Connexions

An online journal of cognitive science

ISSN 1368-3233

EDITORS: TOM DICKINS AND KEITH FRANKISH

Issue 1 January 1997

Contents

Persons, Systems and Subsystems	**Anthony Atkinson**	**2**
The Searle Workout: Connectionism Hits the Chinese Gym	**David J. Harrison**	**9**
Searle's Account of Thought and its Expression	**Sally Perry**	**12**

Sponsored by the Hang Seng Centre for Cognitive Studies

Persons, systems and subsystems

Anthony Atkinson
St Edmund Hall and the Dept. of Experimental Psychology
Oxford

1. Introduction

There is a widely acknowledged (though by no means always adhered to) distinction between the *personal* and *subpersonal* levels of description and explanation (see, for example, Dennett, 1969). In contrast to this traditional two-part distinction, I favour, and will argue for, a distinction between three levels, one of which is the personal level, and the other two of which -- namely, the *system* and *subsystem* levels -- qualify as subpersonal levels, on a certain reading of 'subpersonal'. I will endeavour to show that there is a legitimate system level distinct from both the subsystem level and the personal level. I focus my discussion of these distinctions upon cognitive psychological explanation.

I contend that the correct conception of the information processing approach of cognitive science is that it develops explanatory relations between subsystem and system levels of description and explanation, *not* (at least directly) between descriptions of persons and descriptions of parts of persons. Part of my task in this paper is to argue for this contention.

There are two further sections to this paper. In Section 2 I outline a scientific picture of the world as a conglomeration of hierarchies of subsystems constituting systems and I explain how the explanatory strategy of cognitive psychology fits into this picture. It is in that section that I draw the distinction between the system level of description and explanation and the subsystem level. The distinction between the personal level and the system level is the topic of discussion in Section 3. In that section I consider what is perhaps the most salient and arguably indubitable property of persons that marks a crucial difference between these two levels -- and consequently, between folk psychology and cognitive psychology (or any other natural science, for that matter). This personal level phenomenon is what John McDowell (1994a, p.182) has called the capacity for "reflective responsiveness to putative norms of reason".

2. Systems and subsystems

Rather obviously, the notion of a system and the notion of a subsystem are coupled. A subsystem is a component of a system. A certain sort of circuit board is a subsystem of a CD player, and heart valves are subsystems of hearts. Moreover, many single system/subsystem pairs will themselves be a part of a chain of such pairs. That is to say, there will be a hierarchy of subsystem/system relations, such that a system can itself be a subsystem of a larger system. So a certain sort of circuit board is a subsystem of a CD player, which is in turn a subsystem of a hi-fi system. And heart valves are subsystems of hearts, which are in turn subsystems of circulatory systems, which are in turn subsystems of organisms.

Given this view of levels of organization in nature as series of subsystems constituting systems, it would appear that there is no one definitive system level, nor one definitive subsystem level. I do not think this is quite the case, however. Despite the tight interlocking of systems and subsystems, there are certain entities that will stand out more than others as obvious candidates for characterization as systems rather than as subsystems, depending on the predictive, descriptive and explanatory goals of the particular domain of inquiry. Manufactured goods are prime examples of these standard or conventional systems, as are organisms and indeed, brains. Of course, it will still be possible and may even be desirable, depending on the relevant predictive, descriptive and explanatory goals, to decompose these familiar or paradigmatic systems into subsystems. But the subsystem-level constitution of a system should not detract from the legitimacy of descriptions of that system *qua whole system.* This point can be brought into focus by way of a consideration of the varieties of decompositional analysis and how they play out in cognitive psychological explanation.

Cognitive psychology is information processing psychology. The ultimate overarching aim of the information processing approach is, as one practicing cognitive psychologist recently put it,

> to come up with a set of related hypotheses about the internal structure and functioning of an information processor, i.e., with a model, that makes it possible to explain and predict

> aspects of the information processor's behaviour in terms of its internal structure and functioning (Van der Heijden, 1992, p.31)

In short, information processing theories aim to specify the internal structure and functioning of the information processing brain. The canonical conception of the internal structure and functioning of the brain as information processor is of a structured system of component information processors, memory stores, sensory transducers, and the like. (See Figure 1 for an example.)

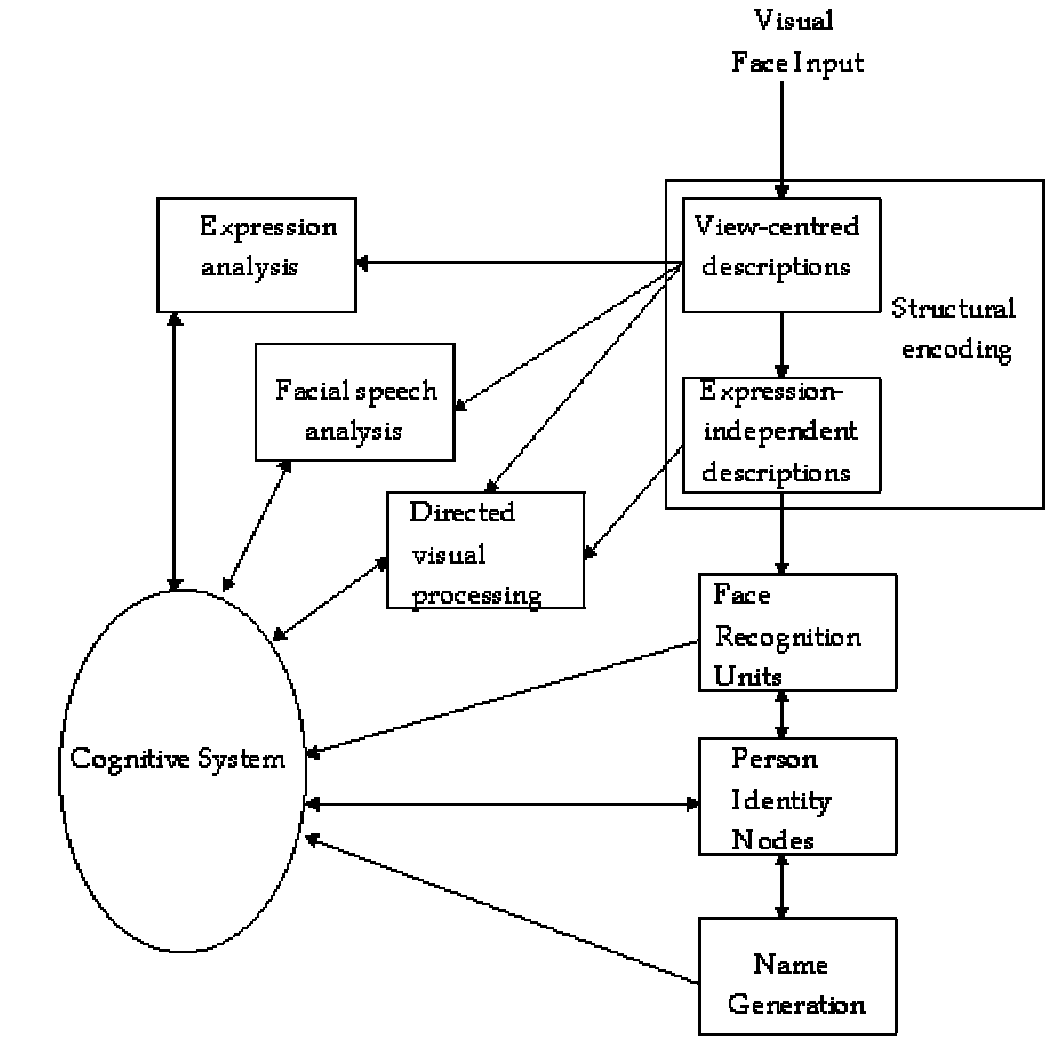

Figure 1: A copy of Bruce and Young's (1986) face processing model.

The basic assumption of the form of theorizing that treats the brain as a complex of information processing subsystems is that capacities of a whole system can be explained in terms of the operation and structure of its parts. This explanatory strategy of cognitive psychology consists of three levels or types of decompositional analysis, as follows.

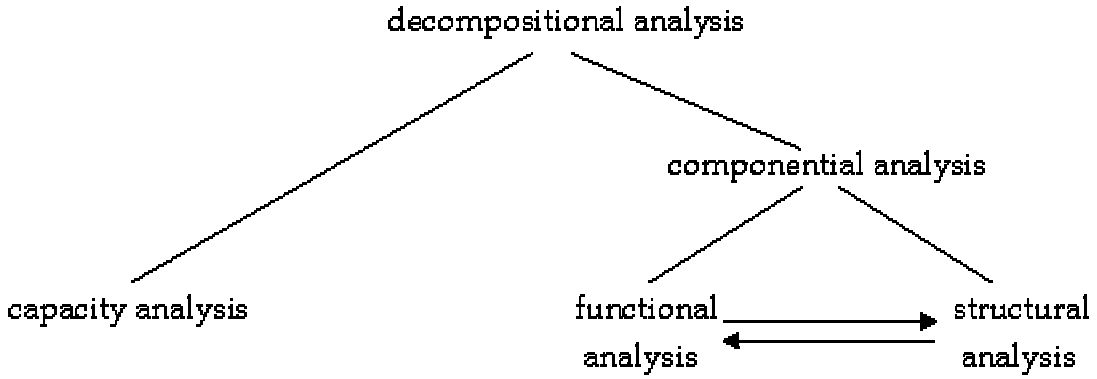

Figure 2: Tree diagram showing the varieties of decompositional analysis

A *capacity analysis* involves the specification of a capacity of a whole system, and the parsing of this capacity into subcapacities of the whole system. So, for instance, a task that a whole system has the ability to perform (e.g., baking a cake) is decomposed into subtasks that that whole system has the ability to perform (measuring the ingredients, mixing the ingredients, etc.). Another example: A capacity of the human organism is visual depth perception. A capacity analysis of our ability to perceive depth in the visual world will reveal a range of stimulus-response pairings or input-output mappings. One such input-output relation would be this: When normally-sighted subjects are presented with two photographs of exactly the same scene, taken from slightly different angles, one projected to one eye, the other to the other eye (via e.g., a stereoscope), the subject will perceive depth in the scene.

Crucially, then, a capacity analysis is a theoretical endeavour at the *system level* of description and explanation. That is to say, there is a way of talking about systems that makes reference only to activities and properties *of the whole system* or to things that happen *to the whole system.* And capacity analysis is a descriptive and explanatory enterprise that trades only in this system-level language. Within a capacity analysis, there is no commitment as to the constitution of the system -- that is, to what the system's functionally or structurally individuated subsystems are. For to make any such commitment is to change the topic from talk of whole systems to talk of components of systems.

In contrast to capacity analysis, *componential analysis* is concerned with the constitution of the system under investigation. Componential analysis is the decomposition of systems into subsystems. Hence we can mark off a proprietary level of description and explanation for componential analysis, namely, the *subsystem level.* As indicated in Figure 2, there are two varieties of componential analysis, viz., *functional analysis* and *structural analysis*. In brief, functional analysis consists in the delineation of subsystems by functions, whereas structural analysis consists in the delineation of subsystems by material structure. (There is a close interplay between functional analyses and structural analyses, as indicated by the arrows in Figure 2.)

A functional analysis is a specification of (a) what gets done by one or more components of a system (without any commitments as to what material apparatus the components consist in), and (b) the order of and relations between these functions or effects (i.e., a process, where the specified functions or effects are stages in that process). A functional analysis is also an explanation of how it is that the system under investigation has the capacity or property that it has. More specifically, a functional analysis is an explanation of a system-level capacity or property insofar as it is concerned with the functional architecture of the system -- that is, with the system's components or *subsystems*, functionally individuated -- that enables it to have the said capacity or property.

Consider, for example, the human ability to shadow speech (i.e., the ability to repeat immediately what is spoken to one). Cognitive psychologists have proposed a first-pass functional analysis of this

system-level capacity according to which there are three routes from auditory input to speech output (see Ellis & Young, 1988; and Figure 3 below). Each route consists of a small series of distinct information processing operations. One route, for example ('Route 2' in Figure 3), consists in the following four operations: (a) the extraction of individual speech sounds (e.g., phonemes) from the sound wave, (b) the registration of familiar spoken words, (c) the conversion of the representations of the auditory form of words to representations of the spoken form of words, and (d) the segmentation of the representations of the spoken form of whole words into representations of individual speech sounds (e.g., phonemes). The general proposal is that the system-level capacity of shadowing speech is realized in virtue of the performance of the information processing functions in any one of the three routes by some as yet unspecified material components of the whole system.

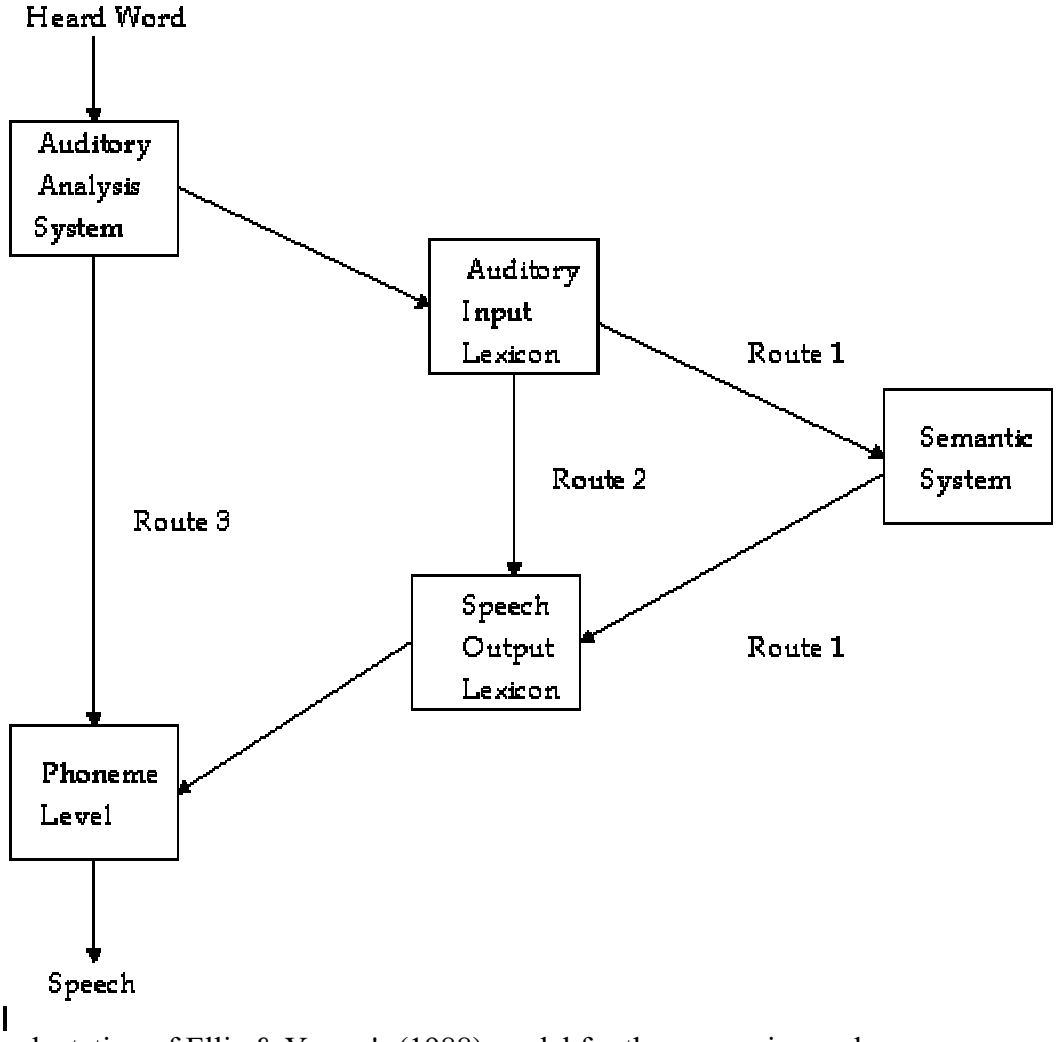

Figure 3: An adaptation of Ellis & Young's (1988) model for the processing and repetition of spoken words

This example of an actual functional analysis demonstrates two important points. First, the notion of function at work in functional analysis is what Robert Cummins (1975, 1983) called the *containing-system* sense of function. The containing system in this case is the human brain, so (a), (b), (c) and (d) -- phoneme extraction, familiar word identification, auditory word to spoken word conversion, and whole word to phonology segmentation -- are containing-system functions of components of that system. The second point is that in many cases, what has been offered as a functional analysis often looks to be a plausible capacity analysis. For example, it seems plausible to suppose that each of the four functions in the above

partial functional analysis of the ability to shadow speech is something that the *whole system* does. That is to say, a *capacity analysis* of speech shadowing might propose that one way in which the system performs this feat is that the *whole system* does the following things: (a) it extracts the individual speech sounds from the sound wave, (b) it recognizes familiar spoken words, (c) it converts the auditory form of these words to their spoken form, and (d) it articulates the relevant individual speech sounds. It is, of course, an empirical question whether one or more of a set of information processing operations suggested in a functional analysis corresponds to a subcapacity of the whole system as revealed in a capacity analysis. But it would be wrong to *assume* that this is the case.

Turning now to consider structural analysis, we can note that it involves breaking down a system into subsystems defined by their physical structure, and specifying the organization of and relations between these subsystems. More particularly, a structural analysis is a description of the material apparatus that implements the functions and processes detailed in a functional analysis. Neuroscientists are in the business of carrying out structural analyses of the brain (though arguably they are in the business of carrying out functional analyses too).

Now let us return to consider the larger picture constituted by these three types of decompositional analysis. The cognitive psychological theories that result from the deployment of these varieties of decompositional analysis are attempts to explain the capacities of brains *qua* information-processing *systems* in terms of the structure, organization and operation of those systems -- i.e., in terms of their *subsystems*. Derivatively, cognitive psychological theories are attempts to explain the capacities, properties and behaviour of organisms with brains, and especially of human organisms. What I now want to consider is a claim to the effect that cognitive psychology cannot yield complete explanations of the capacities, properties and behaviours of humans *qua persons*.

3: Persons and cognitive psychology

Human beings qua organisms are exemplary systems. Consequently, it might seem a small step to say that *persons* are also paradigmatic systems. After all, our everyday talk of ourselves and our fellow humans is of us as beings considered as a whole, as self-contained, behaving individuals. That is, if there is a personal level distinct from the system level of description and explanation, then it would be right to say that the personal level cuts the world into chunks of approximately the same size as talk of organisms-as-systems does; in other words, the personal and organism-as-system levels are pitched at the same level of aggregation.

However, there are some plausible reasons for resisting the collapse of the personal level to the system level. One reason is this. System-level characterizations of humans are descriptions of the human *organism*, a system that is ideally suited to a decompositional analysis into subsystems (i.e., organs). In contrast, commonsense psychology, the everyday practice of attributing mental states to oneself and others, is a predictive and explanatory enterprise whose target systems are, in the first instance, *people*. And persons, as characterized in folk psychological terms, are not systems of the sort that can be subject to the decompositional analyses employed by cognitive psychologists.

Now, although this reason for resisting the collapse of the personal level to the system level is appealing, it would be difficult to defend if one accepts my above argument that the system level is, in its own right, a legitimate level of description and explanation. So what further or alternative reason might one have for thinking that the personal level is distinct from the system level? Let us consider McDowell's answer to this question.

A familiar claim is that it is part and parcel of our folk psychological view of what it is to be a person that we are not simply rational creatures, but are beings who act (behave, think, judge) in appropriate ways in response to the norms of reason. That is to say, we are creatures who act *for* reasons. And according to McDowell (e.g., 1994a) and some other philosophers, it is also part and parcel of our view of what it is to be a person that we have the ability to engage in critical, self-reflective thought and judgement about our own and other persons' actions.

Persons are rationally responsive creatures who, in the words of McDowell (1994a, p.114) are subjects who are in charge of their thinking, standing ready to reassess what is a reason for what, and to change their responsive propensities accordingly.

McDowell is here describing what he calls the capacity for "reflective responsiveness to putative norms of reason" (1994a, p.182). It is an ability that

> involves not just a collection of propensities to shift one's psychological stance in response to this or that, but the standing potential for a reflective stance at which the question arises whether one ought to find this or that persuasive (1994a, p.125).

Burge (1996, p.98) puts it this way:

> Critical reasoning is reasoning that involves an ability to recognize and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such. As a critical reasoner, one not only reasons. One recognizes reasons as reasons. One evaluates, checks, weighs, criticizes, supplements one's reasons and reasoning.

According to this 'thick' conception of folk psychology, as detailed by McDowell, Burge and others, when we press the folk psychological view, we see that notions of *rationality* and *normativity* are at its core; notions that take on a special significance when applied to persons. But according to McDowell at least, the mode of understanding that is grounded upon these notions is ipso facto irreducible to the mode of understanding characteristic of the natural sciences. Nevertheless, McDowell sticks steadfastly to the claim that phenomena described in personal-level terms are phenomena of the natural world. He argues for a position that he calls "naturalized platonism", a view according to which the personal level is *sui generis*, in the sense that personal level phenomena resist capture in the web of natural laws, but not in the sense that they are supernatural.

McDowell's claim is that the purview of science is limited to what he calls "enabling condition accounts", and that no matter how thorough enabling condition accounts of personal level phenomena are, they will not add up to or replace the mode of understanding given by folk psychology. On this view, subsystem level theories, theories of the functional and structural composition and operation of the nervous system, are not able to provide, or even to contribute to, so-called "constitutive explanations".

McDowell seems to use the term 'constitutive' to mean 'analytic', in the sense of philosophical, conceptual analysis. That is, a constitutive account of X in McDowell's sense is analogous to, if not the same thing as, the conceptual elucidation of X. Moreover, McDowell seems to be of the view that conceptual analysis is set apart from scientific explanation, which is primarily an enterprise in decompositional analysis.

On McDowell's view, enabling condition theories, including cognitive scientific theories, do not tell us *what it is to be* a minded creature, to be a person; only constitutive accounts can tell us that. Rather, cognitive scientific theories make

> it possible to understand how this mindless internal control system [i.e., the brain] enables us to do what it takes to display genuine mindedness, namely, to live competently in an environment (McDowell, 1994b, p.200).

Now, if I am right about information processing psychology's explanatory role -- that is, that it develops direct explanatory relations between subsystem and system levels of description and explanation, not between the level of subsystems and the level of persons -- then it might seem that I am doing McDowell a service by demonstrating that he is right about the explanatory limits of information processing psychology. Appearances can be deceiving, however. While I believe that information processing psychology's first and foremost explanatory relation is to a descriptive and explanatory level at which humans are characterized as systems or organisms, not persons, I nevertheless also believe that information processing psychology is able to reveal more about ourselves as persons than McDowell is willing to allow. On my view, although the personal level does not belong in the scientific, system/subsystem hierarchy of levels of description and explanation, this should not be taken to imply that the personal level lies outside the realm of science. The scientific realm is not confined to considerations of system/subsystem relations, as the concepts of centres of gravity, equators and other such 'abstracta' illustrate. (See e.g., Dennett, 1987, 1991, for a discussion of abstracta.)

In conclusion, while I assent to the thick conception of folk psychology, I am nevertheless sceptical of claims to the effect that a scientific understanding of this thick conception (and in particular, of its central notions of rationality and normativity) is impossible. I am not yet convinced that to uphold naturalism one must choose between "bald naturalism" as McDowell characterizes it (i.e., the view

according to which all natural world phenomena are subsumed under natural laws) and McDowell's 'naturalized platonism'. I think that we will be able to see our way towards a position different from those that McDowell presents us if we give due consideration to the details of the nature and purview of the brain and behavioural sciences that McDowell seems obliged to license. The particular line that I would push is that a version of what Fodor (1995) has called "hairy" naturalism will be seen as a viable option once system/subsystem-level psychological explanation is properly established in the context of evolutionary biology.

References

Burge, T. (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society, XCVI*,91-116.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72,* 741-765.

Cummins, R. (1983). *The nature of psychological explanation.* Cambridge, MA: MIT Press/ Bradford Books.

Dennett, D. C. (1969). *Content and consciousness.*London and New York: Routledge.

Dennett, D. C. (1987). *The intentional stance.*Cambridge, MA: MIT Press.

Dennett, D. C. (1991). Real patterns. *Journal of Philosophy, 88*, 27-51.

Ellis, A. and Young, A. W. (1988). *Human cognitive neuropsychology.* Hove, Sussex: Lawrence Erlbaum.

Fodor, J. A. (1995). Encounters with trees. Review of John McDowell's *Mind and world. London Review of Books, 17 (8), 20 April*, 10-11.

McDowell, J. (1994a). *Mind and world.* Cambridge, MA and London: Harvard University Press.

McDowell, J. (1994b). The content of perceptual experience. *The philosophical Quarterly, 44* 190-205.

Van der Heijden, A. H. C. (1992). *Selective attention in vision.* London and New York: Routledge.

The Searle Workout: Connectionism hits the Chinese Gym

David J. Harrison
Dept. of Philosophy
University of Sheffield

Into the gym

In a 1990 article, Searle once again presents his Chinese Room argument against a computationalist picture of the mind, based on the claim that syntax is not sufficient for semantics. However, this time he also tries to deal with connectionist computational models and so he establishes the Chinese Gym. Instead of a small room containing a monolingual individual shuffling tokens around in an instantiation of a classical program that understands Chinese, Searle opens up the Chinese Gym to exercise our intuitions:

> Imagine that instead of a Chinese room, I have a Chinese gym: a hall containing many monolingual, English-speaking men. These men would carry out the same operations as the nodes and synapses in a connectionist architecture..., and the outcome would be the same as having one man manipulate symbols according to a rule book. No one in the gym speaks Chinese, and there is no way for the system as a whole to learn the meanings of any Chinese words. Yet with appropriate adjustments, the system could give the correct answers to Chinese questions. (Searle 1990, p.22)

For Searle, it seems clear that no matter how well organised and energetic these healthy individuals may be in producing a sweat they will not produce any meaning. So although they are simulating the syntax of a connectionist network, they are not producing any semantics.

Clearly, if Searle's arguments defeat all computational models of the mind, then cognitive science is in trouble Although perhaps the more extreme supporters of the 'dynamical systems' approach are going to be happy, as they claim their approach to be non-computational. So it may be that Searle's argument, rather than totally dismissing connectionist modelling of the mind, will instead merely force it to sever its classical computational ties and fully join the dynamical systems programme, suitably beefed up by its Chinese Gym workout. So let us see the benefits of exercise.

Parallel vs. serial architectures

To apply his anti-computational argument to connectionism, Searle has to first establish that connectionist networks are computational systems that share the appropriate properties of more classical computational systems, such as digital computers. Connectionism is often presented as a very different approach to classical computationalism. Major differences that have been cited are that such networks work in parallel, involve distributed representations and are analog rather than discrete state systems. However, as there are also localist, non-distributed networks, it appears that at least for the moment we can, like Searle, ignore this feature. Searle himself focuses on the parallel nature of neural networks and seems to consider only the more digital-like networks.

Searle describes connectionist neural networks as systems that "have many computational elements that operate in parallel and interact with one another according to rules inspired by neurobiology" (p.22). He adds, however, that

> The parallel, 'brainlike' character of the processing ... is irrelevant to the purely computational aspects of the processing. Any function that can be computed on a parallel machine can also be computed on a serial machine ... Parallel processing, then, does not afford a way around the Chinese room argument.

Clearly then, for Searle, connectionist networks are computational in nature and equivalent to more traditional rial architectures and so fall victim to the idea that you cannot get semantics from formal computations alone. It seems to me though that there are two elements of Searle's argument that the

connectionist can attack. First, the notion of computation can be explored to establish if connectionist networks are truly computational. A second possibility is to attack Searle's equating of serial and parallel; this can be done using part of Searle's own argument.

Exploring the notion of computation is difficult as it is variously defined -- and, of course, making connectionism come out as non-computational may remove its appeal for some and perhaps even drive its supporters deep into the dynamical systems camp and non-representationalism. Of course, in the article itself Searle says that 'humans are, at least in a trivial sense, computers' (p.22) -- as indeed is just about anything, even parallel distributed processing neural networks. However, most people consider the computational nature of networks to be far from trivial. The most appropriate definition of computation to begin with is probably the one outlined by Alan Turing in his 1936 paper on computable numbers: computation, says Turing, is the formal manipulation of (uninterpreted) symbols by the application of formal rules.

There are problems for connectionism with such a definition, though -- problems that also crop up in definitions of cognitive architecture for connectionism. Essentially, it is often argued that:

There is certainly computation in connectionist systems, but such computation lies below the conceptual level. Computation takes place at the level of nodes and connections between nodes, and the individual nodes and connections are not intended to carry any semantic burden alone. The semantic burden of the system lies at a higher level, that of the *distributed representation.* (Chalmers 1992, p.32)

A problem with a connectionist view where the semantic and syntactic levels are separate is that it would seem to be unable to conform to the standard definition since the syntactic processing is not over the symbol tokens of the system. Whatever the network is doing, it is not the formal manipulation of representations according to rules, since in PDP networks it is not the individual units that are the symbolic entities. This actually opens the door for the dynamicist. Dynamicists such as van Gelder hold that computational operations are 'discrete, identifiable steps in which one representation gets transformed into another' (van Gelder, 1995, p.354). When the system is a parallel, analog dynamical system (as the most interesting types of networks are), then it seems we don't have sequential operations over syntactic tokens and so the dynamicist presents a non-computational picture. It seems a case can be made for parallel analog systems being non-computational -- particularly as Turing's own definition is actually only applied to discrete-state machines such as digital computers. At the very least it has to be said that connectionist computation is not the symbolic computation that Searle attacks.

This leads on to the second element. Turing showed that all digital computers are in a sense equivalent, in that a Universal Turing Machine can be programmed to replicate the input-output profile of any other Turing Machine, and this seems to lead Searle to claim that 'Computationally, serial and parallel systems are equivalent ... [So] if the man in the room is computationally equivalent to both, then if he does not understand Chinese solely by doing the computations' then neither do our exercisers in the Chinese gym (p.22). What Searle ignores, though, is a point that he himself spends a great deal of time making. Which is that *simulation should not be confused with duplication.* (Searle 1990, p.23)

The gentlemen in the Chinese gym and the programs run on serial computers are both merely computational simulations of parallel neural networks. 'The computational model is no more real than the computational model of any other natural phenomenon' (Searle p.23). What is missing in such a simulation, as the Churchlands (1990) try to point out in their reply to Searle, are the higher-level features of the system that are said to emerge from the local interactions among units and which carry the semantics of the system. Searle concedes that this is a possibility, but claims that it has nothing to do with computation. 'Computationally, serial and parallel systems are equivalent'. But the equivalency is not yet determined, at least until another definition of computation is found and it seems clear that Searle does not wish to reject the existence of higher-level features of systems.

Intrinsic properties

Most connectionists do not wish to ignore the higher-level features that emerge from the unit interactions. As Dinsmore points out:

> The recognition of higher-level structures is in no way inconsistent with the thesis that cognition emerges from connectionist principles. But like many kinds of physical systems, the connectionist architecture can be assumed to organize itself into higher-level structures and processes with their own properties that call for a higher level of

> description. We should always look for higher-level organizational principles. (Dinsmore, 1992, p.18)

In the same way that the connectionist and dynamicist talk of the cognitive level arising from the neuronal level, Searle has long argued that 'lower-level neuronal processes cause higher level mental phenomena' (p.23). For him 'the brain does not merely instantiate a formal pattern or program (it does that, too) but it also causes mental events by virtue of specific neurobiological processes' (ibid.)

A strong element of Searle's anti-computational view is opposition to the fact that computation is all about the formal properties of tokens. What Searle attacks is strong AI and 'the basic premise of strong AI is that the physical features of the implementing medium are totally irrelevant' (p.25). All that matters is the abstract causal structure that can can be captured computationally -- that is, the behaviour of the system in terms of inputs and outputs.

What Searle says is that what is also required is the 'right causal processes' underlying the behaviour. The actual realization of the behaviour is important -- unlike computational processes, whose formal structure could be realized by anything that could change from the input state to the appropriate output state. So Searle and the connectionists both claim that the processes underlying the behaviour are essential *and* essentially depend upon their realization to have the specific properties that they do. Thus they cannot be merely computational processes, since these are 'are completely independent of any reference to a specific type of hardware implementation' (Searle p.25).

In both brains and networks it is not individual neurons or units and their activity that explain or cause behaviour and cognition, but the activity produced by masses of neurons that self-organise to produce new global forms of behaviour. So it is the activity, interaction and organisation, rather than the formal, abstract properties and features of brains and networks, that are relevant to the study of cognition. The activity of neuronal groups is the representation. Parallel networks and brains are dynamical systems where representation occurs, but it is an activity, rather than a static structural item like a formal token -- an activity resulting from the specific patterns of causal interactions of the components, rather than from specific structural properties of the components. So clearly these are intrinsic properties. They are not the abstract functions and dispositions that classical computational models favour, but neither are they purely material stuff properties as Searle seems to believe they are. In a more dynamical conception of connectionism and the mind we have a picture of mental representations as being intrinsically active and of cognition as their own self-processing.

In conclusion, it seems to me that Searle has no real objections to the connectionist approach as more dynamically conceived. For such an approach does not have formal syntax directly producing semantics, nor does it require an intermediate computational level between the neurophysiological level and the intentional level. Instead, real neural networks, as opposed to mere simulations, explain how mental phenomena are caused by and realized in the physical, dynamical structure of brains and networks. And as long as consciousness is seen as of central importance, Searle can be happy with connectionism, for:

> nothing I say is inconsistent with [the] hypothesis that the appropriate explanatory level for cognition is that of 'masses of neurons' (Searle 1991, p.143).

References

Chalmers, D. (1992) 'Subsymbolic Computation and the Chinese Room' in J.Dinsmore (ed.) *The Symbolic and Connectionist Paradigms: Closing the Gap*, pp.25-48.

Churchland, P.M. and Churchland, P.S. (1990) 'Could a machine think?' *Scientific American*, 262, no.1, 20-25.

Dinsmore, J. (1992) 'Thunder in the Gap in J.Dinsmore (ed.) *The Symbolic and Connectionist Paradigms: Closing the Gap*, pp.1-24.'

Searle, J. (1990) 'Is the Brain's Mind a Computer Program?' *Scientific American*, 262, no.1, 20-25.

Searle, J. (1991) 'Response to Skarda and Freeman' in E. Lepore and R. van Gulick (eds.) *John Searle and his Critics.*

Turing, A. (1936) 'On computable numbers, with an application to the Entsheidungsproblem.' *Proc. London Math. Soc.* 42, 230-65.

van Gelder, T. (1995) 'What cognition be, if not computation?' *Journal of Philosophy*, vol. 91, no.7, 345-381.

Searle's Account of Thought and its Expression

Sally Perry
Department of Philosophy,
University of Sheffield

Language is, essentially, the expression of thoughts. This is to say that only if a verbal performance is an expression of a thought does it constitute a linguistic act. So what is it to express a thought? The undeniable conceptual dependence of language on thought just stated has motivated many, including John Searle to attempt to account for what it is to express a thought in language or by means of any other 'utterance' (in the Gricean sense) in terms of the intentional states of speakers - in terms of speakers' beliefs, desires and intentions. A reduction of language to thought has seemed possible for the reason that there no such undeniable conceptual dependence running in the other direction - of thought on language. Meaning (ultimately linguistic meaning) has been maintained to be explicable in terms of the capacity to have thoughts because although there can be thought in the absence of language you can't be a language user without having thoughts. This asymmetry is expressed in Searle's distinction between the *intrinsic* Intentionality (representational feature) of thoughts and the derived Intentionality of linguistic entities[1].

I take it that Searle's account of expression (1983 Ch. 6) is meant to give at least necessary conditions for an utterance to be an expression of a thought and so for it to have meaning. My objection to the account is that it requires possession of the concept of belief on the part of an agent capable of meaningful utterances and this is not a necessary condition for the expression of thoughts in language. I do not intend to argue for the claim that the concept of belief is not necessary in order to be a language user although I will indicate the direction in which such an argument would go at the end of the paper. Searle, strangely enough, agrees that the capacity to express thoughts is independent of command of this concept. He thinks that his account of expression respects this independence. So my immediate concern is to establish, *pace* Searle, that his own account does indeed demand that the concept of belief be understood by any creature capable of expressing thoughts.

First then, I will outline the account of thought and of what is maintained to constitute its expression. I will go on to illustrate how the account presupposes the concept of belief on the part of the utterer. I will end with some remarks concerning why this presupposition makes the account unacceptable as an account of what constitutes meaningful uses of language.

The points which define the intrinsic Intentionality of thoughts are given by Searle as follows,

1) An Intentional state is a representation of the conditions of its satisfaction (p.177)
2) There is no way an agent can have a belief without that belief having its conditions of satisfaction (p.22)
3) Part of what it is to have the conscious belief [e.g.] that it is raining is to be conscious that the belief is satisfied if it is raining and not satisfied if it isn't (p.22)
4) Intentional states are individuated by their conditions of satisfaction and by their directions of fit. To have a belief is to be in a state with a mind-to-world direction of fit. To have a desire is to be in a state with a world-to-mind direction of fit.
5) When we have ascribed to a creature the capacity to have Intentional states we have already ascribed to it the capacity for relating its Intentional states to objects and states of affairs in the world (p.177)

Utterances derive their Intentionality by virtue of the fact that they are produced with the intention to express beliefs, desires and other Intentional states. Thus, on Searle's account, when an individual expresses a thought she confers meaning on her utterance-act in virtue of performing it with the intention that it have

[1] I follow Searle in using the upper case when referring to content bearing mental states as 'Intentional states' though nothing hangs on this.

the same content (carry the same conditions of satisfaction) as the thought she intends to express. For example, if an individual intends to express her belief that it is raining by means of the utterance-act of raising her arm (we are to think of a primitive individual without the resources of construct sentences) the content of her Intentional state will be,

> (My arm goes up as a result of this intention in action and my arm going up has conditions of satisfaction with the mind (or utterance)-to-world direction of fit that it is raining) (p.167)

This contains the crucial component for meaning, the meaning intention (MI) with the content,

> MI (My arm going up has conditions of satisfaction with the mind (or utterance)-to-world direction of fit that it is raining)

If Searle is correct concerning the implications of his own account the concept of belief should not be required either in order for a creature to satisfy conditions 1-5 or in order for a creature to entertain MI. First, then, let us look at Searle's account of intrinsic Intentionality to ensure that the concept of belief is not being presupposed at this stage.

1) just commits us to a way of thinking about Intentional states and their contents; the belief that it is raining is a 'representation' of the worldly state of affairs of its raining and the belief is satisfied (true) if and only if that state of affairs obtains. 2) just stresses the internal relation between an Intentional state and its particular condition of satisfaction; one cannot have the belief that it is raining without one's state being satisfied if and only it is raining. 4) adds to this framework the theoretical concept of direction of fit. To say that a belief is a state with a mind-to-world direction of fit is to say that it is a mental state which must conform to the world if the state is to be satisfied. To say that a desire is a state with a world-to-mind direction of fit is to say that it is a state to which the world must conform if the state is to be satisfied. So far so good. I propose to leave 1), 2) and 4) alone. I have no relevant objections to them. My main worry lies with 3) and 5). The most natural interpretations of these conditions as stated are ones on which any agent capable of satisfying them, and so any agent capable of having belief, would seem to need the concept of belief. But this, Searle and I agree, need not be the case.

3) and 5) as stated are strictly false of an agent that lacks the concept of belief. No such creature could be conscious *that its belief is satisfied if it is raining and not satisfied if it isn't* I see no other way of construing what it would be to enjoy such a mode of consciousness apart from saying that it is to be conscious of the emphasised content. But to be conscious of such a content would require that the agent have the concepts of belief and satisfaction - concepts which figure as constituents of the that-clause. Moreover, 3) implies that a creature with a conscious belief be simultaneously aware of two opposing contents that it is raining and that it is not raining and that it be aware of the implications for its mental state of each of these contents holding. It therefore implies that a creature with a conscious belief be aware of the possibility that its belief be mistaken, that the creature be aware that its belief may be inconsistent with the world. Now this is surely to presuppose the concept of belief on the part of the agent - it is to presuppose that the creature know that the way the mental state represents the world to be may not be how the world is in reality.

Similarly 5), on one interpretation, presupposes a grasp of the concept of belief. Only a creature that is aware of its Intentional states on the one hand and the states of affairs which they purport to represent on the other, and knows of the relation between these two things can be said to relate its Intentional states to objects and states of affairs in the world.

3) and 5) have to be construed along much weaker lines if they are to allow thoughts to creatures lacking the concept of belief. In fact Searle does employ a weaker interpretation of 3) when he writes that to have a conscious belief is to have 'the capacity for recognizing what it would be to get it right' (p.177). This recognitional capacity is perhaps supposed just to fall out of 1); the belief that it is raining is a 'representation' of the state of affairs of its raining, so a creature with the conscious belief that it is raining will be able to recognise the truth condition of that belief, the state of affairs of its raining. That truth condition will, in virtue of the subject's having the belief in consciousness, be made available to it in conscious thought. If the belief is false and it is not in fact raining the creature will, on becoming aware of this fact, simply drop the belief. The creature would not have to judge of its belief that it is false in order to abandon it. The content it is raining is, because of the creature's subsequent experience to the contrary, no

longer accepted. The content recedes from consciousness and the creature is no longer disposed to consciously judge that it is raining.

Perhaps this lends some sense to the idea of an unselfconscious creature being able to distinguish when a content is correct from when it is incorrect. It is at least part of the story of how a creature without the concept of belief - without the capacity to judge of its beliefs that they are true or false (mistaken) - can still be a creature whose representational states are properly called 'beliefs'.

The above also serves to indicate that my objection to Searle's strategy for explaining language in terms of a prior (or language independent) account of thought does not depend on rejecting the asymmetry pointed out at the beginning of this paper. Rather, as I said, my objection is merely that the account of expression presupposes the concept of belief on the part of those capable of meaningful utterances. I now give my reasons for taking this to be the case.

Expressing a thought is maintained to be a matter of intentionally imposing the conditions of satisfaction of the thought to be expressed from the thought to the utterance-act. The content of the meaning conferring state is given in MI. Now, were we to interpret MI as a literal rendering of the content of the utterer's Intentional state we would clearly be committed to crediting her with the concept of a thought. This is so simply because the content so given contains the concepts of satisfaction or fit, concepts of the relation that Intentional states are supposed to bear to the world. As I said, in order to have the concept of such a relation one needs not only to know that one's mental states are distinct from states of world but also, in the case of belief expression, one has to know that one's belief states are rendered correct or incorrect by states of the world. Thus belief expression would seem to require the concept of belief.

However, in discussion, Searle has claimed that his account is not to be taken so literally. He maintains that MI is merely the theorist's way of describing the state of an agent who is expressing a thought. Of course the theorist needs the concept of satisfaction or fit and so an understanding of the nature of belief, but the agent herself does not. I think this is an obvious evasion, but I intend to approach the matter almost painfully slowly to ensure that the point is made as decisively as possible.

Of an agent in a state with the content of MI it is true that she

i) has the conscious belief that it is raining
ii) intentionally raises her arm when she believes it is raining
iii) raises her arm with the intention of expressing the belief that it is raining

From the agent's point of view i) and ii) can be stated thus

i') It is raining
ii') I (intentionally) raise my arm when it is raining

The problem is how to credit the agent with a conception of iii), the meaning conferring part of the intention with which her action is performed, without crediting her with the concept of belief, that is, without crediting her with grasp of the fact that some of her mental states are correct or incorrect depending on whether the world is the way her states represent it to be.

To be thorough, we should ask whether the agent even needs to have a conception or an awareness of something that amounts to iii). Obviously she does. Firstly, I can't make much sense of the idea of an agent *intentionally imposing* conditions of satisfaction on her utterance acts without having any understanding (in the sense of conscious awareness) of what it is that she is doing. Secondly, without awareness of something that amounts to iii), the agent would, from her own point of view, merely be intentionally raising her arm when (from her point of view) it is raining which by itself doesn't amount to an intentional act of meaning that it is raining. Thirdly, it is a constraint on attributing an intention to a subject that the subject have some conception of the desired end. So the subject needs to know why she is raising her arm in the rain if this is to seem like an intentional/rational/purposive action either from the subject's point of view or from our own. So everything hangs on finding a way to credit the subject with a grasp of the part of the intention with which she performs her utterance-act which is maintained to confer meaning on the act.

Again just to be thorough, we should run through what we can uncontentiously say is part of the agent's awareness in having an intention with the structure of MI. First in having the conscious belief that it is raining (i' above) we can say that the agent is aware of a certain content's holding. Of course, she may be wrong, there is always the possibility that it not be raining. But an awareness of this possibility is not given

to her just in virtue of her capacity to have mental states with intrinsic Intentionality. Although she may be responsive to evidence to the contrary if and when she is confronted with it, (a responsiveness which consists in her capacity to abandon the belief), to say that she is aware of the possibility that it not be raining when she is in the state of belief with regards to the content it is raining would be to credit her with the concept of belief, an understanding that the way her state represents the world to be may not be how the world is in reality, and this, we are supposing, is a concept she does not have[2].

Secondly, in having the conscious intention to raise her arm when (from her point of view) it is raining (ii' above) we can say that the agent is aware of her intention to raise her arm when it rains. To be aware of one's own aim, goal or intention in performing an action does not seem to require the concept of belief. But neither does the awareness of the aim to perform a specific bodily action amount to awareness that one's action is the expression of a belief. The problem is still that we need to credit the agent with something that amounts to the awareness of the aim that the action have conditions of satisfaction.

Searle, (again in discussion), has offered an alternative way of capturing the normative notion expressed by the term 'conditions of satisfaction' which is supposed not to tempt one to the idea that expressing a belief requires the concept of belief. He claims that just as all a creature needs in order to have a belief is to know when the belief is 'Ok' and when it is 'not Ok', so this sort of knowledge is all that is required for a creature to express a belief. But this is too loose to be of help in the present problem. Were we to say that the agent aims that her action be Ok or not Ok *period* we would be leaving things radically underdetermined. We would not have captured the sense in which the agent's action is Ok or not Ok, for, even if we accept that the action is an expression of belief, (which as yet we have found no reason to), it is always 'Ok' for someone to express her beliefs, even in cases when they are false.

We would, then, still need to fill things out by saying something that amounts to the agent's being aware that the action is Ok or not Ok in the same sense that the belief is or that the action is Ok or not Ok depending on whether the belief is Ok or not Ok. We can't say this exactly because this would again be to presuppose that the agent has the concept of belief. But what are we to say? Any way of stating what the agent must be aware of in having meaning as her aim will presuppose this. If, avoiding the direct reference to belief, we say that the agent aims that her action is Ok if performed when it is raining and not Ok if performed when it is not raining, and that this aim is what confers meaning on the arm-raising, we are supposing that the agent has the capacity to acknowledge that it might not be raining even when she is in the state of believing that it is, for, in order to express the belief that it is raining, the agent at least needs to be in the state of belief with regards to that content[3]. But this is just to assume that the agent can acknowledge, of the belief she now has, that it might be false - that it might be inconsistent with the world - and this is to assume the concept of belief on the part of the agent.

In one final attempt to drive home the point that there is no way Searle's account goes through unless we suppose that the utterer has the concept of belief we could see what happens if we do not suppose this of the utterer. The aim that an action be Ok if performed when it is raining and not Ok if performed when it is not raining is not an aim that an agent lacking this concept could have. The concept of belief is, for our purposes, just the awareness that the way things seem to one need not necessarily be the way things are in reality. Now, an agent lacking this awareness could raise her arm whenever it seems to her that it is

---

[2] In fact I think this is where Searle and I disagree. I thi nk Searle is assuming that an understanding of the distinction between the way one's Intentional states represent the world and the way the world is in reality (the is/seems or reality/ appearance distinction) is available to one at the level of perception and action. He just doesn't think grasp of this distinction involves, or amounts to, grasp of the concept of belief. Of course a creature can be aware of the possibility of some action it is performing or about to perform failing to have the desired outcome without having the concept of belief. Thus a cat might know that its attempts to catch a rat don't always result in catching the rat. But there is a difference between this sort of awareness ('either I will succeed in my aim or I will fail') and the awareness that a proposition one holds true may be at odds with reality. The former awareness does not does not require the latter. As for perception, Searle seems to think that even a dog or a cat can learn on the basis of perceptual errors that it can take an object to be something other than what the object actually is. This requires a reflective understanding of one's own thoughts, the capacity to judge of one's perceptual judgements that they either do or do not correspond with the way things are in reali ty. This sort of understanding does require the concept of belief. But for reasons which I shall give later, Searle would do well to refrain from maintaining that such an understanding is available to any creature capable of action, perception and thought.

[3] I have assumed throughout that an as yet languageless creature, and one lacking the concept of belief, can only form an intention to express the belief that p when she has the conscious belief that p. In *Intentionality (p.167)* this is not assumed. Searle has the agent forming the intention to express the belief that p in advance of believing that p. But surely forming an intention to express a belief in advance of having that belief requires the concept of belief. In such a case the reference to belief i n the content of the intention would surely be ineliminable. Moreover, forming an intention to express the belief that p in advance of believing that p would also require that one have some way of (consciously) representing p to oneself independently of be ing in the state of believing that p - it could plausibly be argued that this requires language. I have stuck to the case where someone intends to express the belief that p only when she consciously believes that p because otherwise I do not think that Sea rle has any hope of a coherent account.

raining. But if there are cases when this action is performed when in fact it is not raining this is not something such an agent could ever discern. Since she is (supposedly) expressing her belief that it is raining she will necessarily believe that it is raining and so she will necessarily be unable to distinguish cases when her action is Ok in the required sense from cases when it is not Ok in the required sense. If the distinction between cases when her action is Ok and cases when it is not Ok is a distinction such an agent could not make then surely it is a distinction such an agent could not aim or intend to make. But MI is an intention to make just that distinction. So an agent lacking the capacity to make the distinction is incapable of having a meaning intention.

What makes MI plausible as the content of a state capable of underpinning meaning is the fact that it is a content which requires that any creature capable of it would know the conditions under which its utterance-act would have to be not merely withheld or abandoned, but judged as mistaken, false or incorrect. Since meaning intentions are supposed to be available to an as yet languageless creature the notions of mistakeness, incorrectness and falsity have to be understood by the creature as applicable to thoughts prior to a creature being able to apply the notion to non-verbal or verbal utterances. Meaning does require an understanding of the conditions under which an utterance is incorrect, false or mistaken but what I will say below will nudge us more towards the idea that these notions be understood primarily as applied to linguistic entities.

So what, after all, is wrong with maintaining that language use - the intentional expression of thoughts - requires the concept of belief? Recent developmental literature is replete with studies which purport to establish that linguistic competence - including the ability to give information (express beliefs) by means of language - is temporally prior to, and so conceptually independent of, an understanding of the notion of belief[4]. Young children, it seems, are competent language users before they are able to pass a battery of tests designed to establish whether they understand that either they themselves or others have states which may fail to represent the way the world is in reality. The data casts doubt on the assumptions concerning what sorts of cognitive capacities underlie meaning and expression shared by Grice and his followers, Davidson and (as I have argued) Searle. The data tends rather to give empirical backing to the claim of Wilfred Sellars (1958) that any analysis of meaning in terms of the expression of thought fails if it entails that a being capable of meaningful utterances must have the concept of a thought. Sellars justified the claim purely on the grounds that we can conceive of a linguistic community in which no individual possessed psychological concepts. The empirical data suggests that what Sellars identified as possible is very likely actual.

The empirical data can, of course, be questioned. But even if it were found that the data fails to establish conclusively a lack of belief understanding in some language users we could still pursue Sellars' point with effects detrimental to the strategy of explaining meaning and language use in terms of the sorts of psychological states which require psychological concepts. Sellars' point is one which has also been argued by Michael Dummett (e.g. 1993 Chs. 6 and 7): if linguistic competence is explained in terms of a prior grasp of thought contents then we are committed to conceiving of thought as something which is essentially non-linguistic. If, as in the case of Searle's account, meaning is constituted by the intentional imposition of conditions of satisfaction from the thought to the utterance, then grasp of the thoughts to be expressed must necessarily be independent of the capacity to express them. But isn't it possible that there be some thoughts that can only be grasped via an understanding of the linguistic means of expressing them? Searle (and indeed many Griceans) have admitted that this is possible. Searle has claimed that his account is only meant to apply to 'simple' thoughts. He has conceded that more complex thoughts can only be had by coming to understand the linguistic means of expressing them. But if is the case then Searle's account would seem not to give necessary conditions for meaning. It would seem possible to side with Dummett when he claims that in acquiring language one can come to have the thoughts expressed by means of language - it doesn't matter if the capacity to have any of those thoughts temporally precedes the capacity to express them. For Dummett the possibility that language serves as a vehicle for thought, a means (and in some cases the only means) by which thoughts are grasped, is enough to rule out the strategy of explaining linguistic expression in terms of a language-independent notion of thought.

Of course, this is to assume that the notion of a creature's grasp of a thought being constituted by the hearing/uttering and understanding of a linguistic expression of it - something like Sellars' notion of

[4] See for example Baron-Cohen (1995). The phenomenon of 'mind blind' language users is not found only in young children. Baron-Cohen's book is largely concerned with the lack of belief understanding in individuals wi th autism.

thinking-out-loud - is a defensible one. Given the suggestion that linguistic understanding be autonomous of psychological concepts - the threat that the empirical data poses to the enormously influential accounts of meaning and expression along broadly Gricean lines - the issue is at least worth pursuing.

References

Baron-Cohen, Simon 1995: *Mindblindness*(MIT Press)

Davidson, Donald 1975: 'Talk and Thought' pp. 155 - 170 of *Inquiries into Truth and Interpretation* 1985 (Oxford University Press)

Dummett, Michael 1993: *The Seas of Language* (Clarendon, Oxford)

Searle, John 1983:*Intentionality* (Cambridge University Press)

Sellars, Wilfred 1958: 'Intentionality and the Mental' in H. Feigl, M. Scriven and Grover Maxwell (eds.) *Minnesota Studies in the Philosophy of Science* Vol.II pp. 507-539 (University of Minnesota Press, Minneapolis)