Folk Psychology, Connectionism and Constraints on Believers

1 Introduction

The folk psychology (***FP***) and eliminativism debate comes to a crucial focus as a result of the appearance of connectionism on our scientific horizon. In this paper, I will examine Andy Clark's efforts[1] to counter Martin Davies' Conceptual Modularity argument[2] made in favor of eliminativism. This argument has as its conclusion, the disquieting result (for proponents of ***FP*** at least) that connectionism, if held true as a model of the mind, is in deep and fundamental conflict with most folk psychological platitudes[3] about the mind.

Clark arrives at a position that sounds, looks and feels a great deal like a version of Dennett's first person operationalism; a folk psychology without the ontological commitments. I will look at Clark's arguments for this position and examine the further two requirements that he places on believers:
1. Requirement of Normative Depth[4]:

> The inner workings of an intentional system must be of a kind compatible with the description of that system as capable of making mistakes which

[1]Clark, Andy. Associative Engines: Connectionism, Concepts and Representational Change, Chapter 10, Cambridge, MA, Bradford Books/MIT Press, 1994

[2]Another Propositional Modularity argument is presented by Stich, Ramsey and Garon in "Connectionism, Eliminativism and the Future of Folk Psychology" in *Folk Psychology and the Philosophy of Mind*, Christensen and Turner eds., Hillsdale, NJ, Lawrence Erlbaum Assoc. Publishers, 1993. RSG are committed to the following claim about propositional attitudes. They are:

> functionally discrete, semantically interpretable predicates that play a causal role in the production of other propositional attitudes, and ultimately in the production of behavior.

RSG argue that connectionist models display no such characteristics: superpositional storage as found in a typical distributed, connectionist network renders as nonsensical, the question of whether a particular piece of information stored in a network was the causal factor in a particular action or decision. This threatens us with total causal holism; an idea that any cognitive episode involving our stored beliefs causally implicates all of them. But ***FP*** is committed to the idea that we can causally implicate one belief out of many that a thinking agent might have held. Without this commitment, it's hard to make sense of ***FP*** assertions such as "John went to the opera because he wanted to see Pavarotti embarass himself". In addition, where ***FP*** finds a natural psychological kind i.e all the believers that vindaloo is spicy (Fs such that all Fs possess the attribute $B(V_s)$), connectionist models often represent equivalent information to one another with no features in common. Same content; different form To even ask for a common form is to commit a category mistake.

[3]What ***FP*** is committed to is disputed in these debates of course, but it is reasonably uncontroversial that ***FP*** is committed to propositional attitudes, propostional modularity (which forms the basis for the Ramsey, Stich and Garon argument) and causally productive internal states.

[4]Clark, Andy, *Associative Engines*, page 216

involve the failure to respect those commitments in episodes of on-line processing.

2. Requirement of Consciousness[5]:

> ..it seems to me conceptually impossible for a being to count as grasping a concept and yet be incapable of ever having any conscious experience involving it.

I will argue that while Clark's move to a neo-Dennettian position is relatively unproblematic, his constraints, designed to protect himself from the sorts of counterexamples and arguments made against Dennett's view, actually make matters worse for his case.

2 Davies' Argument for a Language of Thought; Systematicity and Cognitive Processes

Davies' argument against ***FP*** starts as one for the language of thought (***LOT***) and by the time he finishes, a putative case for eliminativism can be said to have been made. The following represents the structure of the argument[6]:

> If a cognitive process is systematic-in a sense yet to be defined-then the inputs to that process have syntactic structure..being a thinker-a believer, a deployer of concepts- involves the systematicity of certain inferential transitions amongst thoughts. Consequently, the inputs to those transitions-thoughts-have syntactic structure; that is, there is a language of thought.

The argument for a ***LOT***, requires then, that thoughts have syntactic structure. More comprehensively, if thinking were to be the act of making *systematic* inferential transitions amongst syntactic structures then we would have a ***LOT***. Or, the act or state of thinking should possess syntactic properties. What kinds of properties should count as syntactic ones? The properties in question should be physical properties; they should be *systematically* related to semantics and they should be determinants of causal powers.

What distinguishes a cognitive process from other kinds of processes? For Davies, a cognitive process' input states have semantic descriptions; its output states have

[5]Ibid, page 217

[6]Davies, Martin, "Concepts, Connectionism and The Language of Thought", printed in W.Ramsey, S. Stich and D. Rumelhart eds., Philosophy and Connectionist Theory, Hillsdale, NJ, Lawrence Erlbaum Assoc. Publishers, 1991, pp 229-257

either a semantic description or an action description. A system employing cognitive processes moves from *information* to *action*. Paradigmatic folk psychological statements then, apply to cognitive processes: "I opened the door because I wanted fresh air". An action here is the result of some information that can be given a semantic evaluation; 'I wanted fresh air' is true and is the causal determinant of my door-opening action. If thinking is a cognitive process, it should possess syntactic properties and if it were systematic, inputs to it i.e thoughts, would possess syntactic structure. Being a thinker involves inferential transitions amongst thoughts (presumably, this much is uncontroversial[7]); if these transitions were to be systematic, we would have a ***LOT***. So a definition of systematicity is important to proceed any further.

Systematic processes are defined relative to a pattern in their input/output relations; given a cognitive process, we could define systematicity relative to its input-output pattern. Davies goes on to use a notion of *causal systematicity* in his description of systematic processes. What is causal systematicity? An explication in terms of causal explanations: if we can describe a pattern in the input-output relation of some physical system and find the same pattern in other input output transitions, then a pertinent question is whether these transitions have a common causal explanation that corresponds to the common pattern just exhibited. If such a causal explanation is forthcoming, then we are justified in describing the pattern at hand as being causally systematic. In Davies' drink machine[8] example, the presence of a common mechanism that services each of the four possible requests would satisfy the requirement of causal systematicity[9]:

> If we think of a physical system [the drinks machine] as containing various subsystems or mechanisms, then the requirement for causal systematicity relative to the pattern described by G is that there should be a mechanism in the system whose presence in the system explains all the input output transitions that conform to the pattern described by G. It is not sufficient that this common mechanism should merely figure as a component somewhere along the way in the several transitions. Rather, the common mechanism should actually mediate between inputs and outputs in accordance with G.

The mediation requirement ensures that the common mechanism can figure in a causal explanation of the displayed pattern. A completely modular system in the drinks

[7] Well, for Davies at least; there certainly is no argument to this effect in his paper.
[8] Ibid, page 234-235
[9] 9Ibid, page 235

machine example that employed separate processors for each drink request would not meet the causal systematicity requirement since a common mechanism is not being employed to ensure the appropriate outputs. A ***LOT*** would require that there not only be such an input-output pattern in cognitive processes but that the required causal systematicity be present as well in the form of a common causal mechanism.

Davies is concerned with systematic cognitive processes; processes that are systematic relative to patterns revealed by a semantic description. Causal systematicity relative to some semantic input-output patterns requires that the input states of the machine should have properties that are correlated with their semantic properties (i.e if an input state in the drinks machine *means* 'drink with milk' then the output state is drink-with-milk; if it *means* 'drink without milk', then the output state is drink-without-milk) and are determinants of the causal consequences of those states given the internal structure of the machine (if the input state means 'drink with milk', it will cause the drink-with-milk producing mechanism to engage; at the same time it will share a commonality of property with any other input state that also means 'drink with milk' since those will cause the 'drink with milk' producing mechanism to engage as well). Because these are physical properties, they meet all three conditions of syntactic properties. These shared properties could be of any kind, they could be as in the drinks machine example, the "squareness or redness, respectively of the token in the slot". The notion of syntactic property as used by Davies is austere indeed.

3 The Generality Constraint and the Structure of Thought

The second part of Davies' argument takes its cue from Gareth Evans' theory of concept grasping. But concept grasping only enters the picture because we need to understand the structure of thought itself. For Davies, the content of thought is conceptualized content. To be a thinker, to be having a thought, is to have grasped the constituent concepts of that content (the content of a thought being some sort of composite of the concepts involved in the thought). The eventual road on this approach to propositional atitude states is clear: to be in a propositional attitude state would require one to be a master of the atomic concepts specified in a comprehensive description of the state.

The Evans generality constraint (one that any concept-grasper must satisfy)[10]

[10]"Any thought that we can interpret as having the content that a is an f involves the exercise of

(***GC***) requires that:

> To entertain the thought that object a is F a thinker must have the concept of being F. If a thinker has that concept, then the thinker knows what it is for an arbitrary object to be F. So if a thinker thinks that a is F and is able to think about the object b, then that thinker is able to entertain the thought-to frame the hypothesis-that b is F.

Clark[11] provides another reading of ***GC*** involving the appropriate application of a concept. In his reading, to know the concept that vindaloo curries are spicy (All Vs are S) one must be able to apply the concept spicy in any case where its use does not involve a category mistake.

Now, the assertion that contents of thoughts have syntactic structure is made plausible by the existence of an analogy with a closure condition on the semantic content of sentences of natural languages: if there is a sentence that means a is **R** to b, then there is one that means b is **R** to a[12]. Much in the same fashion, the domain of thought contents made available to a thinker is subject to a similar closure condition: they are closed under permutation and recombination of constituent concepts. In fact, this much should follow from the ***GC*** itself: if a thinking agent can be said to be thinking that a is F and b is G then we can speak of her as having the capacity to think that a is G and b is F. Similarly for the converse of relational statements. So thoughts have semantic content (this content being of a kind that is subject to the ***GC***) and syntactic structure. But this doesn't take us to a ***LOT*** right away. States can have semantic content (under some appropriate behavioristic description: such and such action means such and such) without having constitutent syntactic structure. But the presence of an analogical closure condition does constitute a reasonable increase in the plausibility of the claim being made.

In the next step of his argument, Davies translates the ***GC*** into a requirement for what Clark terms "functionally localized inner states" with his notion of causal systematicity and the idea that folk psychology is committed to some form of causal systematicity underlying semantic patterns.

Why would ***FP*** be committed to the idea that causal systematicity underlies paradigmatic semantic patterns? Consider the following: we present a potential

---

an ability-knowledge of what it is for something to be an f-which can be exercised in, for instance the thought that b is f". Evans Gareth, *The Varieties of Reference*, Oxford, Oxford University Press, 1982, page 103

[11]Clark, Andy. *Associative Engines*, page 196

[12]Perhaps not. It has been pointed out to me that some strange examples can be constructed, Consider: (i) Tarski uttered "Snow is white" and (ii) "Snow is white" uttered Tarski

believer/desirer with "Tom likes lasagna" , "Dick likes lasagna" and "Harry likes lasagna" . If the believer responds with "Tom likes an Italian dish" , "Dick likes an Italian dish" and "Harry likes an Italian dish" then the believer is exhibiting a certain pattern; we would be inclined to say of the believer that she is a master of the concept 'lasagna'. The agent at hand understands that from the fact that Tom likes lasagna, it follows that he likes an Italian dish and so on for Dick and Harry. So in order to know that Tom likes an Italian dish, our agent must have grasped the concept, lasagna. She must know that lasagna is an Italian dish. The question then is whether this pattern is indicative of a causal mechanism at hand since just the presence of a input output pattern like the one above would not be enough to prove the presence of an underlying causal systematicity. What Davies needs to argue is that ***FP*** is committed to the idea that causal systematicity underlies the patterns shown above. Davies takes this as following from the ***GC*** itself[13]:

> A thinker who has the thought that a is F appreciates that from this thought it follows that a is H, say; and he also appreciates that from the thought that b is F it follows that b is H...It is not just that there is an input-output pattern in the inferences that the thinker is disposed to make. The two inferences are manifestations of a common underlying capacity; namely, mastery of the concept of being F.

Consider the following weak argument for the assertion above: if a thinker can be credited with the thought that a is F and the thought that b is K then that thinker can be said to be in possession of the conceptual resources for thinking that a is K and b is F. The reason this is likely to be unconvincing is that this sort of case could be met with three autonomous, concurrent states: a is F, b is K, a is K and b is F. Instead, the fact that inferential transitions can be made from one thought to another and the fact that thinking is precisely these sorts of inferential transitions is the clincher that ensures the presence of a "common underlying capacity". This capacity of course, meets the causal systematicity requirement and is found in the grounding of these inferential transitions by the mastery of the concepts concerned.

To accept the ***GC*** for a class of thinkers is to accept the causal systematicity requirement for that same class. But this requirement is not met by most connectionist models. As Davies puts it[14]:

> The question is whether the input states of a [connectionist] network have

[13]Davies, Martin. Concepts, *Connectionism and The Language of Thought*, page 243
[14]Ibid, page 247

> syntactic structure relative to the standard or classical semantic description of what the network is doing.

Connectionist models however, utilize context sensitive representation; no one particular pattern of hidden unit activation corresponds to single conceptual-level content. We get a variety of patterns instead, the difference amongst which is established by semantic context[15]:

> ..relative to a semantic description..the input states of a connectionist network with distributed representations will not have a syntactic description.

For a connectionist network[16], the input state with semantic content 'coffee in a cup' produces the output state with semantic content 'warm drink in a cup'; the input state with content 'coffee with sugar' produces the output state with content 'warm drink with sugar' and so on. While the output might be same, the internal representation of 'coffee' in the network will be different in both cases[17]:

> the content of connectionist mental representations..[is] semantically holistic.. the constituent parts of a connectionist mental representation..have no discrete semantic properties..they are wholly context-dependent.

Since the constituent parts of a connectionist mental representation have no discrete semantic properties, they lack "context-independent abstractions" for semantic elements such as 'coffee'. The internal states of a connectionist system that mediate transitions between "Tom likes lasagna" and "Tom likes an Italian dish" may only show a group-wise resemblance to those involved in the transition from "Harry likes lasagna" to "Harry likes an Italian dish". It is never going to be the case that we can say of the network that there is a common mechanism that is responsible for these transactions as opposed to more classically structured system that would channel them all through a single "logical bottleneck". We can't say that an inferential rule is being followed by the network.

So, ***FP*** appears to be committed to causal systematicity, connectionism isn't[18]

[15]Ibid, page 248- 249

[16]I apologize for not providing a more detailed description of connectionist, *parallel distributed processing* or *neural* networks.

[17]O'Brien, Gerard J., "The Connectionist Vindication of Folk Psychology", in *Folk Psychology and the Philosophy of Mind*, Christensen and Turner eds., Hillsdale, NJ, Lawrence Erlbaum Assoc. Publishers, 1993, page 376

[18]A stronger condition than this can be constructed since some connectionist models are *incompatible* with systematicity

and therefore the conditional thesis of the conceptual modularity[19] argument follows: if connectionism is true, then ***FP*** is false and with it so are all the claims that it makes about the existence of an opulent array of beliefs, desires and other mental states.

4 Concepts as Skills

Clark's defense of ***FP*** begins with a demand for a clearer specification of the common causal mechanism so dear to Davies' argument. Its not going to be clear as to what ***FP*** is committed to unless we get some sort of handle on what sort of common causal mechanism this is supposed to be. Maybe the sort of common causal mechanism required by ***FP*** is one that connectionist nets could provide, given a flexible enough reading. What form should it take? Should there be a physical state (***PS***) that corresponds to any inference involving a concept F ? Or should there just be a way of describing and grouping (***DG***) all brain states that are involved in inferences depending on F ? Or should there be a commonality revealed only by a particular mode of description (in this case, ***FP***)?

One can almost immediately rule out the last form as being unpalatable for proponents of realistic causal factors; it doesn't seem likely that someone committed to finding causal explanations in ***FP*** will agree that the so-called causal machinery is only revealed by a mode of description. What about the first two? Consider ***PS***. Surely no one is suggesting that exactly the same physical resources be deployed in every cognitive episode. Clark's example of computation[20] that ensures the same result without using the same physical resources makes that clear. That leaves us with ***DG***. A cluster analysis of the network at a particular level of description i.e one that grouped elements of the network into equivalence classes would do the trick. That is, such an analysis would result in clustered patterns of activity in the net being described by ordinary propositions. Now, Davies allows that a scientifically vindicated method of description of connectionist networks could arise that is permissible and that is compatible with ***FP*** as well. So this might be viewed as a concession to Clark. Not for long though.

This concession doesn't solve Clark's problems in the way he would like (promissory notes in philosophy never seem to). It would be entirely possible that such a

[19] Clark terms this the conceptual modularity argument since it requires a "processing level commonality between inner states which share a conceptual component".

[20] Clark, Andy. *Associative Engines*, page 199

method of description never becomes available to us. Stich and Warfield point out[21] that all sorts of functionally equivalent networks with radically dissimilar structures can be designed and constructed that will not exhibit the "same clustering profile". There is no guarantee that a clustering profile applied to one network will give us the same results when applied to another one. So for Clark, the question then becomes[22]:

> How fragmented can our inner representations turn out to be without compromising the folk-psychological discernment of unities and commonalities? The answer I think, is simple: They can be almost *arbitrarily fragmented* [my emphasis], for the folk unities are properly grounded in folk properties and folk individuative properties.

This is a bit of a move, of course, and one likely to meet with stern opposition. Nothing in standard descriptions of ***FP*** suggests any such looseness about the ontological status of ***FP*** entities. ***FP*** is committed in a fairly straightforward fashion to the existence of what Clark terms "folk solids". What else could the "folk- pyschological discernment of unities and commonalities" entail? Nor is there any ambiguity in the demand for a common causal mechanism required by the ***GC***. It is not quite obvious either, as to how Clark derives the authority to make the claim about the "proper" grounding described above. What sort of claim is this? What does this proper grounding consist in?

Clark suggests that the ***GC*** when further unpacked (finding echoes of Ryle in Evans) points the way out; the ascription of the concept F to a subject is the ascription of a state that is a complex of the exercise of several distinct conceptual abilities. A thinking agent that displays mastery of the concept 'lasagna' is displaying a proficiency comprised of lots of fairly complex skills. Each ability or skill described above (much like Ryle's examples of tripping or the clown) corresponds to a very loosely organized set of "inner computational states". These states are not unified by a common neurological (ruling out ***PS***) or psychological description (ruling out ***DG***). Instead, for Clark[23]:

> ..these several scientifically disunified inner states combine to constitute a single skill because their combined presence enables an agent to successfully negotiate some macro-level domain which interests us in virtue of the form of our daily life.

[21]Stich, Stephen, Warfield, Ted. "Do Connectionist Minds Have Beliefs", printed in *Connectionism: Debates on Psychological Explanation*, Volume Two., Oxford, Blackwell, 1995, pp395-411
[22]Clark, Andy. *Associative Engines*, page 201
[23]Clark, Andy. *Associative Engines*, page 203

These skills are "disunified" since they only pop into view under the folk psychological gaze. It is the existence of a folk domain ('golf') which legitimates our gathering together of this particular domain set of subskills (physical, sensorimotor, verbal, imagistic) under the label of 'skill at golf'. If no one played golf in this world, how would the abilities that we term 'bunker escape skills' ever emerge? When they would , they would be a complex of several factors already present in the agent; some of them physical (wrist strength, hand-eye co-ordination), sensory (visual ability) and so on. No cognitive science program is going to reveal the presence of some common factor which is constitutive of a skill in golf and which is present in all instantiations of golf playing. The important point for Clark is that ***FP*** has something else in mind when it talks about concepts[24]:

> The root of the more traditional claim is just the tendency to uncritically identify the idea of a concept with the idea of an *occurrent state*.

This is what Clark gives up. It's probably a good thing too, since proponents of ***FP*** have been notoriously hard pressed to explain what this occurrent state would be like. A concept is now identified with something more nebulous; an emergent property[25] of sorts, one that cannot be further reduced to atomic constituents. To ascribe a concept to someone is to ascribe them a certain knowledge, a set of skills[26]:

> Talk of concepts elevates mixed bags of inner subskills into folk psychological unities. It does so only against a background of human needs and institutions (just as in more mundane cases, like that of skill at golf). Thus, to possess one such skill may positively require you to embody a varied and shifting panoply of inner states whose unity is not displayed at any further, level of description.

What determines the individuative properties of concepts is dependent upon several, not too clearly defined factors and Clark doesn't have much to say in this regard other than to wave his hand in the direction of the notion of pragmatic utility playing a big role. Which is not too bad a thing to say about any philosophical approach,

[24]Ibid, page 204

[25]A similar view is expressed by Dennett in "Three Kinds of Intentional Psychology", printed in *Folk Psychology and The Philosophy of Mind*, Hillsdale, NJ, Lawrence Erlbaum Assoc. Publishers, 1993, page 139: "the relation any vehicle of content [in this case, a concept]..must bear in order to have content is not just a relation it bears to other similar things but a relation between the token and the whole life-of the organism it serves and that organism's requirements for survival and its evolutionary ancestry." Once you give up the idea of a concept as a state then you also give up the idea that for that state to have content is a matter of it being in some relation to other states.

[26]Ibid, page 205

except that one would have hoped that Clark, after ensuring the 'proper grounding' of concepts in 'folk individuative properties' would have had something more to say in this regard.

5 Concepts as products of Human Cognition

Clark's thesis amounts to dismissing the idea that concepts act as atomic structures in any meaningful sense of cognitive activity. With this reading, the constraints placed by the ***GC*** seem less onerous. Concepts are, instead, products of human cognition; entities whose "integrity is dependent on macro-level folk practices". But this doesn't mean that ***FP*** is false[27]:

> The folk explanations simply occupy a different arena.

Besides simply shoving folk explanations over to a different intellectual arena, Clark's move also makes the ontological status of folk psychological entities dependent on ***FP***. A sort of instrumentalistic reading if you will, in that beliefs and desires now exist to the extent that they facilitate discourse and allow us to engage in mutually profitable behavior. Folk psychological "solids" exist because ***FP*** does; because ***FP*** finds it expedient to postulate the existence of these and because quite simply, people that utter ***FP*** platitudes find it plausible and profitable to do so.

To back this claim up, Clark suggests that work in robotics points to the fallacies of an overly intellectualist explanation of action and behavior[28]:

> the mentalistic vocabulary is simply inadequate and misleading as a design specification for a robot system.

So, the idea of a layer of representations lying between input and action modules in robotic doesn't work and is the artifact of an overly rationalistic model of intelligence (shades of Ryle again). At some point, brute action takes over. Clark considers too, the possibility that work in this area will show that representations of inner states in other systems will not show the kinds of concepts envisaged by the folk picture. But this is still a desirable state of affairs since it implies an understanding of the mind in more advanced fashion that is possible by current "biological neuroscience". Clark's reading of Evans has enabled a move to a position that shifts the focus of the debate somewhat[29]:

[27]Ibid, page 207
[28]Ibid, page 207
[29]Ibid, page 214

> By treating concept ascription as ascription of cognitive skills (and not unitary states), and by treating propositional-attitude explanation as a way of picking out patterns in actual and counterfactual behavior (rather than as necessarily picking out discrete, causally active inner items which participate in the push and shove of creation), we put considerable distance between the folk framework and issues concerning the innards of organisms.

This moves the debate from being an ontological one to one about which kind of explanation suits our inquiries the best; if that sort of explanation needs to postulate "folk solids" then so be it.

Clark's position is of course, getting closer and closer to Denett's position expressed below[30]:

> Any object, whatever its innards that is reliably and voluminously predictable from the stance is in the fullest sense of the word a believer. What it is to be a true believer, to have beliefs and desires, is to be an intentional system.

The standard objections or counterexamples (Giant Lookup Tables, Quantum Fluke Beings and so on) to Dennett's position can be levelled against against Clark as well. But for Clark there is a crucial difference: Clark is a realist about believers as opposed to being a realist about beliefs (an approach that could lead into messy ontological disputes). This position leads him to two further constraints on the class of beings for whom the language of ***FP*** is suitable as a mode of description. These two requirements form a pair of necessary conditions that all believers must satisfy.

5.1 The Requirement for Normative Depth

Clark's normative depth requirement is designed to be inclusive enough to allow for a wide variety in the internal structures or organization of believers and yet rigid enough to block constructs such as the Giant Lookup Table. This should serve to highlight a supposedly unique capacity believers seem to possess; the ability to judge our own performances as meeting certain standards[31]:

> This ability to recognize our own guiding cognitive commitments and to *judge our own judgings* [my emphasis] as correct or mistaken, is I suspect,

[30] As quoted by Clark on page 215

[31] Ibid, page 217

central to the idea of moral agency.

Is this judging of judgings supposed to be only one layered? Or should we able to carry this on indefinitely on backwards? Is there a a particular class of judgings that we consider the ones to be judged? Should all our judgings be potentially judged? Now, this is obviously an incorrect reading of Clark; an affirmative to the last question is probably all that Clark has in mind. All believers have the capacity to judge their own judgings; they don't actually judge their all of their judgings.

Other questions abound however. What kind of requirement is this? Is it stipulative? If so, how does Clark derive the authority to do the stipulation? Is there any particular reason to accept this constraint as a reasonable stipulation? Is the constraint descriptive? It can't be; there is no data to back this claim up, and besides, these are supposed to be constraints. Is this requirement supposed to capture the intuitions of a large number of people? If so, the lack of data to support this claim is problematic. Is this requirement supposed to convey some information about what people mean by 'believers'? By what they claim about believers? So this must be along the lines of another conceptually necessary requirement. But this as will be pointed out again, doesn't really get us very far. Is this a prediction? That all believers will be found to be capable of judging their own judgings? But, again, if that is the case, where is the data to back this sort of empirical hypothesis up?

The requirement for normative depth is far too strong or restrictive in ruling out those systems that are perfectly capable of revising their cognitive performances without necessarily going through the act of judging their own judgings. Examples such as household cats and the average laboratory rats then would not possess beliefs according to Clark's requirement. But is someone going to agree that their cat doesn't believe that the milk lies just beyond the door[32]?

Other questions remain. What kind of data would be relevant ? Some suggestions: empirical research could show that our ordinary locutions don't have the built in constraints that Clark claims or it could show that ordinary folk *never do* this sort of judging. But why would this last be relevant? Why shouldn't it just lead to an eliminativist conclusion? Clark admits that empirical research could make a mess of these constraints, so maybe one could classify his constraint as one that is open to revision. For example, if we were to judge a particular being as a believer based upon our observation of its revision of past cognitive performance and then discovered that this was not traceable to any internal process, we would have to revise our

[32]These examples are due to Stephen Stich, *Deconstructing the Mind*, New York, Oxford University Press. 1996.

classification[33]:

> A total lack of any internal mechanism by which traces of earlier outputs can be stored and later reassessed undermines the surface image of the being as able to judge that it has misapplied the concept.

But if this were to be the case, then Clark is coming back to the "innards" issue again. Surely, if a system has judged that its judgements were flawed, why would it matter (to Clark) whether the system has an internal property that be so localized that there exist an "internal mechanism by which traces of earlier outputs can be stored and later reassessed" . This is a curious requirement to lay down for someone who has just been trying to make a case for moving away from the innards of organisms. Clark makes a fairly explicit demand above. If Clark were to stick to his guns, he should be satisfied with a system behaving as if it were really revising its performances. Would some sort of statistical evidence work? That is, "All believers found so far have also been found to be capable of judging their own judgings"? In one sense, this shows a problem with instrumentalist approaches; they can be half hearted at times. Clark is more a of realist about the internal machinery that ***FP*** is committed to than he would like to admit.

5.2 The Consciousness requirement

Consider next, Clark's requirement of consciousness for a class of potential believers/desirers:

> For someone to able to grasp a concept it is then necessary that the person have conscious mental experiences that involve that very concept.

This particular requirement arises as before with the need to get away from counterexamples to an operationalist position. But it is not immediately clear that this requirement does what it is supposed to do; many of the criticisms raised in the previous section can be raised against this particular requirement as well. What sort of constraint is this? What would make it true? Where does Clark derive his reasons for making this claim? Why does Clark think that it is conceptually impossible that a person not be able to grasp a concept without having conscious mental experiences that involve that concept? In this step he has gone from giving us constraints on class of believers to a conceptual necessity associated with believers. What bearing this sort of conceptual necessity is supposed to have in an empirical theory is not clear.

[33]Clark, page 218

In effect, with the above requirement, Clark is stating that a sentence of the form "All believers are conscious" is analytic. What makes this sentence analytic though? Is its analyticity based upon some the *definition* or *meaning* of the words 'believer' or 'conscious'? There certainly is no argument to that effect yet. Is there even a paradigmatic definition available for either of these? Is Clark basing this conceptual necessity on his intuitions? If so, how come his intuitions get to decide what is analytic and what isn't? Surely analyticity (much in disfavor these days) doesn't get predicated of a sentence unless there is compelling reason to think so. Paradigmatic analytic sentences such as "All bachelors are unmarried" are deemed analytic (in pre-Quinean fashion) as a result of the meanings of the words 'bachelor' and 'unmarried'. Is there any such linkage of meanings to be found in the sentence "All believers are conscious"? Clark thinks that this analyticity is ensured by the meaning of the sentence "X has grasped the content of the concept C". For Clark, this sentence means "X has a conscious mental experience which involves the concept C'. But nothing by way of data or commonly shared intuitions leads us to believe that there is any equivalence of meanings in any sense (the type required to ensure analyticity in the event that one could even take analyticity for granted) in this sentence.

Moreover, doesn't this particular requirement take Clark back to the innards of organisms? For isn't a empirical theory of consciousness going to tell us something about the innards of organisms? An empirical theory of consciousness is presumably not going to be a instrumentalist one. Can a viable theory of consciousness really tell us that "To be a conscious entity is to behave like one"? No, it would lay down certain conditions on what a conscious being would be like. Maybe it would postulate the existence of a particular mental states like the Higher Order Thought theory or a physical state like the Crick and Koch version. If that was the case, and being conscious was conceptually necessary for being a believer, then wouldn't it follow that being a believer would necessitate the existence of a particular mental or physical state in the potential believer/desirer? So that being a believer would in fact, entail the commitment to a particular occurrent state. Is this what Clark has in mind? More problematically, Clark now needs some account of conscious experience that involves each and every entertainable concept. Is such an account likely?

Its hard to avoid feeling that Clark has postponed the mystery just one step further. Its not quite clear what constitutes a conscious mental state or conscious experience as yet. In light of this particular state of affairs the usefulness of a constraint on a class of thinkers that uses this concept is questionable.

6 Conclusion

Clark grants Davies much of his argument; certainly their reading of what ***FP*** is committed to and what connectionist nets are committed to is very similar. The only possible difference between the two is their unpacking of Evan's generality constraint. Evence)? this isn't too radically apart. Davies does after all consider and dismiss, the 'concepts as bags of skills' reading of Evans. Clark uses his to make the case for a Dennett style analysis; Davies is led to a different conclusion. But Clark's commitment to his operationalist position seems half- hearted as evinced by his falling back on the results of internal investigations.

The criticisms of Clark's move to Dennett's position will echo those previously made of any instrumental approaches to intentionality; what is the status of the claims it makes about **FP** or the state of being a believer? What is the evidence for these claims (if they need evidence) ? Clark's further constraints are puzzling and its not entirely clear that this is the most convincing rejection of eliminative conclusions. Its not too plausible that a strategy that relies upon a instrumentalist treatment of folk entities and two "conceptually necessary" requirements of believers is going to convince very many people that the ontological status of ***FP*** entities remains sacrosanct. Indeed, one wonders how much any of this defense will come as comfort to proponents of ***FP***.