged computations are performed Fodor addresses the threat of regress. For if

[5] There are various confusions and conflations in Fodor's discussion of Ryle both in *The Language of Thought* and in his earlier *Psychological Explanation*. Part of the difficulty is that though Fodor (rightly) identifies Ryle as working in a Wittgensteinian spirit, he also (wrongly) identifies Ryle's programme with that of the logical behaviourist; i.e., one, who, according to Fodor, hopes to show that 'logically necessary and sufficient conditions for behaviour being of a certain kind can be given in terms of stimulus and response variables' (8). (Indeed, in this quotation, psychological behaviourism and logical behaviourism are conflated.) He also (wrongly) suggests that the appeal to criteria implies a 'criteriological' theory of meaning which he construes as a type of (property) cluster theory (5, fn).

Sometimes when philosophers invoke the distinction between logical, conceptual, metaphysical, and physical possibility or impossibility, they take the conceptual story to involve what we can consistently imagine to be the case in logically possible worlds that may be at a distance from ours, vis à vis the (or our) laws of nature. On this view, the use of predicates in the relevant domain of discourse tends to drop out of the picture. Ryle's conceptual cartography, by contrast, traces the inflections of meaning, or elasticities of significance of the term 'rational' and its cognates. By his lights it would not make sense to say that a philosopher could explore a concept (such as rationality) and ignore the way the relevant predicates are used. So another mistake Fodor makes in discussing Ryle involves mixing up Ryle's project in conceptual cartography in which concepts are construed as double abstractions from sentences performing their various jobs with those philosophers (Frege, early Russell and Moore) who suppose concepts to be, in the Platonic tradition, independently existing ideals (whose 'essences' can be unpacked by necessary and sufficient conditions for the concepts' application) to which our natural (imperfect) expressions merely gesture. Ryle's criticism of this approach began with his first articles and continued for the rest of his career.

understanding a predicate of English requires representing the extension of that predicate in a language that one already understands, then what about the predicates of the meta-language? If understanding an English predicate requires representing its extension in a language then understanding the predicate of a meta-language presumably requires representing its extension in some meta-meta-language, *ad infinitum*.

Here, Fodor allows that what we may count as 'understanding the predicate of a meta-language' may *not* require representing its extension: 'a sufficient condition [for this] might be just that one's use of the predicate is always in fact conformable to the truth rule' (65). To put this in Rylean terms (in a way Fodor does not) it seems that he is here allowing that understanding a predicate in a meta-language may be a matter of knowing this language *by wont*; that is, without the need to involve prior theoretical operations. But the obvious question, which Fodor acknowledges, is why he will not allow that learning English is a skill that is acquired through training and practice; a kind of knowledge-how or knowledge-by-wont that underlies any explicit higher-order practice of following a meaning-rule.

Fodor fails, after a long discussion, to meet this objection. He compares the language of thought (Mentalese) and a natural language such as English with the two languages used by computers: the input/output (programming) language and a machine language which the machine is *built* to use.

> Roughly, the machine language differs from the input/output language in that its formulae correspond directly to computationally relevant physical states and operations of the machine: The physics of the machine thus guarantees that the sequences of states and operations it runs through in the course of its computations respect the semantic constraints on formulae in its internal language. What takes the place of a truth definition for the machine language is simply the engineering principles which guarantee this correspondence. (66)

Although this kind of correspondence may also hold between physical states of the machine and formulae of the input/output code, this could be effected only by first translating them into the machine language. It is in this sense that the machine 'follows' the input/output code but only acts in accordance with its machine language. The idea seems to be that while genuine rule-following occurs in the programming

language, mere rule-conformity occurs in the machine language. (I shall be challenging this idea in the following section.)

This thought is developed further in a footnote in which Fodor considers what is involved in thinking of an organism as a computer. It is wrong-headed, he contends, to think of the nervous system as issuing commands which must be 'read' and translated into action or behaviour by a further system that intervenes between the efferent nerves and the effectors.

> On the contrary, what is required is just that the *causal* properties of such physical events as are interpreted as messages in the internal code must be compatible with the *linguistic* properties that the interpretation assigns to those events. Thus, if events of the physical type *P* are to be interpreted as commands to effector system *E*, then it better be the case that, *ceteris paribus*, occurrences of *P*-events are causally sufficient for activating *E*. (*Ceteris paribus* means: barring mechanical breakdown and barring events interpretable as overriding countercommands to *E*.) (74)

In the following footnote, Fodor responds to the critic who suggests that if one is willing to attribute regularities in the behaviour of organisms to rules that they unconsciously follow, one may as well attribute, say, Kepler's laws to planets in pursuit of their orbits. The point of the critic's remark, Fodor says, is to suggest 'that the only *real* case of rule following is conscious rule following by articulate organisms' (74). Other organisms merely act in accordance with rules; they do not follow them.

Fodor replies to this imagined critic that what distinguishes unconscious rule following from the rule conformity exhibited by planets is that 'a *representation of the rules [organisms] follow constitutes one of the causal determinants of their behaviour*' (74). Such is not the case for planets: 'At no point in a causal account of their turning does one advert to a structure which encodes Kepler's laws and causes them to turn' (74).

This is an ill-considered criticism of the position, which I shall have something to say about in the next section. In any case, Fodor's tactic is clear: if rules can be seen to be encoded or represented in the structure of the system and constitute one of the causal determinants of behaviour, the organism (or its subsystems) can be said to be following them. Indeed, this move is a now familiar one in cognitive psychology and

computer design for both are faced with what is known as the homunculus problem, which, it is generally accepted, can be avoided. Indeed, some understand this threat to be an articulation of Ryle's regress argument, so it will be worth a short digression to examine this problem in a bit more detail. [6]

The cognitive scientist, whose job is to explain some particular complex form of cognitive behaviour, will attempt to construe the behaviour as an information-processing task. The classical cognitive scientist will construe this, following Turing, as involving rule-governed computations over syntactically structured symbols. In order to say how the information-processing task is accomplished in the system in question, the cognitive psychologist uses a method called forward engineering or functional analysis. The aim of the analysis at this mid-level is to uncover the particular algorithm that is assumed to be 'used' by the system to complete the information-processing task that was defined by the higher-order computational analysis. In other words, since it is assumed that some effective procedure or other is implemented by the system to tackle the information-processing task identified at the computational level of analysis, the goal of forward engineering or functional analysis is to identify that procedure. Whereas the programmer or designer of some device will amass rules, heuristics, and brute procedures for producing the correct responses for some task, a (classical) cognitive scientist will assume that his subject is using some combination of such procedures and attempt to determine which ones, or 'recreate the program that is producing the behaviour' (Dawson, 109). This is the goal of functional analysis.

But if the goal is to analyse or divide complex functions into sub-functions, the danger is that they will never end.

> It would appear that functional analysis leads us directly into Ryle's Regress; each function that we propose to explain an earlier black box will rear its head as an ugly homunculus. We wind up with an infinite proliferation of unexplained functional terms. (Dawson, 156)

Forward engineering must be constrained so that each functional decomposition is decomposable into simpler functions. There has to be some principled reasons for supposing that functional decomposition must stop. And finally, it must be clear how

[6] The discussion below in the text follows Dawson.

the set of functions that exist at the end of forward engineering can be physically implemented. 'It is only with this third claim that a classical cognitive scientist can abolish the ghost from the machine.' (Dawson, 157)

Although Dennett expresses doubt about various aspects of Fodor's project, he explicitly praises this way of avoiding the homunculus problem in his review of *The Language of Thought.*

> …perhaps the *prima facie* absurd notion of self-understanding representations is an idea whose time has come, for what are the 'data structures' of computer science if not just that: representations that understand themselves? In a computer, a command to dig goes straight to the shovel, as it were, eliminating the comprehending and obeying middleman. Not *straight* to the shovel, of course, for a lot of sophisticated switching machinery is required to get the right command going to the right tools, and for some purposes it is illuminating to treat parts of this switching machinery as analogous to displaced shovellers, subcontractors and contractors. The beauty of it all, and its importance for psychology, is precisely that it promises to solve Hume's problem by giving us a model of vehicles of representation that function without exempt agents for whom they are ploys. Alternatively, one could insist that the very lack of exempt agents in computers to be the users of the putative representations shows that computers do not contain representations—real representations—at all, but unless one views this as a rather modest bit of lexicographical purism, one is in danger of discarding one of the most promising conceptual advances ever to fall into philosophers' hands. (102)

I shall be arguing that the suggestion that computers do not contain real representations is *not* a modest bit of lexicographical purism: indeed, the ontological status of Fodor's hypothesized language of thought is at stake (as is, incidentally, the realism of the Representational Theory of Mind, as well as that of the content-bearing mental states that are alleged to play a causal/explanatory role in functionalist philosophies of mind). Here, let us note that it is not at all clear how the digression on machine code language helps Fodor's argument. Fodor turned to computers when asked why we could not say 'we're just built that way' one level earlier — to explain how we learn the English predicate 'is a chair'— thus making the retreat from the natural language to the inner language unnecessary. After introducing the difference between programming language and machine code, Fodor returns to this question and replies that he agrees (presumably with Wittgenstein) that explanation has to stop somewhere, 'but it doesn't have to —and better not— stop *here*.'

> The question of how we learn 'is a chair' *does* arise precisely because English *is* learned. The question of how [a predicate of the meta-language] is learned does not arise precisely because, by hypothesis, the language in which [the predicate] is a formula is innate. (67)

This response simply begs the question; it assumes but offers no support for the claim that we need to posit a meta-language to explain how English is learned. The model can only be considered a 'hypothesis to the best explanation' if we have some reason to believe that an explanation of the kind proposed is required.

The same criticism can be raised against Fodor's language-learning hypothesis. Presumably Fodor would agree that the conceptual story for (what we correctly count as) learning a (first, natural) language does not require *explicit* hypothesis-testing. If we do on occasion test a hypothesis for the application of a predicate it is either without using language to do so or, if we do use language, a fair bit must have already been learned. After all, the same regress arguments can be marshalled to show that hypothesis-testing is not sufficient, because of the possibility of error. And, of course, the (natural) language in which explicit hypothesis-testing would be couched presupposes the very thing the hypothesis-testing is supposed to explain. So it cannot be necessary either. Thus there is no motivation for supposing the existence of an underlying mechanism that subserves the role of explicit hypothesis-testing, for it is has just been conceded that hypothesis-testing is neither necessary nor sufficient. Thus the suggestion that there is a non-explicit, innate, error-free, hypothesis-testing mechanism that does not threaten regress is completely unmotivated. In other words, to argue as Dennett does that the regress is not vicious (because the homunculi become 'stupider' and 'stupider' until they finally drop out of the picture) would at most block the argument from absurdity and allow that prior—mechanical—operations are not ruled out. But this would still yield no support at all for the covert conceptual claim that such *mechanical* operations are necessary to explain the ability of human beings to act rationally or understand language.

Fodor's response to the objection that the retreat from natural language to Mentalese is not necessary simply begs the question against those who think that learning English may involve training, practice, and the (innate, if you wish) capacity to catch

on to similarities that are natural in the circumstances.[7] Nothing in the fact that English is learned requires that this learning must be effected via mechanised hypothesis-formation which in turn would require a meta-language.

. . . . . .

I have argued that the point of Ryle's regress argument against the rationalist or Intellectualist was to show that prior theoretical operations cannot be considered necessary for rational action, language learning, or intelligent abilities in general, for the supposition leads to absurdity. The conclusion is that it must be possible to act rationally, understand a language, or act intelligently without participating in what are in effect higher-order activities requiring the same abilities.

Fodor seems to agree that prior theoretical operations are not conceptually required, but maintains they are 'empirically necessary' for human beings. But this is confused. If such higher-order activities were conceptually required, it would make sense to look for mechanisms that subserve these operations in human beings. But they are not, so it does not. Fodor thus seems to be making the covert conceptual claim that such operations are necessary; not for angels, but for human beings and other 'sublunary' creatures. But we have been given no reason for believing this claim, other than that it would make theoretical sense of work in the cognitive sciences. On the contrary, if the operations are, as Ryle envisaged, higher-order rational activities, then they are ones that assume the ability to be explained, and we have been given an overwhelming reason for rejecting it.

---

[7] Georges Rey (1997; 5) gives similar short shrift to the objection by parodying the attitude—crystallized in Wittgenstein's (1953, §1) comment that explanations have to come to an end somewhere:

> [O]ne could… [rationally]… never ask for explanations of anything at all: we *could* just say that it is a natural capacity of lightning to burn what it strikes. The question is whether we *can also* rationally ask for slightly deeper explanations, and, if we can, what those explanations might be.

But this also is too quick, since Wittgenstein's discussion of rules, like Ryle's battle with the Intellectualist, constitutes a sustained argument for the limits of rule-following explanations or those requiring prior theoretical operations.

In assimilating Ryle's regress with the homunculus problem in cognitive science and computer design, Fodor attempts to maintain that the regress is not a threat because it is not vicious. I have argued in response, that even if this assimilation were in order, it could at most block the argument from absurdity by retreating to some other kind of operation; one that is not a higher-order intellectual activity. But we have no argument that such an operation is necessary. Indeed, with this move to a different kind of (because now mechanised) operation, would are bereft of the original 'overwhelmingly plausible' account of how some behaviour is decided upon or, generalising to cover the case of hypothesis-testing, how some languages are learned, since *that* model has been replaced by one with very little indeed in common, as we shall see. The result is that we are left with no positive reason for believing that mechanised 'deliberation' or 'hypothesis-testing' is necessary for rational action and language-learning in any creature.

To be sure, some stretch of activity may be photographically or telephonically indistinguishable from that which is assessable as intelligent, rational, meaningful, etc. We still need a way of distinguishing these cases, and for this we need to appeal to norms, or the criteria upon which such activity may be assessed.[8] But Ryle is right that acting intelligently, etc. does not require that the agent casts a sideways glance at rules that embody these criteria. I have argued above that the supposition that these rules are hardwired is unmotivated; below I shall argue that mechanising the rules renders the norms which they embody unusable as a tool for assessment.

III.

Why would anyone think that prior deliberation—or something like it— is necessary for rational action? What motivates the rationalist, for example, to think that intelligence in general requires prior theoretical operations?

[8] The intentions and propositional attitudes of the agent are indeed relevant in distinguishing these situations. But to attribute an intention is not a matter of naming a particular kind of mental state, roughly in the brain, that takes propositional content as its object. It is to deploy a linguistic tool – several steps up the 'discourse' ladder of sophistication – that puts a marker down on just those assessment conditions that are relevant. See my (in progress, 2005, 2008, and 2009) for more on the explanatory function and logic of intention and other propositional attitude attributions.

The rationalist, or Intellectualist, might argue as follows: what distinguishes one bodily movement from another imperceptibly different movement (in one sense of 'imperceptible') is the fact that one is intentional, the result of agency, the result of reasons; the other not. On the rationalist construal, this difference amounts to the occurrence of a non-perceptual, mental feature (an 'intention', 'deliberation', 'reason') that plays a causal role issuing in behaviour. Similarly, what distinguishes the meaningfulness of a person's utterance from a phonetically similar sound made by, say, a parrot is the addition, in the first case, of a mental act of 'meaning'. What distinguishes acts of hearing from acts of listening is the mental accompaniment of 'understanding'. What distinguishes an inference from a mere string of statements is that the first was, but the second was not, made 'under the influence' of the rules of logic.

According to Ryle, this construal of what is required for rational action, language understanding, or intelligence may have derived from Plato's doctrine of the tripartite soul in which Intelligence is a special faculty by which internal acts of thinking— particularly the consideration of regulative propositions— are exercised. This construal of what is required seemed to be encouraged by the Enlightenment idea that mathematics and natural science set the standard as human accomplishments. Impressed by the analogy, the rationalist supposed that it was the capacity for theorizing that constitutes the intellectual excellence of man, together with

> the idea that the capacity to attain knowledge of truths was the defining property of a mind. Other human powers could be classed as mental only if they could be shown to be somehow piloted by the intellectual grasp of true propositions. To be rational was to be able to recognize truths and the connexions between them. To act rationally was, therefore, to have one's non-theoretical propensities controlled by one's apprehension of truths about the conduct of life. (15)

But, as we have seen, since theorizing is itself an intellectual ability it cannot be required for intelligence, on pain of regress. What of the suggestion that grasp of truth or of regulative propositions controls one's non-theoretical propensities in a merely mechanical way? This would at least seem to eliminate the threat of regress. The trouble with this suggestion is that two different kinds of explanation are conflated. The first, an explanation by appeal to standards or norms that are codified

in the performance-rules that govern some activity or practice; the second, an explanation by appeal to causal relations or to the laws of nature which are supposed to subsume them.

Rules have different explanatory functions from those of natural laws. The rules that govern rational action or logical inference, for example, help us to identify what is incorrect in certain performances and correct in others. Such explanations work by showing how the performance attains or fails to attain the standards or norms set for the practice. These are explanations relatively high up Ryle's 'sophistication ladder'. They are descriptions of the referee, theorist, or tactician.

If someone (as I once did, when very young) puts a cup of milk into a mixture for oatmeal cookies, and wonders why the consistency is wrong, pointing to the recipe that says 'add 1T milk' is one way of diagnosing the problem. If someone who cooks by intuition tosses into a mixture of sugar, butter, oatmeal, etc. a small amount of milk, and the cookies are good, then measuring the amount and writing it down as part of a recipe could be part of a useful second-order practice for those who want to cook but do not have the natural talent. Similarly, if someone is confused by a complicated argument, showing that the argument, once symbolized, is not valid would be one way of showing what is wrong with it. Indeed, explaining that it conforms to an accepted rule of inference is a way of showing that it is correct. But as Ryle says, the rules of logic, like the rules of etiquette or the rules of cooking are performance rules: 'only performances can be or fail to be in accordance with them' (Ryle 1946/2009, 241).

This, incidentally, is why the appeal to Kepler's laws mentioned earlier was inapt. The point of the critic's remark, recall, was to suggest that the only *real* case of rule following is conscious rule following by articulate organisms: other organisms merely act in accordance with rules; they do not follow them. Fodor replied that what distinguishes unconscious rule following from the rule conformity exhibited by planets is that 'a *representation of the rules [organisms] follow constitutes one of the causal determinants of their behaviour*'(74). Such is not the case for planets: 'At no point in a causal account of their turning does one advert to a structure which encodes Kepler's laws and causes them to turn' (74). But laws and performance rules are

being conflated here. Kepler's laws are not performance rules; it makes no sense to talk about them being followed even by a person who is perfectly capable of representing them, let alone a planet that is not.

Performance rules, do not, as Ryle reminds us, debar mistakes from happening. In this way they are different from laws of nature or those events which are supposed to be causally determined, which

> rule out certain imaginable conjunctions of happenings or states of affairs in quite a different way. Hence while there can and do occur breaches of logical rules, there cannot and do not occur breaches of laws of nature. It makes sense to speak of someone obeying or disobeying a performance-rule, none to speak of things disobeying or obeying laws of nature. (1946/2009, 239)

The rules that govern normative practices are performance rules. I suggest that these rules be construed as encoding norms which allow an observer to pinpoint whether a particular performance has lived up to a set standard, *whether or not the one credited with the ability to participate in the practice casts a sideways glance at these rules*. This, I claim, is the upshot of the regress argument: it is possible to credit someone with an ability to participate in a practice even if he is not guided by performance rules. It has to be so: consulting performance rules does not guarantee success or that the performance will live up to the standards which the rules attempt to codify. The second-order practice guarantees nothing because each of the constituent, subordinate actions involved in following a rule, let alone in representing it, can go awry. The rules represented in a logic book, for example, may have been misprinted or mistaken; they may be misread, misinterpreted as applying to a certain situation and not another; once consulted they may be ignored, or may not anticipate novel circumstances, and so forth and so on.

Fodor's imagined critic was half right to say that it only makes sense to attribute the ability to follow a performance rule to conscious, articulate organisms, but the 'organism' had better be a full-fledged rational agent since following a performance rule (as opposed merely to acting in accordance with it) involves its own series of rational actions. And though we may describe someone as following a performance rule if he tells or shows us that he is, telling and showing are yet other (third-order) actions higher up on the sophistication ladder from following the rule, so it cannot be required for it.

Fodor's argument began by giving a model of explicit deliberation that he suggested was a plausible account of how at least some behaviour is decided upon. But this model, which involves a host of constituent, subordinate actions exploits principles, which, like many of the performance rules with which we are familiar, have come about in the first place because theorists crystallise them from practices (most of which are constantly evolving) already up and running. Just as a recipe writer will make decisions (with a particular audience in mind) about what moves in a chef's performance are important for the production of a dish so will a mapmaker take a look at what the villagers do automatically in order to provide information of a different kind to one who does not know his way about.

On this way of looking at things, the practice comes first; the codification of the practice comes later. As Ryle reminds us,

> There were reasoners before Aristotle and strategists before Clausewitz. The application of rules of reasoning and strategy did not have to await the work of their codifiers. Aristotle and Clausewitz were, in fact, only able to extract these rules, because they were already being applied. The crystallisation of performance-rules in rule-formulae is, in some cases, not the condition of their being applied [i.e., not necessary in order for something to be an observance of them] but a product of studies in the methodology of the practices in which they have already been applied. (1946/2009, 243)

This is why Ryle suggests that when performance rules are applied (as opposed to followed) this is because of the intelligence of the theorists—that is, it is due to the good judgment of the logician, recipe-writer, or mapmaker who has distilled the rules (with his particular audience in mind, as well as his particular purposes for distilling them) from those performances that are deemed successful. [9]

How would we decide if someone is following a performance rule as opposed merely to acting in accord with it?[10] Among the features we would point to in justifying our

[9] For more discussion see Tanney 2005/2009 and 2008.

[10] In this discussion I am, in effect, suggesting that 'following a rule' should be reserved for actions that involve the second-order activity of consulting a performance rule. On this way of speaking, the fact that one acts in conformity with the norms, or acts in accordance with a performance rule, will sometimes suffice for deeming his action intelligent, etc. and sometimes not, i.e., when the action comes about by accident, for example. I have not discussed the relevant considerations here (but see note 8, above). Others (notably Wittgenstein and Ryle) seem to allow two senses of 'following a rule'

description of someone as performing the second-order task of following a performance rule would be the existence of some representation or expression of that rule. Other features we would expect to find would be the subordinate actions in which following a rule consists: for example, consultation of the rule-representation, and the application of it to the appropriate situation in action. And if a rule can be consulted it can be misread or misunderstood. If it can be applied it can be misapplied. The existent of these constituent, subordinate actions, and the possibility of error that they entail, completely disappear from the mechanised model; and yet genuine rule-following, and not mere rule-conformity is supposed to be essential to the model.

This discussion of performance rules gets, I think, at the heart of the normativity objection against 'naturalistic' theories. The kind of explanation in which we trade when we advert to norms —those that govern the practices for which we are assessing their ability to participate— is inextricably tied up with the idea that performances may fall short of the norm as well as conform to it. The normativity of these practices and the possibility of error are two sides of the same coin.

Indeed, if the goal of computer design is to build a machine that conforms to the norms that govern some practice, why would we need to build it to consult performance rules—ones in accordance with which it is guaranteed (barring mechanical failure) to act? Why not merely build it to conform to the norms encoded in these rules? And why should Nature design our brains to consult performance rules – ones in accordance with which she guarantees, barring mechanical failure, we will act? Why not merely build our brains to cotton on and eventually to conform whatever will turn out to be the relevant norms?

The computationalist is committed not only to representations ('tokens' of beliefs, etc.) over which computations are performed; she is also committed to the representation of the rules that govern these computations. If we construe the organism as a Universal Turing Machine, then the mental representations over which

---

or 'applying criteria' so that one may be following a rule or applying criteria in the sense that they are performing intelligently, etc., without following a rule or applying criteria in the sense of casting a sideways glance at a representation of that rule or codification of that criteria.

computations are performed correspond to the 'data' symbols on the tape; the representation of the rules that govern these computations correspond to the machine table of the particular Turing Machine that the Universal Turing Machine is mimicking; or, in short, the particular programme computing the particular data. In order to block the suggestion that if the rules themselves are represented there must be further rules explaining how to interpret them, the suggestion is that the implementation of the rules that govern the computations are built into the machine code (they correspond to the Universal Turing Machine's machine table which is built into the head or scanner). Computationalists are obliged to mechanise the rules governing the higher-order ability (i.e., they are obliged to maintain that the rules that govern deliberation and hypothesis-testing are built in) to avoid the explanatory regress. But they insist that rule-representations are being genuinely followed, and not merely instantiated: this is required by the thesis that human beings are Turing Machines and are not mere instantiations of them. But once the rules are 'mechanised' — or viewed as causal determinates of the behaviour they are alleged to explain — their role as codifications of norms and the tools that are used to guide the first-order performance evaporates.[11] This, indeed, is explicit in Fodor when he claims that all that is required for rule-following is just that 'the *causal* properties of such physical events as are interpreted as messages in the internal code must be compatible with the *linguistic* properties that the interpretation assigns to those events.'

Indeed, the computer analogy lets the representationalist down at just this point. I asked rhetorically why we would need to build a computer that consults performance rules and is guaranteed (barring mechanical failure) to act as they require? Why not merely build it to conform to these rules? But this is of course what we do. Computers do not *compute* rule-representations if this means that they consult performance rules as to how the computations should be effected. Rather the computers' mechanism is built so that certain sequences of operations are applications say, of an inference rule, which, to paraphrase Ryle, is a fact about the efficiency and intelligence of the programmer or designer and not the computer. Nor do computers

[11] As Wittgenstein objects when his interlocutor suggests that a meaning rule (or ostensive definition) could be set up for an (allegedly) private object, 'What is this ceremony for? For that is all it seems to be!' (§258).

manipulate *representations* if these are to be construed as symbols with content or meaning; for the whole point of syntactic operations is that their semantic properties are inessential.

When Dennett suggests we ignore what he calls a 'rather modest bit of lexicographical purism' he is asking us to ignore an illegitimate equivocation on the expression 'representation' which makes all the ontological difference. If the symbols are performance rule-formulae and function as such, then the use is legitimate. But for them to function as such, not only must they be interpret*able* as rule-formulae, they must be so interpret*ed* and this interpretation must guide the interpreter's performance.

In other words, and to repeat, we may agree to call uninterpreted, but interpretable symbols 'representations' as he suggests, but if we agree to stretch the meaning of 'representation' in this way, and we extend the computer analogy to the case of human agents, it would present the computationalist with a renewed homunculus problem. For if physical events are interpreted as messages in the internal code, which are in turn compatible with linguistic properties that '*the interpretation* assigns to them', an interpreter, it would seem, is still required to do the interpreting. If however, the rules that govern computations are only represented in the weak sense that the causal processes may but need not be so interpreted, then this is to admit that the rule-representation is not necessary and that the uninterpreted causal processes are sufficient for explanation. This collapses the rule-following/rule-conformity distinction and threatens the ontological status of the rule-representations alleged to govern the computations that operate over them.[12]

The Intellectualist had required the existence of higher-order practices in order to explain the intelligence of the lower-order ones; this presented him with a vicious explanatory regress. The computationalist, in an attempt to avoid the regress, collapses these practices into an algorithm in which there is no genuine latitude as to

[12] Although I cannot discuss this here, a similar argument, it seems, can be run for the first-order representations ('tokens' of beliefs over which the computations are allegedly performed). For the same vehicle-cargo model (as Ryle calls it) is adopted by any view that conceives mental concepts to signal the existence of states or events (the 'attitudes') which carry propositions or meaning as content. This means that the representationalist as well as the functionalist is threatened.

how the 'rules' of the programming language are to be implemented.[13] Whereas the Intellectualist is guilty of requiring the existence of second-order practices which pushes back the explanatory problem; the computationalist is guilty of forgetting that the performance rules belong to the second-order practice and if they figure at all, they figure for the computer designer. Dennett admits that sophisticated switching machinery is required to get the right command going to the right tools, and that it may be helpful to treat parts of this switching machinery as analogous to displaced shovellers, subcontractors and contractors. But the importance of this remark is unnoticed. If it is helpful to treat parts of the switching machinery thus, it is for the benefit of the designer. For if any rule-following occurs here, as opposed to mere rule-instantiation, it is the designer who accomplishes it. She is the one deciding how the rule is to be interpreted and implemented; it is she who will have to anticipate new applications.

Could it be that because there has been some success in building expert systems in limited domains, cognitivist scientists have been over-excited by the possibility that we ourselves are organic examples of similar computational systems? But the fact that expert systems can be built is no evidence that we are anything like them. The designer of the system is distilling performance rules and heuristics from some practice in which norms or standards that help identify correct or successful moves can be discerned.[14] In building a machine that instantiates the programme, she is in effect making decisions about how the rules are to be followed. Just like the directors of *Final Fantasy*, computer designers have looked at what people do to build their expert systems. The problem is that the cognitive scientists who are inspired by computer design have attempted to throw the community of rule-followers away in

[13] Even the representationalist is forced to say that mental representations have their content intrinsically in order to avoid a homunculus threat. In functionalism, the crux of the problem comes when we see whether it is the mental (semantic) or merely the causal properties that play an explanatory role.

[14] This is clear to see in PDP architectures, considered by some to be competitors to other psychological models as a means of explaining psychological data. In connectionist systems learning rules are used to train networks to give the 'right' outputs, given the inputs, by adjusting connection weights. According to Dawson, in order to make the transition from computer design to cognitive science, the connectionist must deduce the algorithm that the trained network is 'using' and give some evidence that this is biologically sound. But, of course, what counts as the 'right' during training is decided by the trainer. If PDP architectures are to be models for human cognition, do connectionist cognitive scientists suppose, with their classical colleagues, that what counts as 'right' for humans is determined by the 'natural' environment, independently of human practices?

supposing that all of this may be contained within the individual's brain as a part of Nature's design. But, as the computer designer should be happy to admit, the artificial machines they build are only pixellated ghosts of human practices.

What reason is there to think that the 'natural' machines cognitivists imagine us to be are not the same?[15]

REFERENCES

Fodor, J.A. 1968. *Psychological Explanation*. Random House, New York.

—— 1975. *The Language of Thought*. Harvard University Press, Cambridge Massachusetts.

—— 2000. *The Mind Doesn't Work that Way*. The MIT Press, Cambridge, Massachusetts.

—— 2008. *Lot 2 – The Language of Thought Revisited*. Oxford University Press, Oxford.

Dawson, M.R.W. 1998. *Understanding Cognitive Science*. Blackwell, Oxford, UK and Malden, Massachusetts.

Dennett 1978. 'A Cure for the Common Code' in *Brainstorms*. Bradford Books, Montgomery, Vermont.

Rey, Georges 1997. *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Blackwell, Oxford.

Ryle 1946/2009. 'Why Are the Calculuses of Logic and Arithmetic Applicable to Reality?' Reprinted in *Collected Papers*, vol. 2. Routledge, London. (Pages references are to the Routledge edition.)

——1949/2009. *The Concept of Mind*. Hutchinson, London. 60th anniversary edition reprinted by Routledge. (Pages references are to the Routledge edition.)

Tanney, J. 2005a "Reason-Explanation and the Contents of Mind", *Ratio*, Vol. XVIII, No. 3 (September), 338- 351.

——2005b 'Une Cartographie des Concepts Mentaux', Preface, Ryle, *La Notion d'Esprit*, Payot, Paris, 2005, 7-70.

——2008a 'Real Rules' *Synthese*. http://www.springerlink.com/content/k7j7h38m04343712/fulltext.pdf

[15] Thanks to David Corfield, Laurence Goldstein, Ken Westphal, and Jon Williamson for discussion.

——2008b ‘The Colour Flows Back: Intention and Interpretation in Literature and in Everyday Action’, *Journal of European Studies*, vol. 38, no. 3 (September).

——2009a ‘Reasons as Non-Causal, Context-Placing Explanations’ in *New Essays on the Explanation of Action*, edited by Constantine Sandis, (Palgrave MacMillian), 94-111.

——2009b ‘Rethinking Ryle’: Critical Introduction to *The Concept of Mind*, Ryle, *The Concept of Mind* (60th anniversary edition), Routledge, ix-lvii.

—— in progress ‘Ryle and Thick Concepts’.