Implicit knowledge

How thinking of concepts as sets of instances throw new light on the sciences and other expert disciplines.

ABSTRACT

Existing cognitive science theories of concepts have difficulty dealing with phenomena associated with the growth and dynamics of mature scientific disciplines. This thesis attempts an initial cross fertilisation between promising theories of concepts and concept dynamics in the fields of History and Phiosophy of Science (HPS) and Sociology of Scientific Knowledge (SSK) on the one hand, and existing cognitive science perspectives on the other. To this end, certain aspects of contemporary Cognitive Science theories of concepts are briefly reviewed, with an aside regarding connectionist theories of the mind. Two seminal works of the sociology of scientific knowledge are also reviewed, Barry Barnes "TS Kuhn and Social Science" and J.R Ravetz's "Scientific Knowledge and its Social Problems". The social finitist ideas of Barnes, and the accounts of the implicit knowledge of Ravetz are fused to create the idea of concepts as sets of instances. This in turn allows for the explanation of a number of phenomena either unexplainable or difficult to explain by more traditional, computational models of the mind. Examples include contextual effects on learning and the anomalous nature of expert skills. Finally, it is suggested that these findings apply beyond the mature natural scientific disciplines to all knowledge-based traditions of practice. Two sets of reflections on such traditions-Zen training in the military arts and 'MIT AI Hacker' Culture-are offered as openings to future research.

To Trang, Mum, and Dad,
For giving me the instances I needed
to be who I am.

Acknowledgements

Thanks go to Dr J, for a truly excellent job of supervising me, even when I did decide to do crazy thiings like revues, to Dave, for being the idea-bouncy guy, and for keeping it real in letting all his net connection belong to me, old skool! To my proofreaders, Louise, Dave, and Trang, for keeping me honest just as I thought I was done. Thanks also to Lisa, for reminding me when I'd left dirty dishes in the sink and thus helping to keep Trang sane, to my parents, for dealing with me not seeing them for months on end and most of all to Trang, for making sure I did important things like eating and showering and in general giving me far more support than I had a right to expect.

Preface

This project began with a vague feeling of uneasiness, with a dissatisfaction with the ideological differences between those who study science from a social or sociological point of view, and those who study science from the philosophical, or logical point of view. The dissatisfaction is with people who refuse to listen to ideas because they come from the other side of the fence, with philosophers who will not admit that implementation of ideas is necessary at some point, and with sociologists who will not admit that implementation is not all there is.

I use the word ‘implementation’ because I am a child of the computer age. For me, using computers is as natural as breathing. Why is this relevant? Because with modern computers, you have a system built using philosophical ideas of universal machines, that at the same time manages to exhibit strangely complex behaviours. The problems that arise with today’s computers arise out of a system of such vast complexity that, it seems, emergent behaviours of the sort spoken of by sociologists become the norm.

It was with these feelings in mind that I began my honours year. The first project I was assigned was to read two books, Jerry Ravetz’s *Scientific Knowledge and its Social Problems*, and Barry Barnes’ *T S Kuhn and Social Science*. After having read both it seemed that these two authors had been trying, in their own way, to do exactly what I wanted to do, that is, to look at what actually happens in the dynamics of science. Furthermore, that the two books seemed to fit together very well, with the strengths of one respectively relieving the weaknesses of the other. Barnes’ strength lay in his attempt to sociologise Kuhn, to make him more social by, paradoxically, introducing some more philosophical underpinnings, and Ravetz’s strength lay in his highly detailed description of the actual processes of scientific inquiry.

It seemed to me that this type of fusion was exactly what I had been seeking, and so I set about trying to relate the ideas in the two books, extending them and updating them with current cognitive science. This was the germ of my project, and some of my reasons for doing it.

As I studied, another thing that became clear to me was that science is, and should be treated as, an expert discipline, similar in many respects to other non-scientific expert disciplines. Because of this feeling I began looking at other expert disciplines, and immediately found one I was already a part of: ‘hacker’ culture. As I will explain later, hacker is not used here in the perjorative media sense, but in a sense that implies mastery in the same way that a blacksmith implies mastery of that craft. In this case, the craft is that of maintaining and programming computers. Not only a child of the computer age, I am also a computer programmer to some degree a member of the hacker community. This is important, because it enables me to see another expert discipline through the eyes of a member, and extract illustrative links to the processes of science.

This project was always going to be hard to compress into an honours thesis, an I also felt somewhat presumptuous in dealing with such big issues at such an early stage in my career, but I felt that it was well worth attempting, even if only to provide pointers to areas that require much more in-depth study. But, as always, one must begin somewhere. In the words of the Zen master Dōgen[1],

Dig the pond
Without waiting for the moon.
When the pond is finished,
The moon will come by itself.

[1] From Cleary (op cit 131), p 43

Chapter 1 – Introduction

Why another *theory of concepts?*

There are and have been many theories of the working of concepts[2]. From Hume and Kant to the modern theories of cognitive science and the influence of computers, the constituents of our thinking are things that have (ironically enough) been the subject of much thought.

So why attempt to upset some or all of this? I suppose one could blame the tendency of youth to rebel against its elders, but I would like to think that it is because I do have something useful to contribute. In much of today's thinking about concepts and the way the mind works, the disciplinary divide is too large. Sociologists and historians of science seem to be so caught up in an ideological battle with the philosophers and methodologists of science that they both fail to see that each could contribute something to the other. From sociology and history, we can get an awareness that things happen in a real world that doesn't behave in an ideal fashion, and that this should be embraced and dealt with in thinking about the world. From philosophy and methodology, we can realise that the ideas of logic, abstraction, and a fascination with rigour are very powerful, and should be recognised as such.

To summarise before we begin: the primary problem with many of the past theories is that they make no effort to explain human conceptualisation in terms of a *process*. Kuhn pointed out that the sciences could be thought of as fields or disciplines in a process, dynamic, sometimes growing, sometimes shrinking entities. Part of the reason for the dynamic nature of science may lie in the dynamic nature of human concepts. That is, it would seem natural that science would change if the concepts being used by the scientists are capable of change.

---

[2] There is certainly a vast literature about the history of theories of concepts. One example would be Weitz, M. *Theories of Concepts: A history of the Major Philosophical Tradtition*. Routledge, London and New York, 1988.

Because of this, and because of what was noted previously about ideology and professional interests in the 'meta-science' disciplines, I will be traversing a number of disciplines in this thesis. I will be looking at cognitive science, as modern cognitive science is the field where theories of concepts are most often discussed. Contrasted to this will be the sociology of scientific knowledge, or SSK. From Kuhn to the present day, a major argument in favour of SSK is the importance of processes in science, but this is just one of the facets of SSK I will consider. Hopefully, by looking at the two fields in contrast, it may be possible to throw new light on each, to deepen our understanding of both. I also intend to contrast some key areas in cognitive science, between what Daniel Dennett[3] called 'High Church Computationalism', or what Jerry Fodor[4] calls 'Representational Type Mapping', and the newer upstart child, connectionism. Both of these sub-divisions will benefit from some time spent thinking about some of the issues raised by members of the SSK community, and how those relate to disagreements between themselves. From all this travelling, I hope not to make the reader tired, but instead to create a sketch theory of concept dynamics. I also hope to show in some 'case study' reflections how considering concept dynamics independently of science can lead to a greater understanding of science as an expert discipline, related very closely to non-scientific expert disciplines.

However, some explanation of the previous and current ways of thinking about concepts is required. I will briefly look at the history of theories of concepts, with particular reference to Cognitive Science theories. After this I will look at some constraints on theories of concepts enunciated by Jerry Fodor. After this I will look at and answer some objections Fodor has to the use of similarity relations between concepts. This has some important implications for Barnes and Ravetz, namely that Barnes and Ravetz are both

[3] In Dennett, D.C. *Brainchildren: Essays on designing minds*, Penguin, St Ives, 1998

[4] In Fodor, J.A. *Concepts: Where Cognitive Science went wrong*, Clarendon Press, Oxford, 1998

founded on ideas of open, revisable concepts, and either explicitly, in the case of Barnes, or implicitly, in the case of Ravetz, rely on similarity relations.

Two important things I am *not* doing in any of this are:

1. Implying that mind is in some way ultimately mysterious and non-understandable.
2. Suggesting that what I have to say is the be-all and end-all of the subject. I am only aiming for a sketch theory at the moment, and while I know I have some valid points to make, I know that they are most certainly not in their final form.

Having clarified my stance to some extent, it is time to begin the history lesson.

Modern theories of concepts: Classical Cognitive Science vs. Connectionism

There have been many theories of concepts in the history of philosophy. Indeed, any theory of mind relies to some extent on a theory of what the constituents of the mind are. Plato, Descartes, Kant, and Turing, are but some of the proponents of theories of mind that entail certain theories of concepts. However, it is since the advent of digital computers that interest in theories of concepts has really exploded. The reason for this? Computers create the possibility of actually building the systems theorised about, and seeing if they work as they are supposed to. That is, Turing's beliefs about the equality of computation and thinking have shaped what has now become cognitive science, precisely because it became possible to build machines that would act in the way described by Turing's Universal Machine, machines that could perform operations on symbols and thus mimic any logically possible process.

Dennett describes the basic aims of cognitive scientists as:

> the explanation of the aboutness or intentionality[5] of mental events in terms of systems or organizations of what in the end

[5] That is, the way that mental events seem to express intentions.

> must be brain processes. That is, they take it as agreed by all parties to the discussion that what we want, in the end, is a materialistic theory of the mind as the brain.[6]

As Dennett points out, the more interesting debates are concerned with how we get to this goal. For the purposes of this work, the different schools' ideologies lead to very different theories of concepts.

Dennett maps the cognitive science arena in terms of two poles, the Eastern and Western, in reference to remarks made by Fodor about how the Massachusetts Institute of Technology (MIT) is the Eastern pole, because everything in cognitive science that does not come from there is Western, and thus slightly strange. Dennett calls the Eastern Pole "High Church Computationalism"[7] (HCC), and the Western, "Zen Holism". The adherents of HCC claim, according to Dennett, that 'intervening between folk psychology[8] and brain science will be at least one level of theory quite "close" to the high level of folk psychology that is both "cognitive" and "computational".'[9] Dennett gives three 'defining dogmas' of HCC:

1. Thinking is information processing.

2. Information Processing is computation (which is symbol manipulation)

3. The semantics of these symbols connects thinking to the external world.[10]

These principles tend to give rise to theories that are philosophically rigourous, and also to give some abstract concept-token primacy as being

---

[6] Op. Cit 1. Ch 13, "The Logical Geography of Computational Approaches: A view from the East Pole", p 216

[7] I would be more inclined to use the slightly less polemic Classical Cognitive Science, and will do so after this discussion of Dennett.

[8] Dennett and Fodor both use this term. It seems that they both mean the naïve view of psychology shared by those who haven't really thought about it, that is, the personal theories that we all create to deal with other people.

[9] Op. cit. 3. p 217

[10] ibid. p 217

what "having" a concept consists in having.[11]

Dennett defines "Zen Holism" as the contradiction of these principles, namely that 'thinking is something going on in the brain all right, but is not computation at all; thinking is something holistic and emergent – and organic and fuzzy and warm and cuddly and mysterious.'[12] However, as I said before, it is not my intentto argue that mind is mysterious in this way. I think mind is mysterious in an explicable way, and thus I will not consider Zen Holism, except in this consideration of Dennett.

Zen Holism is not Dennett's creed; it seems from the foregoing quote that he finds it slightly silly. What Dennett is more in favour of is "New Connectionism" (again, his term). He summarises their view like so: 'they are looking closely at neural architecture and trying to model much closer to the brain than to the mind.'[13] Later he gives a more thorough summary of common properties of connectionist models:

1. "distributed" memory and processing, in which units play multiple, drastically equivocal roles, and in which disambiguation occurs only "globally".

2. no central control but rather a partially anarchic system of rather competitive elements.

3. No complex message-passing between modules or subsystems

4. A reliance on statistical properties of ensembles to achieve effects.

5. The relatively mindless and inefficient making and unmaking of many partial pathways or solutions, until the system settles down after a while

[11] See op. cit. 4, Ch. 1 for some ideas from Fodor's theory of concepts, and see Fodor's refutations of other types of theory, in chapters 3, 4 and 5 for more information on this.

[12] ibid. p 217

[13] ibid. p 225

> not on the (predesignated or predesignatable) "right" solution, but only with whatever solution or "solutions" "feel right" to the system.[14]

One can immediately see the differences Dennett is trying to explicate. HCC is interested in semantics, symbolic manipulation, and computation in the 'Turing machine' sense of the word.[15] Indeed, HCC seems in large part to be descended from Turing's universal machine. Connectionism, on the other hand, is more holistic, though not to the degree of Zen Holism. It emphasises the importance of individual units that are not capable of much by themselves, but when grouped together become capable of far more.[16] Theories of concepts by those with connectionist ideologies tend to emphasise the importance of the connections between the units in storing everything, including concepts. Also, connectionist theories tend to emphasise the properties enumerated by Dennett when considering concepts as well.[17]

It can be seen that the field of cognitive science is certainly a field that must be considered if one wishes to write intelligibly about concepts. To further this theme, I will now look at some constraints on theories of concepts put forward by Jerry Fodor, with a view to using them to evaluate my own ideas at a later stage.

Some suggestions from Classical Cognitive Science – via Jerry Fodor

In his book, Fodor's objective is to advance his theory of concepts. However, as he begins, he gives five conditions on a theory of concepts. Although there

---

[14] ibid. p 227

[15] That is, computation as in the idea created by Turing, of the universal machine capable of mimicking any logically possible machine by operating on symbols that are themselves instructions to the machine. I mentioned this above.

[16] This seems similar to R. Buckminster Fuller's ideas about *synergy*, that is, relationships where the whole is greater than the sum of the parts. For explication of this see Fuller, R. B. *Synergetics: Explorations in the geometry of thinking*, Macmillan, New York, 1975.

[17] Dennett speaks of this again in 'Mother Nature versus the Walking Encyclopedia: A Western Drama', ch. 2 in Ramsey, W, Stich, S.P. & RumerlHart, D.E, *Philosophy and Connectionist Theory*, Lawrence Erlaum Associates, Hillsdale, 1991. Another good survey paper is Boden, M. A., 'Horses of a different color?', ch. 1 in the same volume.

is the impression he would feel that this project is utterly wrong-headed[18], I think that his conditions raise valid points about what must be covered by any theory of concepts. I will thus consider these conditions after I have expounded my own theory as well.

> 1. Concepts are mental particulars; specifically, they satisfy whatever ontological conditions have to be met by things that function as mental causes and effects.

This is untendentious. All this condition says is that concepts are mental things.

> 2. Concepts are categories and are routinely employed as such.

By categories Fodor means that 'they apply to things in the world; things in the world "fall under them".'[19] Again, relatively untendentious, in the sense that it is very difficult to argue that concepts are not categories for experience in some way.

> 3. Compositionality: concepts are the constituents of thoughts and, in indefinitely many cases, of one another. Mental representations inherit their contents from the contents of their constituents.

The first thing to note with this condition is Fodor's use of the word 'thoughts'. By 'thoughts' he means 'mental representations analogous to closed sentences'[20], while concepts are 'mental representations analogous to open ones.'[21] Here Fodor seems to be using 'closed' and 'open' in a pseudo-linguistic way, i.e. to distinguish between 'complete' and 'incomplete' sentences, between sentences with all of the necessary structures, and those that are missing some. Thus Fodor thinks of concepts as being like words or phrases, and both CAT and NOT CAT are concepts. In this he is following the

---

[18] This would be because of his objections to similarity relations (see later section on this), and also because of his dislike of anything that smacks of relativism. See op. cit. 4, p 29 for an example.

[19] op cit 2, p 24

[20] ibid. p 25

[21] ibid. p 25

program I outlined earlier, of Classical Cognitive Science, in thinking of thinking as the calculation and manipulation of symbols.

The main point of this condition is compositionality. This is the requirement that concepts can be constituents of one another, as well as the constituents of thoughts. This is very important because it allows for the creation of concept hierarchies, which seems to be the way most folk psychology views concepts.

Figure 1 is an example. Here you can see a hierarchy of concepts under the broader concept of 'vehicle'. The point implied by this diagram is similar to that implied by any similar structure: the tree must have a root. That is, there must be a primitive concept, or a set of primitive concepts, that are at the root of the tree, are thus required to learn anything. To put it another way, there are required to be primitive concepts without which it is impossible to conceive concepts existing. A good example is equality. It is almost impossible to build an account of the acquisition of a concept of equality, as it is required for the comparison of concepts themselves. Due to the primitive nature of the concepts that tend to form the root of the tree, they are most likely to be innate. This brings us to the next constraint.

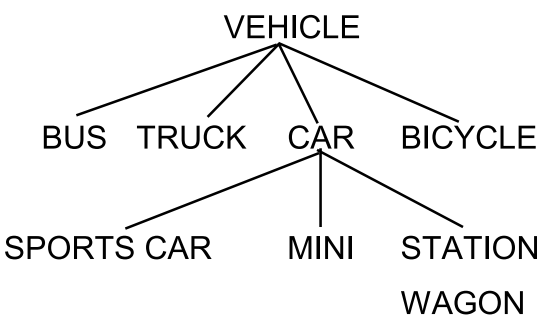

Figure 1- Conceptual Hierarchy

4. Quite a lot of concepts must turn out to be learned.

There must be *some* concepts that are not learned, or else no learning would be possible. However, the number of primitive concepts is very small, according to Fodor. He notes that 'people have Very Strong Feelings about what concepts are allowed to be innate'[22], but that there seems to be no rules

[22] ibid. p 28

as to what concepts are allowed to be considered as innate and what are not. Plainly, this is not an optimal state of affairs, and I hope to address this in my own fashion in the third section, Chapter 3 – The Importance of Implicit Knowledge.

> 5. Concepts are *public*; they're the sorts of things that lots of people can, and do, *share*.

This is the *publicity constraint,* perhaps the most important condition to consider in relation to this work. To put it another way, concepts are something that must be *shared* on some level. Fodor feels that, for this publicity constraint to be correct, concepts must be *identical* between people rather than just *similar*. But what does it mean for concepts to be identical?

Fodor uses the example of water. 'If everybody else's concept WATER is different from mine, then it is literally true that only I have ever wanted a drink of water, and that the intentional generalization "Thirsty people seek water" applies only to me.'[23] So what Fodor is saying is that, unless we speak of concepts in terms of identity relations when considering concept publicity, then we are speaking nonsense.

And so we begin to see what Fodor is driving at. Admittedly, his constraints on concepts are primarily designed to further his own theory, but I think that it is important to have considered them in this thesis.

It is important for two reasons. Firstly, on a 'big picture' level, as the requirement for identity relations is one of the ways Fodor defends against relativism. Obviously, if concepts have to be identical amongst people on some level, then relativism cannot be the case. Given that I am arguing for a more relativistic theory of concepts, it falls to me to justify my stance. Secondly, I intend to make large use of Barry Barnes' work, a keystone of which is an idea which he calls "learned similarity relations between finite

[23] ibid. p 29

clusters of events'. Obviously if I am going to have you, the reader, give my work any credence whatsoever, I must justify the use of similarity relations to some extent. This is the purpose of the next section, to justify to the skeptical reader why, given Fodor's reasons for accepting identity relations between concepts, I think we should instead accept similarity relations between concepts as the focus around which my argument will develop.

A philosophical debate: Identity relations vs. similarity relations

As mentioned above, one of Fodor's constraints on theories of concepts is that of *publicity*, that is, concepts have to be able to be shared. Fodor argues that only identity relations between concepts are justifiable, that is, that people must have the same concept-token on some level. Fodor believes that this is required in order to fulfill the publicity constraint. He also points to theories that try to fulfill the publicity constraint using similarity relations instead, and attacks their weaknesses.

Fodor's objections to Similarity Relations

Fodor's first objection is that similarity is difficult to define without a notion of identity. That is, it's hard to define things as being similar without an idea of two things being identical. This is important because, if concept similarity relies on concept identity, then obviously anyone trying to create a theory of concepts based on concept similarity rather than concept identity is bound to fail. However, I am of the opinion that concept similarity does not necessarily rely on concept identity, and will justify and argue this shortly.

This objection to similarity relations relies on the question: Are shared beliefs *literally* shared? Fodor considers that this raises an unavoidable dilemma.

> If he says that our agreed upon beliefs… …are literally shared, then he hasn't managed to do what he promised; viz. introduce a notion of similarity of content that dispenses with a robust notion of publicity. But if he says that the agreed beliefs aren't literally shared (viz. that they are only required to be similar), then his account of content similarity begs the very question it was supposed to answer: his way of saying what it is for

> concepts to be similar but not identical presupposes a notion of beliefs with similar but not identical contents.[24]

That is, the similarity theorists are merely pushing the problem up a level. Obviously this is insufficient, and does not answer the question.

Secondly, Fodor points out that theories involving the similarity of beliefs rely on a notion of partial overlap of beliefs, which is problematic if the beliefs are not identical between people. That is, I cannot say that your concept CAT overlaps with my concept CAT without the concept CAT being the same in some way between both of us. A caveat here: I do not really think that Fodor is intending to imply that all aspects of the concept must be the same. I get the impression that it is the concept *token*, that is, the actual mental thing that you think when you think CAT, that is the same, not its constituents.[25]

Why similarity relations?

Given all of this, why would I be making a case for similarity relations? Well, I think there are a few reasons to take similarity relations as the default option, rather than identity relations.

Firstly, there is the impossibility of 'clean' communication. That is, it is impossible, in the real world, to pass information between two people without some noise. I think that concept identity theories do not allow for miscommunication, or for the prevalence of misunderstanding so common in the real world, as opposed to the idealised world philosophy operates in.

Secondly, and most importantly, similarity relations allow for a simplified view of science *dynamics*, how science changes, and indeed how human points of view can change over time. If you wish to advocate identity relations, then it falls to you to explain why all of the past concepts were wrong, why yours are

[24] Fodor, p 31-32. (Material from George Washington example excised for clarity.)

[25] Evidence from this would be Fodor's American President example from Fodor (op cit. 4), p 30-32, where one person's ideas about the American President are listed as being different from another. Thus, in this case, both people still have the concept of 'the American President', even if it does have slightly different constituents.

right, and how you will never be proven wrong. Similarity relations allow for changes in the meanings of concepts over time; between subsections of a community, between individuals, or even over time in the same person.

A corollary to this is the prevalence and power of analogy and analogical learning[26]. We can learn by analogy, and then end up knowing more or less than the person who was teaching us. If identity relations are fundamental, then analogy should only ever be able to provide the same amount of information to a learner as is possessed by the teacher. That is, it should not be possible for a student to eventually know more than his teacher, as the teacher must have mastered the concept in order to be able to pass it on.

How to get out of the tangle

My response to Fodor would be that if you consider concepts as non-atomic, then the similarity problems are alleviated. That is, if you think of concepts as generalisations from sets of instances, judged by contingent decisions as being members of the concept, then both of Fodor's objections evaporate. Also, for both groups and individuals a concept may change over time. If this seems unclear, remain calm, as I will elaborate both here and in later sections of this work.

To answer the first objection: It is true that two people have to understand that they are communicating about the same concept. However, in a thoroughgoing similarity theory, the only way for two people to establish what they are talking about is to agree that they are talking about the same thing. For things that vary little from person to person (e.g everyday physical objects), this is generally not a problem. When it is, people misunderstand each other, and must negotiate back to common ground in order to be

[26] A good summary book is about the importance of analogical reasoning in concept learning is Vosnaidou, S and Ortony, A (Eds.) *Similarity and Analogical Reasoning*, Cambridge University Press, Cambridge, 1989. See also the classics: Hesse, M. B. *Models and analogies in science,* London, Sheed & Ward, 1963, and the more recent Leatherdale, W. H. *The role of analogy, model, and metaphor in science*, Amsterdam : New York : North-Holland Pub. Co. , American Elsevier Pub. Co., 1974.

understood. For things that are highly abstract, for example, scientific concepts and ideas, misunderstandings are more common, as there is less commonality of experience. An example should help to clarify this.

A table would be a good physical example of an everyday object. Everyone who has a good idea of what the word ‘table’ means has probably seen a reasonably large number of tables, and can thus extract from their experience the idea of what constitutes a table. A person who has never seen a table would have difficulty in understanding exactly what was meant by the word until they had seen it. However, those who have seen tables have seen a set of physical objects with reasonably stable properties (some of which would be legs, a flat surface on which you can place things, and so on.), even though the details may vary between the instances. Communication about physical objects is generally unproblematic, as there is a large commonality of experience about them.

As the things being talked about become more abstract, the amount of commonality of experience decreases. Another good example is gravity. Most people who have been educated today have some idea of what gravity is. It’s the force that makes objects fall to the ground. As has been pointed out by many people, this is a very theory-laden statement. It presupposes a lot of knowledge about the nature of a ‘force’, (among many other things) that a person not educated in this western, industrialised culture would not have. For members of this culture, there is a large commonality of experience, in that most have been told similar stories about gravity during their education. Moreover, those who have studied physics further (for example, at university) will have more information available to them about gravity, about its foibles and details. They will thus have less in common with others, and thus use the word ‘gravity’ to signify something different. However (and this is a big ‘however) the only way for a person who does not share their background to figure out that they have a different concept of gravity is by the way the trained physicist uses the word in communicating. As long the physicist uses the word in a manner that is consistent with the untrained person’s ideas about gravity, there will be no confusion, *even though the people may be* *talking about different things*.

This also applies to the second objection. It is perfectly valid to say that my concept CAT overlaps your concept CAT if the only way each of us *can* know what the concept CAT consists in is via our own experience, and the agreement that we are both talking about CATs. The overlap is only evident when we begin generalising common properties of our respective CAT concepts. That is, as long as we don’t *know* there is a difference between our CATs, then, effectively, there is none, and all of the concept identity material applies. I should point out here that, if differences in the instances of concept CAT in one person cause the person to use CAT differently to another, then both people will know that they have different CATs and the negotiation process can begin. In the gravity example, if the physicist speaks of gravity in terms of space-time curvature, the other person will realise that the physicist has a different concept, and the person is then free to request clarification. Of course, the physicist may not be willing, or some other factor may interfere, but the parties will still be aware that they may be talking about slightly different things.

Basically, it comes down to this: As long as the parties involved in any communication *believe* that they are talking about the same thing, *then functionally they are*, even if the underlying data (that is, the actual experiences of the communicators) is different. In practice, of course, any major differences in the underlying data will create differences in usage which can be noticed by the communicators, and thus negotiation can begin. This leaves us with the question of ‘*How* are the concepts different between people, and between the same person at different times, and between contexts?’ To which Lawrence Barsalou has an answer.

Barsalou and context dependency of concept structure

Barsalou[27] 'demonstrated that certain knowledge about a category – *context-dependent information* – becomes active only if relevant in the given context.'[28] Barsalou's contention is that the context[29] in which a concept is used changes the structure of the concept. 'Different people do not use the same representation for a particular category, and a given person does not represent a category in the same way across contexts.'[30] Obviously, for anyone who is interested in concept *dynamics*, this is very interesting, implying as it does concepts that change over time.

To account for these effects, Barsalou posits a retrieval-based framework for dynamic knowledge representation. That is, he proposes a theory that:

> a concept is simply a particular individual's *conception* of a category on a particular occasion. And rather than being definitional – as they are often assumed to be – concepts simply provide an individual with useful expectations about a category based on long-term past experience, recent experience, and current context.[31]

This theory is similar in its outlines to my own, which is partly why I mention it here. However, I will go over its structure in a more thorough fashion in

27 Barsalou, LW, 'Intraconcept similarity and its implications for interconcept similarity', Ch 3 in ibid, p 76-121

28 ibid, p 77

29 It's important to note here that Barsalou's setting of the context involves using sentences that indicate one or another of a word's attributes is to be activated; it is thus simpler than what might be found in the real world.

30 ibid, p 86

31 ibid, p 93-94

Chapter 3 – The Importance of Implicit Knowledge, as I describe my own work.

Having circled through some cognitive science issues and ideas, it's now time to have a look at some SSK. In the next section, I will look at Barnes and Ravetz, the similarities and differences between them, and how this fits into my argument.

Chapter 2 – Barnes and Ravetz.

This part of my thesis began as an essay on the textual comparison of two books, *TS Kuhn and Social Science*, by Barry Barnes[32], and *Scientific Knowledge and its Social Problems*, by Jerome Ravetz[33]. As I worked on the essay, ideas stimulated by the symmetries I percieved between the texts drew the paper into my thesis.

Barnes is best known for his sociologising of the works of Thomas Kuhn, and *TS Kuhn and Social Science* is his book on the subject. Ravetz is not as well known, but his project was to explain the special character of scientific knowledge, and provide some explication of how it could lead to problems (e.g. thermonuclear war). Ravetz seems to have come from a more 'hard sciences' background, and as such is very good at looking at the smaller details when studying the sciences.

In any case, my focus here, is that some aspects of the finitism of Barnes, namely his learned similarity relations between finite clusters of events, and Ravetz's theme of implicit craft knowledge, most especially his sets of intellectually constructed classes of things and events, are very similar and support each other. Thus, a synthesis of these two sets of ideas may prove interesting, and important for the philosophy of science, and sociology of science. The results of the analysis will hopefully be of assistance to all of the areas I weave through on this quest for a new sketch theory of concepts.

The most important idea to come out of SSK in general, and from Barnes and Ravetz in particular, is the acknowledgement that human beings are social creatures, and that social forces are as important as any other, and have to be considered when studying any area of human experience that involves

[32] Barnes, B. *T S Kuhn and Social Science*, Macmillan, London, 1982

[33] Ravetz, J.R. *Scientific Knowledge and its Social Problems*, Second Edition, Transaction, New Brunswick, 1996 (originally published 1970)

more than one person. This is especially true in the case of modern, industrialised science[34], where there is not only a very large specialist culture of scientists, but also a very large amount of interaction between the scientific community and the rest of the community.

Serious recognition of the fact of social interaction in science, and the operation of sciences, began (as with many other things) with Kuhn. Barnes points this out, and then takes the idea even further, incorporating philosophical ideas about finite conceptual sets (finitism), and using them to attempt to create some sort of coherent account of how scientific knowledge can be valuable in the absence of a neo-Platonic universe where the rules are waiting to be discovered by anyone who looks for them in the right way.

Ravetz is also concerned with studying what makes science special; why it seems to work so well. In this, both books are following a classic question in studying science. However, Ravetz is particularly interesting in his documentation of some of the phenomena at the lower, more ‘coal-face’ levels of the disciplines, at the levels of tool-use, of claim-making, and so on. Above this, what makes both of these books interesting is the realisation that social factors are important, even crucial, and must be explained in some fashion.

While reading and comparing the two, and having noticed this point about allowing for social factors, I realised something about each text. Barnes' project seemed to be focused on providing some sort of more philosophical basis for post-Kuhnian ideas of social construction of scientific knowledge, while Ravetz seemed to be focused on how the social construction of scientific knowledge actually works, with particular reference to ideas such as implicit knowledge, the 'craft-like' nature of much scientific knowledge, and also what makes scientific knowledge special. Ravetz’s ideas on intellectually

[34] I owe a lot to Ravetz for developing this idea of modern, highly industrialised science, especially with reference to post-war science.

constructed classes of things and events were also important in generating my ideas.

As I studied the two, I found that many aspects of one book would complement what the other book was *not* looking at, and thus a picture began to emerge. It is a picture that takes some of Barnes' ideas about finite clusters of instances and concepts, and also takes some of Ravetz's ideas about the importance of implicit knowledge and about the intellectual construction of objects of inquiry, and attempts to put them together to create a coherent picture of how social factors can be important in scientific knowledge, while still retaining the special character of scientific knowledge.

In this section, I will first briefly describe what Barnes and Ravetz have to say, and then point out the similarities between them, and the places where they support each other. In doing this, I will show the reader these ideas, (for those who have not encountered them before), explicate them somewhat (for those who have), and finally show how they can be put together to shed some new light on theories of concepts, and on how we think about expert disciplines.

Barry Barnes, learning by ostension, and finitism

By the title of Barry Barnes book, 'T S Kuhn and Social Science', it seems obvious that his project is to 'sociologise' Kuhn to at least some extent.[35] Whether or not he is successful in this will not be addressed here.

What I am interested in is some of the points that come up in Barnes' discussion of training, namely those that describe Barnes attempts at fitting the framework of finitism with extra sociological appendices.

Barnes follows Kuhn in proclaiming the importance of paradigmatic examples:

---

[35] In addition, in his theories of learned similarity relations, he is sociologising Hesse's conceptual webs. These were outlined in Hesse, M, *The structure of scientific inference* ,London, Macmillan, 1974.

> Abstract verbal presentation of concepts, definitions, rules and laws are pedagogically unsatisfactory, and take second place in science to teaching through paradigmatic examples.[36]

That is, merely presenting the laws will not work as a teaching method, but they instead require paradigmatic examples to illustrate those laws.[37] In the case of science, these paradigmatic examples are taught through the solution of textbook problems, or other problems given as practice (e.g. tutorial problems etc.):

> The exercises of the science textbook are akin to the finger exercises of the pianist, pedagogically preferable to the music of actual research.[38]

The obvious question to ask is: How is this done?

Teaching by paradigmatic examples: How?

Barnes uses Kuhn's example[39] of a boy learning what a swan is while walking through a park with his father. The child points to birds as they walk through the park, and names them according to previously learned types of birds. So, when the child sees a swan, if it does not know what a swan is, it will probably simply call it a 'bird', and the father, 'who can be taken as a source of the accepted usage of his community'[40], will correct him, informing the child that that particular bird is a swan. Then, when the child sees another type of bird, it will either call it a swan or a bird, and will be corrected by the father. Thus, if the child sees a duck, a goose, and a swan, it will learn how to identify these birds.[41] This is learning by ostension, or by example.[42]

---

[36] op. cit. 1, p 18

[37] The idea of examples *instantiating* laws is important in my ideas, so please take note (this will be explained further later in this work). A quick summary would be: since concepts can be thought of as sets of instances, a single instance of a teacher telling a student the mathematical representation of a law is obviously less desirable than that same instance, with a number of associated instances of the law in action. More on the details of this later.

[38] ibid. p 18

[39] From Kuhn, T.S. *The Essential Tension*, University of Chicago Press, 1977, p 307-319. Quoted in Barnes (op cit. 32), p 23

[40] ibid. p 23

[41] Example from ibid. p 23

Barnes contrasts this with learning by rules and definitions, that is, learning by the application of a linguistically specifiable rule, such as a scientific law, or in Barnes' example, an assertion such as 'A swan is a large, white, orange-billed bird.' As Barnes points out, this implies that the rest of language can be learnt in a similar fashion. Barnes notes that Kuhn responded to this by observing that this kind of learning does not always apply. Processes of ostension, says Barnes, are 'a form of knowledge acquisition with practical advantages over learning by rule and definition and thus they cannot, without loss, be replaced by processes solely involving rules and definitions.'[43] What those practical advantages are may become clearer as we look further at Ravetz's ideas on crafts and craft practices.[44]

However, Barnes goes on to say 'learning by ostension and learning by rules and definitions are not competing strategies between which a choice is always possible.'[45] Later, he continues, 'we can see now how false is the contrast between learning by ostension and learning by use of definition and rule. The real contrast here is between learning that relies *directly* upon ostension, and that which relies upon it *indirectly*. Nothing can be learned *ab initio* purely by verbal means.'[46] Barnes also suggests the idea that 'what any given term in such a[n] [ostensive] system refers to can never be characterised without reference to learned similarity relations.'[47] Thus, the importance of learning by

---

[42] ibid. p 23

[43] ibid. p 26. Barnes does not explicitly state here what the advantages are, but later he speaks of terms being 'learned directly', without 'involvement of existing language'. (both ibid. p 26) It would seem he means that it requires little to no language to be present previously, and thus can get around infinite regress issues.

[44] In addition, some musing about the practicalities of sets of instances should throw some light on this.

[45] ibid. p 26

[46] ibid. p 27. To link this with the idea mentioned in footnote 37 (about examples instantiating laws), one could suspect that learning by ostension must rely upon the creation of both excitatory and inhibitory relationships between instances, not only those classified as being part of the concept, but also those outside, to some extent. Thus, the ostensive learning events involve adding the experience to a web of experiences, both similar and dissimilar. Another relevant quote from Barnes: 'Any putative new instance will at once differ from, yet be similar to, the instances in both clusters.' (ibid. p 28.)

[47] ibid. p 28

ostension and the forming of implicit links between instances of a concept are emphasised. After this, the project of understanding how teaching by paradigmatic examples can work becomes understanding how people can learn similarity relations between finite clusters of events. This is Barnes' term for encapsulating his system of finitism with sociological appendices. But what exactly is this 'finitism'?

Finitism

Barnes defines finitism as follows:

> Finitism denies that inherent properties or meanings attach to concepts and determine their future correct applications; and consequently it denies that truth and falsity are inherent properties of statements. 'True' and 'false' are terms which are interesting only as they are used by a community itself, as it develops and maintains its own accepted patterns of concept application.[48]

He also puts it another way: Finitism 'denies that correct usage[49] can be discovered or inferred.'[50] In this view, concepts are open-ended, revisable clusters of instances that have been classified by ostension, are associated via similarity relations.

[48] ibid. p 31

[49] Here it seems that Barnes is using this term in a 'universal truth', Platonic sense.

[50] ibid. p 30

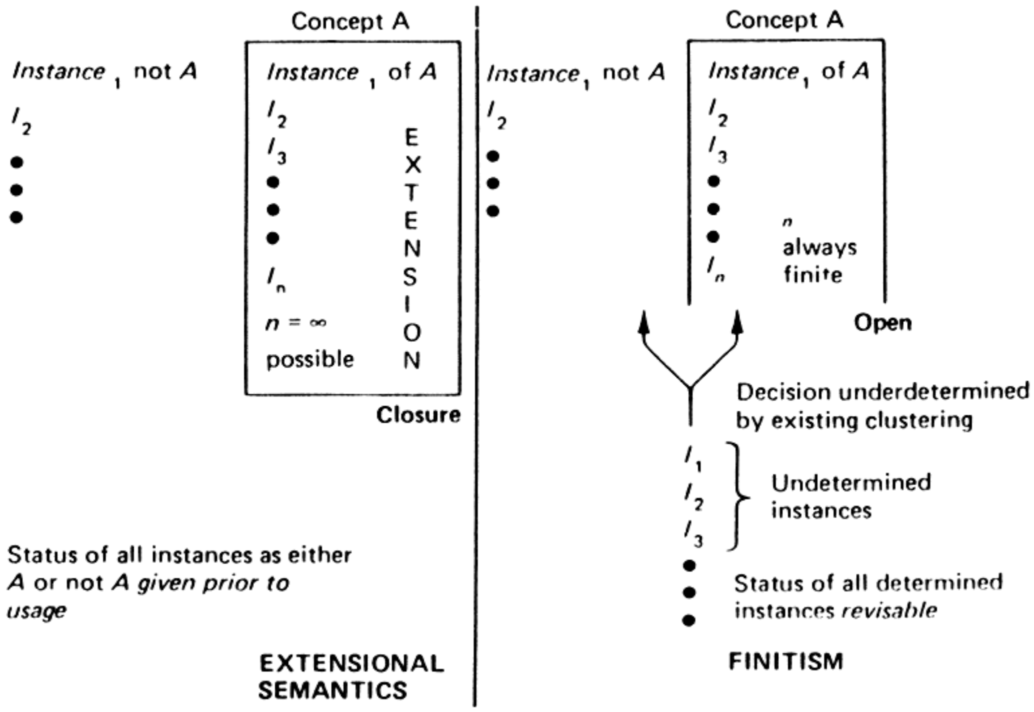

Figure 2- Barnes' Diagram comparing Finitism and Extensional Semantics – from p 31

Here is one of the diagrams used by Barnes to explicate his stance on finitism. Barnes argues that, in the case of 'extensional semantics' any judgements about what does and does not constitute an instance of the concept is predetermined, or 'given prior to usage'. In the case of finitism, however, Barnes indicates that instances will be experienced that are 'undetermined', that is, not classified as being the concept or not. Also, even once instances have been judged to be a member of a concept, their membership is *revisable.* This is important, as it relates to one of the key goals of Barnes' work, namely explaining the fact of science dynamics.

The other important thing to note about Barnes' diagram is the openness of the concept. New instances are always allowed to enter the concept, and it is through this that concepts may change over time. Of course, this is what Barnes is seeking, as he is trying to provide a philosophical basis for the dynamic nature of science as documented by Kuhn. Barnes maintains that this 'opens all instances of concept application, and all acts of acceptance or rejection of applications of concepts, to sociological study as contingent judgements.'[51] That is, whenever you are applying a concept, you are making a judgement that is determined ultimately by your social interactions, past and present. The question that remains after we understand what Barnes means by 'finitism' is: How are the judgements made? Barnes' device for understanding this is the idea of learned similarity relations between the finite clusters of events dealt with by finitism.[52]

Learned similarity relations between finite clusters of events

The reason Barnes gives for his adoption of the 'finite clusters of events' part of this phrase is: 'for all the complexity and richness of language, experience is immeasurably more complex, and richer in information.'[53] That is, it is of course only ever possible to observe a finite number of instances of anything, let alone of one particular concept.[54]

The 'learned similarity relations' part is there because Barnes is interested in how we create associations between events that are identified as being instances of a concept. As Barnes makes clear, we learn the similarity and difference between events identified as being 'within' the cluster, and we also learn similarity and difference between clusters.[55] If we look at Barnes' figure, reproduced here as Figure 3, we can see that Barnes considers the instances

---

[51] ibid. p 32. Text excised: ' '

[52] Of course, there is another question that really should be raised here, namely exactly what is meant by the word 'instance'. Personally, I am inclined to let 'instance' mean 'event that has been classified under a concept', but Barnes doesn't use it in that way. He uses it in the diagram to describe any event that is classifiable, but later (see reference in footnote 73) he refers to an instance as an application of a concept. The difference between Barnes and myself is subtle but present. The upshot of all of this is that I will, in Chapter 3, be using 'instance' in a manner slightly different to Barnes. Actually analysing exactly what Barnes means when he says 'instance' is at least a project of this length in itself, so I will neglect to do it here. However, I think that what we have so far is enough to go on.

[53] ibid. p 28

[54] Also note the implicit equivalence in Barnes' language in the previous quote between a concept and the language used to communicate it. This appears to be anomalous compared to the rest of the text, but it is something to keep in mind.

[55] from Barnes Figure 2.1, ibid. p 28, reprinted on the next page. Here you can see why I felt it necessary to mount some defence of similarity relations. Barnes is not talking about the exactly the same things I am, but however, I feel that, in thinking of concepts as non-atomic, Barnes was ahead of his time.

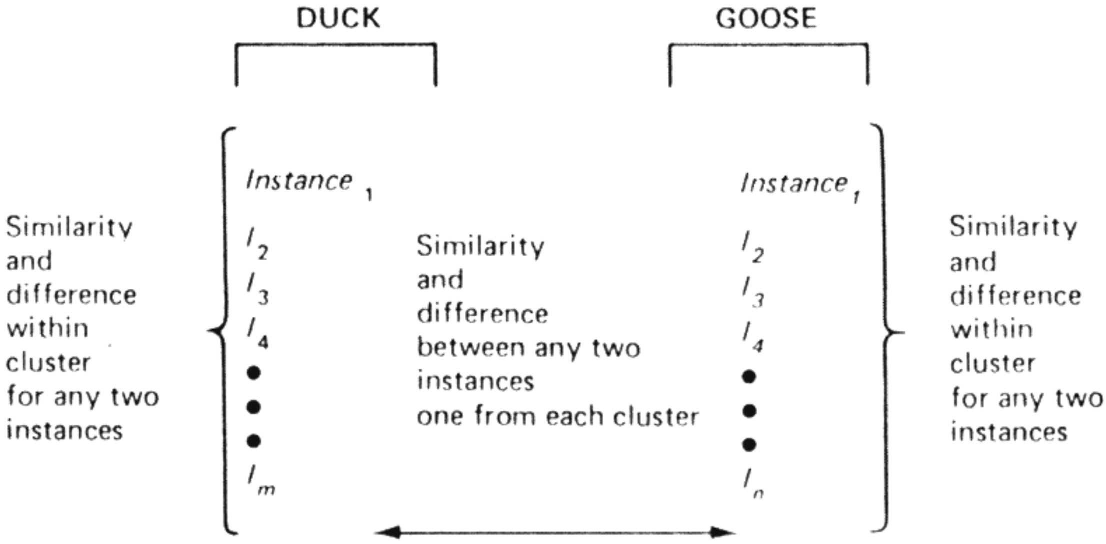

Figure 3- Barnes' similarity relations diagram (from Barnes, p 28.)

to be related with similarity relations within the concept set, that is, between the instances that have been judged as members of the set. Also, and perhaps more importantly, he considers there to be similarity relations between members of complementary sets, in this case, duck and goose. That is, there are similarity relations between both the instances that are members of the conceptual sets, and between instances that are not. Thus, there are similarity relations between *all* instances that have been judged to be the member of *any* concept. That is, instances are part of a web of experience that is capable, by its interconnected nature, of representing conceptual sets as groupings of instances that are associated in a positive fashion (as the instances in the concept 'goose' would be), and of distinguishing these sets from others via instances that are associated in a negative fashion (as the relations between instances from the two complementary concepts would be.)

Barnes also notes that 'past experience and the past usage of a concept can never suffice to determine future usage.'[56] Thus, all of the differences and similarities between the new and old instances, and among the past instances are important in the judgement of what the new event is. That is, the

[56] ibid. p 28

differences and similarities are important in the judgement as to whether the event is an instance of a concept, but are not the only factor. Barnes: 'Formally... [the] ...assertion that an instance falls under a term is only his contingent judgement to the effect that similarity outweighs difference.'[57] In other words, the judgement is *contingent*, dependent on already existing factors. Implied is that the judgement is contingent on previous judgements, and other determinants of the mental state of the judger. Here Barnes also notes his belief that 'there is no natural or universal scale for the weighing of similarity against difference.'[58] That is, Barnes has created a rationale for the incommensurability of Kuhn.

Learned similarity relations and Communities.

It would seem that the natural question to ask, if the previous is the case, concerns the father in the example. Who performs the analogous role if there is no universal authority figure for larger groups to refer to in order to set concepts? Given the title of the book, and the importance of community in Kuhn's work, it is hardly surprising that Barnes ascribes this role to the relevant community.

> Proper usage is simply that usage communally judged to be proper, and is no more predetermined than idiosyncratic individual usage. Concepts cannot themselves convey to us how they are properly to be used. We ourselves must always agree or seek to agree that the application of a term to an instance is justified, that similarity should outweigh difference in that case.[59]

Barnes goes on to explain how a combination of our biological make-up and social upbringing makes the application of many concepts so routine as to appear innate.[60] He also addresses the issue of problematic instances, such

[57] ibid. p 28

[58] ibid. p 29

[59] ibid. p 29

[60] This is an interesting point, as it implies that it is possible that *nothing* is innate! I will consider this more later.

as: Is deuterium oxide water? In cases like this, says Barnes, the community must negotiate its way to a decision as to whether the instance is to be included or not. Barnes: 'And although the new routine may assist a community in its dealings with nature, nature itself sets no constraints on the form of the routine which is produced.'[61] And later: 'Meaning changes, or stays the same, as the community wishes to have it.'[62]

Barnes on Research

Following from all of the work on training is a picture of research itself. Given when Barnes was writing this, and the fact that he was explicitly 'sociologising' Kuhn, it makes sense that Barnes then begins to explain paradigms and how these are maintained within a finitist framework.

To begin with, Barnes is talking about normal science. 'In agreeing on a paradigm scientists do not accept a finished product: rather, they agree to accept a basis for future work, and to treat as illusory or eliminable all its apparent inadequacies and defects.'[63] That is, scientists are well aware that there are problems with any theory, and these can be thought of as either opportunities for new work, or not worth considering.

Specifically, when a theory is first proposed, 'there is felt a need to improve techniques of measurement, and to extend them to more cases.' This is one type of problem that can be looked at in the normal procession of science.[64]

---

[61] ibid. p 29 This quote is important, as it implies how some communities can limit themselves, and some can succeed more than others. Basically, unless the community rates high correspondence with *all* of nature as important, the community can be very isolated from the natural world, and end up in all sorts of odd (to us) places. For example, one might suggest that animistic religions in general weren't out to understand *everything*, in the same way science as a culture is, so they (as a culture) were satisfied with much less sophisticated explanations.

[62] ibid. p 30. This is a nice encapsulation of the idea Barnes is trying to convey.

[63] ibid. p 46. Ravetz also has a lot to say about the treating of defects as eliminable, especially in relation to his ideas about 'pitfalls'

[64] I find this idea of 'normal' science a bit too much like Kuhn's, that is, a bit too focussed on puzzle-solving, but Barnes' point seems to be that 'normal' science, far from being the stagnant puzzle-solving that people sometimes think of it as, is a growing, changing thing, somewhat similar to revolutionary science. The difference lies in the

Another type of problem is the extension of the paradigm to new areas, via 'the reconstruction of existing problem-solutions so that they fit new situations.'[65] In both types of problems, note that each is concerned with fitting new data into the existing structure, with making an analogy to existing solutions. In addition to these two extremes, there are many kinds of problems that combine elements of the two in varying degrees, but what all of the problems have in common is that they rely on analogy to extend the theory.

So, at this stage we have: a paradigm is accepted, with all of its flaws, and these flaws are either ignored or used as a basis for further work. This business of extending the theory by analogy, however, is very, very important in Barnes work. A later quote:

> Scientific inference, like empirical inference generally, is not deductive. It proceeds from particular to particular on the basis of resemblance and analogy. Knowledge is built up and extended a bit at a time by the revisable clustering of instances and applications.[66]

And again: '*Any 'deduction' about empirical phenomena involves a hidden analogical step.*'[67] That is, anyone who claims to be proceeding 'deductively' is, in reality, making an analogy to an already existing problem-solution in order to give themselves a frame of reference. Barnes makes the point a bit more clearly:

> Training displays similarity relations by linking instances to instances; research extends similarity relations by linking instances to instances. The links are made in both cases by analogy.[68]

And again, focussing on research:

---

degree of the change. This is a thoroughly post-Kuhnian SSK idea.

[65] ibid. p 47

[66] ibid. p 122

[67] ibid. p 122, Barnes' italics.

[68] ibid. p 52

> By seeing the unknown in terms of a known problem, inductive inference is possible: variables in the unknown situation are calculated by assuming that it behaves analogously to a known one. Thus science proceeds by analogy and induction, with the former licensing the latter.[69]

Thus Barnes feels that analogy is vital to scientific learning, and indeed learning in general. Analogy would seem from all of this to be the way in which individuals can extend similarity relations without the application of some authority. That is, without some authority to inform as to the way the community says things are, the only basis for a person to learn new things is by analogy to judgements they have already been informed about. In this fashion, the boy in the swan example could classify other things as swans only on the basis of analogy with the instances previously observed and categorised by authority, but which are of course contingent and revisable at both the group and individual levels. Thus analogy becomes the most important component of any learning, and training becomes the learning of accepted instances to form a basis for the analogy that is used in research. It becomes the basis of the mechanism for unsupervised learning, and thus for research.

There is one remaining point that is important to note in considering Barnes, in relation to the purpose of this essay, and that is that there are no objective criteria to evaluate claims:

> But a piece of normal science does not generate a candidate problem solution which is then independently validated.[70]

This quote is specifically regarding the lack of an independent validation criterion for claims arising from the research process, and thus is related to what Barnes has to say about analogy. This also has echoes in Ravetz's ideas about validation criteria, but this will be discussed later.

---

[69] ibid. p 53

[70] ibid. p 50

Another quote:

> There is no logic to help determine the relative technical advantages of the alternative strategies of concept application: people simply have to agree which generalisations they will take account of, and agree in their practice how they will be taken account of.[71]

The most important parts of this quote are 'There is no logic' and the notion that people have to agree on generalisations. I have included the two quotes to illustrate the following point: There is no way to know before you begin working whether or not your question is answerable. This also applies to communities, in that there is no way to know without argument and discussion what the 'right' way is. The notion that 'There is no logic' is not as remarkable now as it was when Barnes book was written. Really, this is just a statement that there is no formal theory of how theories are arrived at.[72] The notion of agreement on generalisations is a reflection of the conventional character of knowledge in Barnes conception of science.

Obviously the other important point here is that people have to agree on what things will be taken account of and how this will happen. To return to the point I made in rebutting Fodor's arguments against similarity relations, as long as people do agree, then there is no problem, and the system proceeds *as though* the concepts were atomic and related with identity relations.

To summarise with a quote from Barnes:

> In training, the scientist learns the accepted similarity relations by exposure to successive particular instances, or applications, of terms. He is judged competent when his subsequent development of similarity relations runs along the lines

[71] ibid. p 109

[72] Indeed, this would create a paradox, because then you would be able to know what you would know. That is, if a logic of scientific theories was possible, then, using this logic, it would be possible to derive all the knowledge that it is possible to know, and thus know what you would know. See Popper, K. R. *The Open Universe: An argument for Indeterminism*, Hutchinson, London, 1982, p 62 for this idea as an argument against determinism.

> accepted by his community; this is indicated by the way he solves routine problems not completely familiar to him but familiar to his teachers. In research, the scientist solves problems by modelling them on existing solved problems or paradigms; and he thus again develops similarity relations to cover further cases. The difference is that this time the cases are as unfamiliar to the community as to him, and he must hope that the community comes to accept his work as sound.[73]

Thus Barnes establishes a means by which a thoroughgoing similarity theory can allow for the creation of traditions of practice, of expert disciplines, via a finitist conception of concepts as being sets created through the classification of instances, instances that are connected via similarity relations into a web of experience. The part of this quote that it is important to take notice of is the remarks about the judging of competency via the solution of routine problems, and of research being the same thing, only more uncertain, as these sorts of ideas are vital to Ravetz's work.

Jerome R Ravetz and implicit craft knowledge

If ever there was a book whose title clearly states the unorthodoxy (for its day) of the ideas contained therein, it must be 'Scientific knowledge and its social problems.', by J. R. Ravetz.[74] "What's that you say? Science causes *social problems*? But how can that be, when science is the right way to do things?" Of course, this is not as contentious now as it was when first published, although I am sure that there are a lot of people who probably should have read this book that have not because of its title.

A large part of Ravetz's work, and the part which I will be focusing on in this essay, concerns the craft-like properties of science. Ravetz points out many things about how science is actually done that seem very similar to ideas of

---

[73] ibid. p 52. It should be noted here, following from footnote 35, that in sociologising Hesse, Barnes puts himself in her tradition, and thus is following Duhem and Quine in many respects.

[74] op. cit. 2

'craft-type' knowledge. I will proceed to explicate this somewhat.

To begin my explanation, Ravetz summarises his position like so:

> 1. Scientific Knowledge is a craft.
>
> 2. The objects of this work are not natural things, but are intellectual constructs, studied through the investigation of problems.
>
> 3. The work is guided and controlled by methods which are mainly informal and tacit, rather than public and explicit.
>
> 4. The special character of scientific knowledge is explained by the complex social processes of selection and transformation of the results of research.[75]

Although all four items are of interest, the fourth is mainly of interest with reference to Ravetz's motives in writing the book, that is, explaining the special character of scientific knowledge. The second is perhaps the most important for this work, being concerned with the objects of science, that is, the tools, methods, and concepts used to perform it. An important item to note with respect to the second item is that the objects of inquiry in science are intellectually constructed. Ravetz in fact uses the term 'intellectually constructed classes of things and events' to refer to these things, that function as the objects of inquiry for research and are typically modified as the outcome of research[76]. This term is also important in Ravetz's conception of tools as being objects of inquiry that have been standardised and solidified, but are still open to renegotiation through debate. However, I will explain this further below. Before this, however, we need to review Ravetz's conception of science as craft-work.

[75] ibid. p 71-72

[76] It is also important here to grasp that in Ravetz's account, other objects of inquiry must be appended to an argument advancing new or altered claims about the objects under study.

Science as Craftsman's work.

Ravetz uses this phrase as the title of one of his chapters, and it is illustrative of a large part of his argument. However, if we are going to consider the sciences as craftsman's work, we need to understand what Ravetz holds a craftsman to be. A quote from Ravetz:

> The craftsman works with particular objects; he must know their properties in all their particularity, and his knowledge of them cannot be specified in a formal account. Indeed, no explicit description of a craftsman's techniques, and of the objects on which he works, can be more than the simplest elements of the subject. They can be useful for the beginner, but he must develop a personal, tacit knowledge of his objects and what he can do with them, if he is to produce good work.[77]

This outlines Ravetz's conception of a craft. The most important things to note here are the importance of implicit, unspecifiable knowledge, and the relative unimportance of explicit, systematised, 'scientific' knowledge.[78]

Ravetz feels that the experience gained in solving 'textbook' problems is essential to learn the implicit rules of use for all the apparatus used, whether the apparatus is physical, as in laboratory instruments, or abstract, as in mathematics.[79] This includes information about how and when not to use the apparatus as much as information about actually using it. In Barnes' language, this would be because of the learning between all of the instances, including both those who were judged correct and incorrect by the teacher. This idea is also communicated in Ravetz's idea of pitfalls.

---

[77] ibid. p 75

[78] I use this term in quotes to be deliberately provocative. However, I think it would be difficult to argue that the idea of 'scientific' knowledge does not imply systematisation and an explicit nature. Also important here is the emphasis on 'particular' objects. This is important for Ravetz's ideas of intellectually constructed things and events: see later section on this.

[79] see the discussion on the transformation of data into results (ibid. p 76-92) for examples of this. A primary example of the abstract kind of apparatus is statistics, which Ravetz uses repeatedly throughout the book.

Pitfalls are 'concealed traps for the unwary'.[80] These are errors that are easily identified as such in retrospect, but are difficult or impossible to detect when first using an apparatus. Thus, the education of a practitioner in the pitfalls of the apparatus is vital if the practitioner is not going to fall into 'simple' errors and end up looking amateurish.[81] However, as Ravetz points out, it is not possible to completely eliminate pitfalls from scientific inquiry, but it is possible to avoid them. This can be done in two ways: 'by the charting of standard paths which skirt them, and by each investigator becoming sensitive to the clues which indicate the presence of the special sorts of pitfalls he is likely to encounter in his own work.'[82] The first is communicated to students via explicit description, textbook problems and the like, but the second is wholly dependent on the investigator being proficient enough with their tools to recognise 'odd' data. As Ravetz says, 'when an individual scientist explores beyond the range of the well-established techniques, his craft knowledge must necessarily be more subtle and personal, for the pitfalls he is likely to encounter are peculiar to his particular materials and tools.'[83]

With respect to actual research, as opposed to training, Ravetz feels that the main differences are that 'a major part of the work is the formulation of the question itself; there is no simple rule for distinguishing a 'correct answer from incorrect' ones; and there is no guarantee that the question, as originally set, can be answered at all.'[84] Anyone who has ever prepared a paper (or for that matter an undergraduate report!) will confirm this is the case.

Classes of intellectually constructed things and events

The idea of classes of intellectually constructed things and events is used by

---

[80] ibid. p 95

[81] 'At every stage of our exploration of the unknown, we are at risk of being mistaken, and of remaining in ignorance of our mistakes until irretrievable damage has been done.' ibid. p 95

[82] ibid. p 97

[83] ibid. p 97

[84] ibid. p 99

Ravetz to provide a part of his rationale for the 'special' character of science, while maintaining the importance of social processes. More than this, they serve as a foundation to base the rest of his argument on. Ravetz's argument is that science is the only field of human endeavour that operates wholly on objects that are not real in the same way as they are in handicrafts. They are negotiated and constructed via social processes in the scientific community. Thus, they are ‘intellectually constructed’.

Also, and more tellingly, the objects themselves are special in some way distinct from those in, for example, theology, where the objects of inquiry are also intellectually constructed. According to Ravetz, the reason for this is the special relationship of the scientist to his tools, which is similar to the craftsman’s intimate knowledge of *his* tools. However, the scientist’s tools are varied, and are special because they are the result of the research process, that is, they are classes of intellectually constructed things and events themselves. To state this again: The objects of inquiry in science are intellectually constructed classes of things and events previously ‘researched’ by someone, and then negotiated by the community until they are accepted as standard, or ‘encapsulated’.

Ravetz speaks of the ‘lengthy process of work’ required to do this, ‘with the individual phase of the investigation of a problem followed by the social phase of the testing, through use, of its solution.’[85] After the social testing phase, the work can attain ‘tool’ status, and be taken as given, to be taught to students as self-evidently true. However, says Ravetz, ‘a solved scientific problem is not, and cannot be, a closed and perfect structure.’[86] It is possible to re-open discussion and debate about any scientific object, questioning the assumptions that were made in generating it, or questioning the data used, the tools used, and so on.

---

[85] ibid. p 191

[86] ibid. p 193

The important point here is that intellectually constructed classes of things and events are not only the objects being researched, but also the outcome of research. It seems that Ravetz's argument is that this feedback is what allows science to be special. That is, according to Ravetz, it is the objects of inquiry being encapsulated into tools for further research that makes science special.

Ravetz then explains further about the ways in which the objects of inquiry can be encapsulated into tools:

> In these fields, the intricacy of the relations into which people get must be grappled with; and procedures are developed so that complex practical problems can be solved in consistent fashion….
>
> …These intellectually constructed classes derive ultimately from common sense experience; but in their detailed elaboration, the need for systematic coherence weighs at least as heavily as the retention of the original link. In any sane system of this sort, there is a place for common sense ideas to be injected directly into the decision-making process; but this is rightly kept as a last resort, or else the whole system would collapse…[87]

That is, the system can tend to maintain itself in preference to complete accuracy. It seems to me that here we see definite links to Kuhn's paradigms and revolutions; in Kuhnian normal science, the system is self-maintaining to some degree, but in Kuhnian revolution, it is recognised that the classes have drifted from their empirical bases, and need reevaluation. This is possible because the objects of inquiry have been intellectually constructed, and thus are open to renegotiation.

Ravetz uses the definition of 'substance' as an example of an intellectually constructed class that is reasonably concrete. The concept represented by the word 'substance', as Ravetz puts it, 'is not a formalised description of a

[87] ibid. p 113-114

unique collection of material; rather, it is a class of things, the members of the class being defined by their possession of certain properties.'[88] Thus, to determine whether a new thing or event is an instance of a class, i.e. of a particular substance, we must decide whether the object is a 'sample'. To do this, we simply test for the defining properties of the particular class. An obvious problem here is how to test for the defining properties of a class if they are not explicit. Obviously, then, we must generalise from some sort of prototype or exemplar, if no list of properties is readily available. But prototypes and exemplars raise other issues amply enumerated by Fodor[89]. Fodor feels that prototype theories do not allow for compositionality, as the prototypes themselves do not decompose from more complex concepts to more primitive ones. This especially holds for Boolean concepts such as ‘not a cat’ where it becomes difficult to argue what the prototype or exemplar would be. However, by considering concepts as sets of instances, this problem can be removed.

'Intellectually constructed classes of things and events' thus refers to objects of inquiry that have been constructed via a process of negotiation conducted through social channels such as published papers, concepts that are constructed and negotiated via social processes, and are open and revisable, if the community wishes them to be so. However, the community can also decide to let the objects stand as given, and hold them to be closed. This is of course very similar to Barnes’ learned similarity relations between finite clusters of instances. Now, to continue finding the parallels between the two works, we need to look at Ravetz’s views on methods.

Methods: more implicit than explicit

In discussing method, Ravetz of course adheres to the view that, just as there

[88] ibid. p 111. Of course, in the text this example is more concerned with the classification of a particular substance, but I think that the argument generalises quite well.

[89] Op. cit. 2, ch 5. “Prototypes and Compositionality” Fodor outlines in this chapter a straw-man system of this type and then goes through the issues with such a system.

is no 'Science', there is no one 'Method' that scientists follow. Rather, there are many methods, with a very large portion of these methods being transmitted in an implicit, craft-like fashion. That is, most of the methods are transmitted via social processes, most especially the teacher/student relationship.

Ravetz does feel that some idea of ‘method’ is necessary:

> There is no doubt that without an appropriate ‘method’, in some sense of the term, scientific work is impossible. A trained scientist can instantly identify the traces of the bungling amateur, or the crank, by the absence of ‘method’ revealed in a report of his work.[90]

Ravetz also points out that some sciences are obsessed with ‘method’, while in other fields, ‘method’ is not an issue at all. Ravetz goes even further:

> The root of the difficulty in any discussion of method is that it involves an attempt to render explicit that which is largely tacit.[91]

This is very important to Ravetz’s argument; as he continues:

> Hence any explicit analysis of scientific inquiry must be incomplete, at best a schematic anatomy representing a complex physiology.[92]

That is, *most methods are implicit, tacit, and not directly specifiable.* Anything that is specifiable is only at best a bare-bones treatment of the subject, if not downright misleading. However, says Ravetz, all is not lost.

> But scientific inquiry differs from handicraft work in using a body of methods which are sophisticated and subtle, but which

[90] op.cit. 33, p 147

[91] ibid, p 147

[92] ibid, p 148

> are a social possession, not restricted to the private craft wisdom of a master.[93]

Therefore, the scientific community serves as a 'group memory', through the acceptance of methods as 'standard' and the promulgation of these methods via standardised teaching. The objects of inquiry have associated methods of measurement and manipulation that are negotiated as the objects themselves are. In this way, methods can be negotiated in the same way as any other intellectually constructed objects of inquiry. Using this effect, a large portion of the mastery of any one individual can be passed to the preceding generations via the creation of standard teaching curricula. But via what processes are objects of inquiry negotiated? To understand this, Ravetz introduces the ideas of the criteria of Adequacy and Value.[94]

These criteria are set only through social processes, such as negotiation and the application of authority and thus are intellectually constructed and revisable themselves. The criteria of adequacy are used for the judgement of both arguments and evidence, and are an indicator of the maturity of a field.[95] As Ravetz writes, regarding certainty in fields:

> For we have the historical knowledge that some fields of science do achieve objectivity and near-certainty in their results, while others do not... ...The difference between them does not lie in this logical aspect of their arguments and conclusions, but in the particular circumstances of their development.[96]

That is, the particular circumstances of their development that resulted in the creation of a certain set of criteria of adequacy are part of the reason that some sciences, or indeed some expert disciplines, seem to succeed in obtaining objectivity and near-certainty. It can be seen that this idea of criteria

---

[93] ibid, p 148

[94] Ravetz's discussion of these is on ibid, p 152-169.

[95] Ravetz speaks of this on ibid. p 159

[96] ibid. p 155

of adequacy only makes any sense when it is established and maintained by the community of scientists, or indeed any community of experts distinguished by their acquisition of some kind of esoteric knowledge that requires a substantial amount of training.

The criteria of value are similar in some ways, and if anything are more socially determined than the criteria of adequacy. They are the criteria the members of a discipline use to judge what is worthwhile, and what is not. The criteria of value are obviously vital to a field, as they define, to some extent, the questions that are interesting. An excellent example of a criterion of value would be when Copernicus introduced a criterion of elegance in *De Revolutionibus*[97]. That is, he felt that another reason to accept his theory over the Ptolemaic was that his theory did not use equants, and could explain things like planetary retrogression in a simpler way than the Ptolemaic[98]. By doing this, Copernicus, in the Ravetzian view, was introducing a new criterion of value, which made it difficult to argue with his opponents, because they held different criteria of value.[99] And so the story of the 'astronomical revolution' is also a story of the acceptance of a new criterion of value.

Another important point about criteria of value is that they are even more dependent on the conditions of the field than criteria of adequacy, on such things as the amount of resources available for use, the perceived cost of the research, and the perceived reward. They are important for the 'health and future prospects'[100] of a field, says Ravetz, and because they involve the prediction of the future, they require a lot of attention from sociologists of scientific knowledge.

---

[97] This very quick summation of a mountain of historical work is based on Schuster, J.A.S., *The Scientific Revolution: An Introduction to the History and Philosophy of Science*, Schuster, Sydney, 1995, p 50-51

[98] i.e. without using epicycles for retrograde motion.

[99] Here is the incommensurability of Kuhn in another guise. Indeed, it would seem that this was a primary reason for its adoption.

[100] ibid. p 159

The criteria of adequacy and value are important in terms of this thesis because they are part of the shared knowledge of the community, the implicit knowledge that is so important for Ravetz, and that it seems so necessary to provide mechanisms for. They are also an excellent illustration of the processes of negotiation that lead to intellectually constructed objects of inquiry.

Ravetz's implicit knowledge: Other points

Ravetz's book often mentions of the idea that implicit craft knowledge and the associated intellectually constructed objects of inquiry are vital to science, and by extension with the assertion that we should thus study this implicit knowledge: its origins, maintenance, and so on. It is because this is so important, and because Barnes' project is an attempt to understand some of these things that this analysis and synthesis of the two is possible.

However, before finishing this summary of Ravetz and moving to the textual comparison of Barnes and Ravetz, I feel that I should illustrate some remaining points that Ravetz makes. These will hopefully allow better explanation of the synergies and 'meshing' of the two texts in the textual comparison.

Firstly a quote on the originators of fields:

> Hence ... there will be a natural tendency for the author to believe that the 'method' characterizing the new approach is as easily communicated to a wider audience as are the scientific results of the work. The subtle, particular, tacit component of the work will accordingly be neglected in the reflection of its principles, since this would obscure the message and decrease its prophetic effect.[101]

This quote concerns the founder(s) of a field, and how it is easy for them not

---

[101] ibid. p 170

to realise the implicit knowledge that goes into any method, and thus to produce an explicit account of their 'method' that fails to capture the nuances so important to performing actual work in the field. Of course, according to the Ravetzian account, it is the nuances that are not linguistically communicable that are the real basis for any method. Ravetz continues:

> But they [methods] have one feature in common: except for straightforward techniques, they are all largely informal or even tacit knowledge; and they are transmitted through the interpersonal channel of communication, rather than through the public channel of printed reports.[102]

In this quote, Ravetz again emphasises the importance of implicit information. However, he is also using the 'channels of communication' terminology that he coined earlier. The interpersonal channel of communication is the primary means of transfer for implicit knowledge, (as in the master-student or peer-peer relationships), while the public channel of communication is the record of published works.

Thus the interpersonal channel of communication is very important for the communication of implicit knowledge, as it is only in the close working with the subject that two people can come to understand the subject in similar ways.

Thus these two quotes serve to illustrate a point about the importance of implicit knowledge, and indeed a point that I am unsure if Ravetz meant to imply. The underlying consequence of all this talk about implicit knowledge is that: *It is possible to know something without knowing that you know it*. In other words, it is possible for your brain to store information without that information being directly accessible to your introspection. This belief underlies all of Ravetz's work about implicit knowledge, and also some of Barnes' ideas about finitism and conceptual sets. Here is an item that is flatly contradictory to classical cognitive science theorists. Implicit knowledge consists in mental representations that are not directly accessible to the

[102] ibid. p 173

conscious mind's introspection. However, the mental representations show themselves in many ways, just as the subatomic structure shows itself in many ways, even in the absence of the ability to directly observe the structure. This point about implicit knowledge is fundamental to both the comparison of the two works, and to my own ideas.

And so we come to a point where a consideration of the two works is possible; Barnes the finitist sociologist-philosopher, and Ravetz the practitioner-philosopher.

Tacit knowledge

What seemed to draw the two books together the most was the synergy between the authors; this idea of knowledge that is not explicitly transmittable, but instead only transmittable by experience and example. That is, the idea of knowledge that you do not know that you know. Thus, Barnes' finitist ideas seem to mesh well with Ravetz's conception of science as craft work, in the following ways:

Barnes speaks of learned similarity relations[103] both between concepts and between instances of a concept. Ravetz speaks of implicit knowledge[104], and of the objects of science being intellectually constructed classes of things and events. The way in which these two ideas merge may not be obvious, so I will consider an example, that of the solution of a textbook problem on the motion of a pendulum. This problem usually involves the measurement of the period of a pendulum's swing for a number of lengths of string.[105]

Barnes' point is that each of the instances of measuring the motion of the pendulum are classified as part of the set of instances representing the

[103] op. cit. 1, figure 2.1, p 28

[104] op. cit. 2, p 71-72

[105] It should be noted here that I use the word 'problem' in the Kuhnian normal-science sense, i.e. an exercise given that is in principle solvable. Here the purpose is ostensibly to train the pupils in the law under consideration, namely the law of constant acceleration in a gravitational field.

concept of pendulum motion[106]. Ravetz argues that, when solving these sorts of problems, what is also learned are the proper modes of use for the apparatus, in this case, a pendulum, ruler, and clock of some kind[107]. Barnes argues that what is learned is not simply the overarching similarity relation - that of the observed motions to the ideal of pendulum motions (which is given as an abstract law, and also as an implied relation between instances), but also a multitude of other similarity relations between all these instances, keeping track of what is both the same and different to all the others for each[108]. This provides a mechanism for Ravetz's learning of implicit knowledge of apparatus operation. Those instances in which the apparatus is used 'correctly' are labeled as such by an authority (teacher or lab instructor), as are those in which the apparatus is used 'incorrectly'. These labellings enable the learner not only to gain an idea of the concept of pendulum motion, but each instance also serves to educate regarding the correct modes of use of the equipment. In the pendulum example, a student may get unexpected results, only to be told that he was using the stopwatch they were given incorrectly. The student now not only has increased their understanding of pendulum motion, by understanding how the results they obtained differed from the established norm, but also their understanding of the norms of stopwatch-use.

I will be the first to admit that this example is somewhat inane, given its simple subject matter, but my aim is to communicate the idea that what I call intra-event learning, what psychologists call 'contextual effects'[109], is important in learning science, and indeed in learning anything.

---

106 op. cit. 1, figure 2.1, p 28

107 see previous footnote, number 34

108 op. cit. 1, figure 2.1, p 28

109 Effects of the context in learning have been 'shown' (to the satisfaction of psychologists) with respect to many effects on animals. Most modern psychology of learning assumes this type of effect. (evidence for this is from my own experience with studying the psychology of learning at UNSW. See Westbrook, R.F. 'Lectures on the psychobiology of animal learning', Course Material, PSYC2081, Learning and Developmental Psychology) See also the work of Barsalou, which I mentioned in the introduction and will be speaking in more detail about in chapter 3, Chapter 3 – The Importance of Implicit Knowledge.

What is the most important to note with respect to this material is the way in which Ravetz tends to say *that* an effect exists, and provide some descriptive evidence for *how* the effect works, Barnes concentrates on *how in principle* and *why* such effects exist. In this way, I feel that Barnes provides a kind of philosophical 'basis' for Ravetz's ideas of craft knowledge. Another area where both books differ from conventional philosophy of science and concepts is in the ascription of a high importance to analogy.

Analogical inference

Given the background of social studies of science shared by both works, it is hardly surprising that both share the idea that science relies heavily on analogy[110]. In Ravetz, the idea is less explicit[111], but I take it to be very similar to Barnes' idea that 'a scientist must actively construct an analogy between the known and the unknown'[112], and his assertions that *all* scientific inference is based on analogy.[113]

Again, this is another area where Barnes provides a deeper explanation for Ravetz's detailed descriptions of science as a craft.

Mutation of questions

Both Barnes and Ravetz discuss the idea of the mutation of the problem under investigation as it is further studied. In contrast to other areas, Ravetz goes deeper into this idea than Barnes does, with Barnes only giving a brief

---

[110] Analogy seems to feature heavily in much of SSK. It would seem that an acknowledgement of the importance of social processes predisposes one towards analogical explanation. Obviously this would follow from the links between similarity relation theories of concepts and social theories of science, but it also makes sense on a lower level, in that acknowledging the importance of social processes increases the importance of communication in science, and the analogy is used very often in communicating scientific thought.

[111] see op. cit. 2, p 111, in that to test for membership in a set, one must establish similarity in important properties with the other members of the set. That is by establishing similarity in important properties with other members of the set, you are comparing your current situation to other, known situations, and thus using analogical reasoning.

[112] op. cit. 1, p 49

[113] see previous footnotes, numbers 66, 67,68, and 69

explanation of the idea that problems do not generate candidate solutions automatically[114], whereas Ravetz looks into the idea that a major part of the work of research is in the formation of the question. The idea that your question can mutate as you work relates to the changing of concepts over time, through the obvious mechanism that your question is representative of your conceptual make-up as you consider a new area. As you learn more about this area, your concepts about it will inevitably change, changing the question you are asking. This principle also holds for disciplines as a whole, via the mechanism of the encapsulation of intellectually constructed objects of inquiry into standard 'facts', which allow for the field as a whole to have the results of work feed back into itself, changing the questions being asked. This is thus important to explain in any theory of concepts, being yet another instance of concept dynamics.

Finitism and craft knowledge

What underlies all of these particular items is the idea of science being about concepts that are open-ended, revisable, and set via social processes. Barnes uses the finitist idea of open-ended concepts in order to ground his ideas about the social factors in the addition and subtraction of instances to and from concepts, and, most importantly, about the learning among instances of the same concept. The idea that relationships between events that are considered to be instances of a set are important is what provides a basis for Ravetz's ideas about the importance of implicit knowledge, that is, knowledge that is neither explicit nor specifiable. In those respects it is similar to the knowledge possessed by practitioners of a craft. This type of knowledge has a number of attributes, some of which are: being learned through a long training phase, and being difficult or impossible to learn without a long period of training in mundane practicalities. This is *very* important for this thesis, because it implies that the greater part of human knowledge is not directly communicable. A large part of my thinking is an effort to understand

[114] op. cit. 1 p 50

how this could be the case.

Both of the works, Barnes and Ravetz, have at their core a recognition of the importance of social processes, and a recognition of the dynamic nature of science, and indeed everything. Both Barnes' learned similarity relations between finite clusters of events and Ravetz's intellectually constructed classes of things and events have at their heart notions of openness, revisability, and construction and negotiation via social processes. In their details, they deal with slightly different levels of explanation and theorising, but that is indeed what makes the synthesis of the two more interesting.

Now that we have some idea of Barnes and Ravetz and how the two fit together, we can begin to consider the picture that emerges from that synthesis, and to feed some of the information that comes from cognitive science into the picture as well, to create a sketch theory of concept dynamics.

Chapter 3 – The Importance of Implicit Knowledge

Now that we have looked at Barnes and Ravetz, and also circled through Cognitive Science, it is time to integrate this information into a whole, to create a sketch theory of concept dynamics – of the nature, acquisition, use, and evaluation of concepts.

From Barnes comes the idea of finite sets of instances making up a concept, Ravetz provides an enumeration of some of the concepts that are created, negotiated, and maintained, connectionism supplies some more detail to Barnes, and Classical Cognitive Science provides critical conditions and expectations.

Barsalou's retrieval-based framework

Earlier I mentioned the work of Barsalou, and his retrieval-based framework for dynamic knowledge representation. His theory states that people store an enormous amount of information in long-term memory about each concept, but only a small amount of information is abstracted from this each time the concept is used.

A more accurate quote:

> A person possesses a tremendous amount of loosely organised knowledge for a category in long-term memory… …However, only a very small subset of an individual's total knowledge for a category is ever active on a given occasion to represent the category in working memory… …Although certain core information may occur in most subsets of a category, much of the information in a subset is either context-dependent or reflects recent experience. Because contexts and recent experience are rarely the same, the same subset of information is rarely, if ever, activated as its representation.[115]

[115] op. cit. 27, p 93

Thus, concepts are structures created in working memory from the knowledge in long-term memory, and from recent information and contextual information.

However, Barsalou does allow for the existence of a 'conceptual core'[116] Of information which has become context-independent via repeated usage, and is relatively invariant over time[117]. Thus, highly discriminative properties are most likely to become context independent. More context-dependent information is either not as discriminative, resulting in low association except when certain contexts activate it, or is inferred from higher-order concepts. Barsalou also posits the idea of recent context-dependent information. That is, when a context is activated, it predisposes that context to be activated for some period of time, raising the context-dependent status to context independent for that period. Barsalou gives an example: 'if someone had frog legs at a French restaurant one evening, an encounter with a frog in the back yard the following day might bring the edibility of frogs to mind. But encountering a frog a week after consuming frog legs may no longer do so.'[118] Hence, the three categories Barsalou uses for the classification of where the information from a concept comes from are: context-independent, context-dependent, and recent-context-dependent.

Barsalou then notes that most theories of concept similarity consider intercategory similarity, but 'if category representations are unstable… then theories of similarity must consider another kind of similarity'[119], what Barsalou calls *intraconcept* similarity. That is, the degree to which the knowledge about the concept is similar across the three categories of information Barsalou defines. According to Barsalou, it is this kind of measure that can measure the differences in concepts between cultures:

---

116 ibid. p 94

117 He does this while quoting another study by himself and Bower (Barsalou and Bower (1980)) {TODO: Insert full reference.}

118 op. cit. 27, p 97

119 ibid. p 102

> Intraconcept similarity depends on (a) whether concepts are constructed by members of the same or different populations, (b) whether concepts are constructed by the same or different individuals, (c) recent experiences with the category, and (d) current context.[120]

Also, this type of consideration of intraconcept similarity implies that *inter*concept similarity is dynamic, variable. This account is relevant to the study of science when we consider sciences to be, to some extent, different cultures, with different rules and implicit information. Thus, any account of concept dynamics also needs to explain these effects.

These contextual effects are important, as they allow to a large extent for the effects documented by Barnes and Ravetz of implicit learning and of learning similarity relations between finite clusters of events. The main reason for explaining Barsalou's work is that my own is similar to it in some respects, and some of his ideas, such as the three categories of information he uses, have been useful in writing the next section, my own work.

The process of learning using sets of instances

My theory is that concepts should be thought of as sets of instances that are operated on at the time of use by a process that extracts relevant features from all of the instances, based upon the network of similarity relations among the instances, both those 'in' the concept and those 'outside'. This process I will call the 'generalisation process', as it generalises from the large set of data (instances) to form a token (representation in working memory, to use Barsalou's language.)

'Instances' is thus also a problematic word. I will be using it to describe events that have been classified as relevant or not relevant to any particular concept. Unclassified events I will refer to as 'events'. Using 'instance' in this way allows for instances to serve as all three of Barsalou's types of information.

[120] ibid. p 114

That is, instances can act as context-independent information, by being positively associated with a large number of other instances, as context-dependent information, by being associated to the other classified instances via a group of contextual instances, and as recent-context-dependent information, by being recently acquired instances, which are thus slightly more related to the other instances.

Now, we shall begin at the beginning, the acquisition of an instance that is representative of a new concept. The process would go something like this (keeping in mind that there is much to be explained from this simplified version):

An event is observed that cannot be classified by similarity to other events. It is thus unclassified.[121] At some time equal to or later than this, the event is classified by some process[122] as being an instance of something. The event is now an instance of a concept. At the time of the first classification, because there is only one instance, it is very difficult for the generalisation process to extract the features that are crucial for concept membership. Thus, the probability of other events being judged as instances of the concept is low, without some outside force intervening, as there is no information on what members of that concept have in common.

At a later time (or set of times), more events are observed and classified as being instances of the concept. As the number of instances increases, it becomes easier to generalise the common properties of the concept from its instances, and thus easier to judge on the basis of experience what should and should not be considered instances of the concept.

[121] or possibly classified as 'unknown'. However, this would go against my definition of 'instance' for this work, and thus consideration of this must be deferred.

[122] To avoid paradox, and make the system work at all, it is necessary for some classification (mainly the early stages) to be imposed by an outside force. In practice this is generally an authority figure of some sort. It may be a teacher, a parent, or even a textbook. I will discuss this in more depth soon, but one can see the outlines of this argument in Barnes' examples of the importance of learning by ostension.

It can be seen from this that the process is self-sustaining once a certain amount of information (viz. number of instances) is available for generalisation. However, what is of primary interest in relation to the Ravetz material in particular is the effect of having many, many events classified as instances. As the number of instances grows, not only does the general concept become clearer, but sub-concepts, intra-concept groupings, begin to occur. Take the swan example used in Barnes. The child who is shown two instances of swans, the young person who has some experience of swans, and the world-renowned swan expert, are going to have very different ideas of 'swan'. The child's understanding will be vague and unformed. The young person will have a fair idea of what is a swan, (maybe even a knowledge of the black swan case), and the swan expert will have an intimate knowledge of *exactly* what a swan is, possibly to the extent that they may avoid using the more general case in favour of the specific subtypes of swan. The effect can also occur in the other direction, namely, the child can be told that a swan is a bird, and thus be able to include all of the 'swan' instances in the 'bird' set. Also, once the child's ideas about swans become clearer, and he has chances to experience other animals the community judges to be 'birds', then he will learn that all of these other animals are also referred to by the name 'birds', and thus he will be able to build up a representation of a higher-level concept from sets of instances of lower-level ones, with all that is required being ostensive learning of the meta-category.

For another example, consider a set of events. All of these events have in common, to us the objective-for-purposes-of-explanation observers, experiences of water. In each event, the person we are watching with our omnipotent gaze is able to judge that something about the event is related to water. It might be an event early in the person's life, and a parent or other authority figure is explaining that the word for the clear liquid is 'water'. It might be a number of experiences involving similar liquids before this, or it might be a teacher explaining that water is made up of Hydrogen and Oxygen and thus can also be called H_2O. Similarly to the previous example, it might also be an environmental scientist judging by sight alone that a water sample, indistinguishable to another observer, looks odd. Also, as the student learns more chemicals, she will learn the classifications of these chemicals as new concepts, with their instances being the sum total of each of the 'chemical' concepts, that is, Hydrogen, Oxygen, Water, and so on.

The important things here are the classification of similar events under a heading. It is imperative that the heading come from 'outside' the observer to begin with, or circularity will ensue. That is, logically, the system is dependent on a previous system. This would seem to lead to an infinite regress, a 'chicken and egg' problem. However, the nature of the system itself precludes this. Because perfect certainty is unobtainable (it would require an infinite number of instances), concepts can be used without understanding them on the expert level, or even on any level above the initial, one-instance level. Because of these two effects, it is possible for the (initial) learner to end up understanding the concept far better than the (initial) teacher. In this way, the system, over generations of concept-making, can bootstrap itself up into higher levels of functioning, into more complex concepts. Also, the individual, once he has learned a number of meta-concepts, or higher-level concepts, can take the instances of having learned the higher-level concepts, and create a new concept that represents the process of creating higher-level concepts. If the person wishes, this process can proceed *ad infinitum*, creating rules about rules about rules about rules, but in practice, it seems that most people pick a level and stop there.

Next, in communicating to others the idea of a concept, for example 'water' or 'H_2O' it is obviously impossible to communicate every nuance of every instance the observer has experienced. This creates a problem, because without every nuance of every instance, any generalisations for the instances will be different, in the fine detail.[123] That is, because the generalisation process relies on the differences and similarities of all of the instances, small changes in the set as a whole could possibly create large changes in the

[123] Note that this effect can work between people, and between occasions for the same person.

generalisation. However, this problem is solved in practice by the generalisation process itself.

The process of generalisation is, by definition, the condensing of a large number of instances, or a large amount of data, into a smaller, more compact item of information. In the case of concepts, generalisation is achieved by extracting relevant features from the data and creating an item of information that can be used to approximately reproduce these features.

In the case of concepts, this is generally a word or other token. Words are used more for communication, while some other token appears to be used in thought, as it is possible for two people to think of an extremely similar set of instances using different words (for example, the different language's words for 'up' and 'down' something which all people born and raised on Earth will have in common).

Thus, I am suggesting that concepts be thought of as sets of instances that are generalised from to create another instance, a meta-instance, which can be thought of as a concept. These concepts function in the same way as the instances below them, that is, they are grouped together by similarity relations, and they can be generalised about to form higher-level, complex concepts. The broader concept can also, with a sufficient number of instances to provide detail, be subdivided into more specific sub-concepts. Thus thinking of concepts as sets of instances provides a means of generating the tree-like conceptual structures seen in classical cognitive science, without requiring the building process to begin at the top or at the bottom. This is an important point, as one of the greatest problems of the conceptual tree is what is at the root. If we think of concepts as sets of instances, it can be seen that the tree can be built from any point, rather than only from the top or the bottom.

A more formal explanation

Given that I have been working with these ideas for some time now, I can see that all of this may be unclear without further explanation. To this end, I will elaborate, with some graphical aids, and some pseudo-mathematical symbols. These two aids have helped me enormously in clarifying my own ideas about sets of instances, so hopefully they will also be of assistance to you.

Let N_i be the number of instances classified under the concept. Let $P(\in)$ be the probability of correctly judging, that is, judging according to the rules of the community the judger is a part of, a new event to be an instance of the concept. (Being a probability, $0 \leq P(\in) \leq 1$) Let G be the difficulty of the generalisation process. This varies in inverse proportion to $P(\in)$.[124]

Step 1: The first instance.

Here is a graphical representation of the concept 'space' after the acquisition of the first instance. In this 'space', points are taken to represent events. The dimensions these events are classified on are dependent on the events themselves. In these diagrams, the distance between two points (events) indicates their degree of similarity. Obviously, in this case, the concept space is only two dimensional, whereas in actuality, it would be n-dimensional, where n is dependent on the concept under consideration, that is, where n is the number of salient properties in the concept. Thus, shapes drawn in the two-dimensional space are representative of sets of instances grouped by similarity. I should emphasise here that I *do not* think that this is what *actually* occurs in the learning of concepts. However, I think that this serves as a useful mnemonic device, a useful set of instances, if you like.

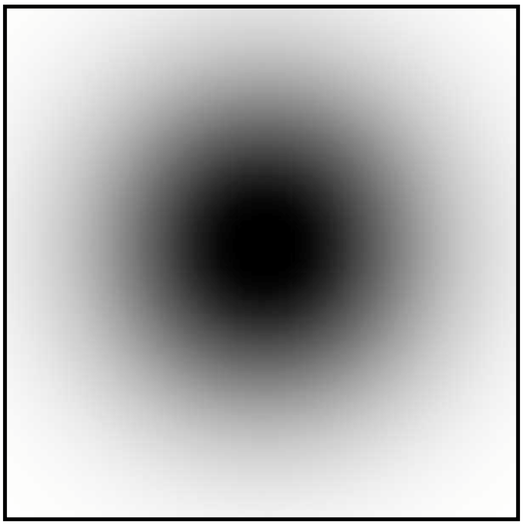

Figure 4 - Initial Stage N_i = 1

In the middle of the dark area is the point representing the initial instance. The darkness of any one point

[124] More correctly, G varies as follows: $G = \dfrac{1}{\sqrt{(P(\in) - 1)^2}}$

represents the certainty that the event that fits in that point will be correctly judged to be a member of the concept. That is, this diagram is a graph of P(∈). The other point to be made is that, for this case, G is effectively infinite, because there is only one instance, and thus there are no properties in common between instances. In this case, then, we have a newly acquired concept, that will be difficult to generalise properties from. This is indicated by the fuzzy outline.

Step 2: After the addition of some instances

Next we have the same concept, after the addition of a small number of other instances. Here it can be seen that the solid dark area in the centre has increased (that is, P(∈) has increased in that area), and also that the 'fuzzy' area has decreased in size. That is, P(∈) has increased overall. Also, in this case, G has decreased significantly. So, concept learning will be easier than in the first case and less subject to error (with respect to the appropriate community).

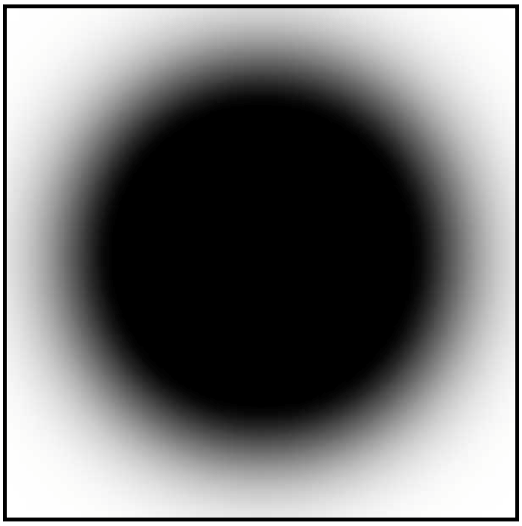

Figure 5 - $N_i > 1$

Step 3: Mastery

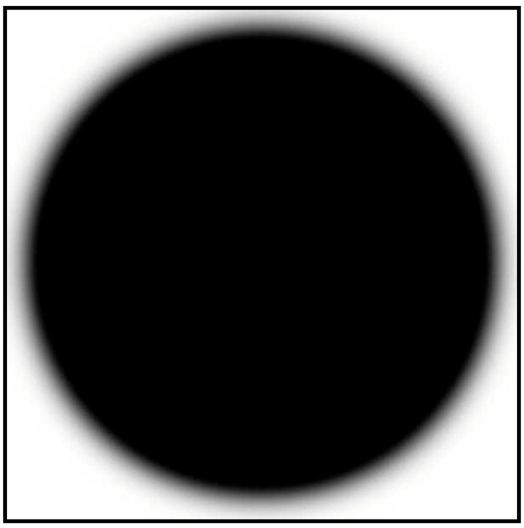

Figure 6 - $N_i >> 1$

This diagram is of a master's conception of the concept. It can be seen that the solid black area is very large, and the fuzzy area is very small. In this case, P(∈) is large, close to 1 in fact. Also, G is very close to zero. As more instances are added, some very efficient process comes to bear, and generalisation actually becomes more efficient. Thus, the master is able to deal with most cases without any uncertainty, much faster than someone without his intimate knowledge would be able to.

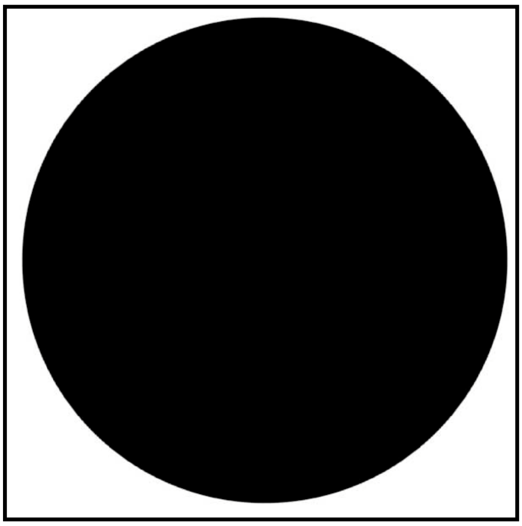

Figure 7- $N_i = \infty$

Step 4: Limiting Case

This is the limiting case, i.e. $N_i = \infty$. In this case, $P(\in) = 1$, and $G = 0$. Obviously this is impossible, but it is useful to consider this limiting case. In this case, every possible instance of the concept has been experienced, and judgements of new instances are thus meaningless. However, like many limiting cases, it is useful to look at to give an idea of the direction the process is heading.

Why do this?

Hopefully, the graphical aids will have cast some light on my suggestions about the gradual clarification of the concept. Of course, this simplified version did not include intra-concept learning, the subdivision of concepts, or the building of concepts into higher-level concepts, which I wrote about in the previous section.

Problems with sets of instances.

A problem with this type of explanation lies in the failure of generalisation to create exact replicas of information. If words are generalisations, and generalisations inevitably result in loss of information, which might or might not be relevant, how can we be sure that we have the truth, in the hard, philosophical sense? Answer: we cannot be sure *in principle* that our concept is the literal 'truth', the same everywhere throughout the universe. However, we can be sure *in practice.* That is, if we have a set of experiences that are very similar to other peoples, for example of the effects of gravity, then we can be *reasonably* sure that we have the right idea about gravity, at least in our own experience. (If we study general relativity, that is a whole different story, as we will eventually have a lot more detail about gravity, and will be able to differentiate between various types of instances of gravity, i.e. those that can be approximated to Newtonian mechanics, and those that cannot.) This idea is invaluable, as it is the only way out of the infinite regress that would otherwise ensue, that is, with the student needing a teacher, who needed a teacher, and so on. If concepts can be used *in practice* without obtaining a perfect (or even very good) understanding of them, then the conceptual system, considered as a whole over generations of people, will be able to bootstrap itself to higher and higher levels of sophistication, as concepts are assimilated into the cultural tradition,as mentioned previously.

It is this failure of generalisation to create exact replicas of information, however, that allows for the phenomena of implicit knowledge. Implicit knowledge as thought of by Ravetz is the product of a learning process, the reason that masters of their discipline are better than those who only know the basics. In teaching, the teacher uses her own understanding of a concept, based on her own set of instances, to create a generalisation which she then teaches the student. To student, to begin with, has a small number of instances compared to the teacher's (one of which is the teacher giving them the teacher's generalisation), and thus does not know the fine detail of the concept. As the student's experience with the concept grows, however, he will have more instances to generalise from, and will have more detail in his understanding of the concept, until he knows the concept in all its particularity, to use Ravetzian language. The concept would also seem to begin dividing into sub-concepts as the student journeys from novice to master.[125]

Sets of instances and Fodor's criteria

Now that the theory of concepts as sets of instances has been explained, we

[125] This is a requirement for compositionality, and I have discussed it before, on page 52.

need to look at the theory in the light of Fodor's criteria for theories of concepts, and provide an account either of how this theory satisfies the conditions, or why the conditions do not need to be satisfied.

Firstly, 'concepts are mental particulars'. That is, concepts are mental things. I think that the theory of sets of instances satisfies this untendentiously.

Secondly, 'Concepts are categories and are routinely employed as such.' Again, the sets of instances are by definition categories, so this criterion is fulfilled.

Thirdly, 'Compositionality: concepts are the constituents of thoughts, and in indefinitely many cases, of one another.' I have already explained (p 54) how the similarity relations between sets of instances allow for both the creation of sub-concepts, and the creation of meta-concepts, without requiring any concept to begin with. In this respect, I think that sets of instances more than satisfies this condition.

Fourthly, 'Quite a lot of concepts must turn out to be learned.' The above conclusion about conceptual hierarchies implies that is possible that everything is learned. That is, because of context effects and the ability to create meta-instances, it is possible to learn seemingly innate things like what constitutes equality based on a number of instances of being told that things are 'equal'.

I am not sure whether there are *actually* no innate concepts. However, it would seem to run very counter to many observed effects to note that *nothing* is innate. Unfortunately, further discussion of this must wait for another project.

Lastly, publicity. 'Concepts are *public*; they're the sorts of things that lots of people can, and do, *share*.' I have explained[126] why I think that this constraint should be deprecated. That is, concepts act as though this constraint holds true as long as there is no awareness of misunderstanding. *As long as people believe they are talking about the same thing, then it is as though they are, regardless of any 'identity' status.*

It can be seen that Fodor's constraints, even though one may not agree with them, are useful to consider, if one wishes to create any theory of concepts. Cognitive scientists have spent far more time thinking about this issue than sociologists, and so it is valuable to include their input.

However, it must be remembered that a large reason for the creation of this project was that thinking of concepts as sets of instances allows us to create an explanation for the implicit knowledge effects enumerated in Ravetz, amongst other things. How?

Implicit knowledge and sets of instances.

Explaining my theories about sets of instances is enables me to use that material to shed some light on phenomena such as implicit knowledge, or other phenomena involving over-learned[127] tasks, that is, phenomena common in expert disciplines.

Ravetz's use of implicit knowledge is in the context of learning the methods of science, as part of the objects of inquiry: he likens these methods to the work of craftsman, in that they are learned only by dint of much practice, and that they can be amazingly quick to operate once learned.[128] The same also goes for any other objects of inquiry, for the same processes underlie all problem solving on intellectually constructed objects.

[126] On page 12.

[127] I ran into this word several years ago whilst studying psychology. It is used in psychology to describe a task that has been performed so many times that its operation becomes automatic to some extent.

[128] See the quotes referenced in footnotes 77 and 83, and my own discussion on page 53.

To explain these phenomena, I propose to add to the previously outlined picture the assumption that some kind of processing is possible without the generalisation process, with the corollary being that this processing is much faster. This is, of course, a reasonably tendentious statement, and so I will now explain why I am proposing such a thing.

Processing without generalisation.

The way I see generalisation, it operates to abstract a sort of ‘mental shorthand’ from a large set of instances. This ‘mental shorthand’ is what is usually spoken of as a concept token, that is, that which you must have to ‘have’ a concept. So the generalisation process is involved whenever a person is thinking in terms of concept tokens. In the real world, these would seem to most often be words. Thus, generalisation is occurring whenever the person is using words.

However, my suggestion is that this sort of tokenised processing is not the *only* sort of processing that can occur. If the person is not thinking using words (or concept tokens), then the generalisation process is not running, and the person may well be proceeding in a similar fashion to connectionist models in AI. That is, at these times, there is no Turing-machine-like sequence of operations on symbols, just inputs to a system (events) being processed via a set of connections between the units of the system (previous instances), resulting in addition of units to the system, and as a side effect, behaviour without conscious thought.

In support of this, I suggest an exercise. Think of last time you were performing an exercise that is reasonably automatic, for example driving, walking, or riding a bike. What were you thinking about? Chances are, you were not thinking explicitly about what gear you were in, which foot gets lifted next, or how to shift your weight to stay balanced. You just did it, and this allowed you to also be doing other things, like listening to music, thinking about your day, or planning an honours thesis!

These are examples of behaviours that seem to involve this lack of linguistic thought. It is no coincidence that all of them are behaviours that most people have learned and practiced until they became second nature. This is the effect of overlearning. It pushes the processing requirements below the operations-on-symbols level, into the connectionist, distributed processing level. It is also no coincidence that processes like these seem to be the ones that work best when simulated via connectionist models: any task involving pattern recognition especially.[129]

However, the area of automatic tasks that are still difficult to simulate are those involved learned movements of the muscles. Any athlete will tell you that there is a huge difference between knowing intellectually how to do something, and actually being able to do it when the task involves muscular movement. Riding a bike is a good, everyday example. It is difficult to tell someone how to ride a bike, but once they get on, one can watch what they are doing, and give them ideas on how to improve.

There is nothing really new here, so far. All of this has been said one way or another before. I have just synthesised it here. However, what I am suggesting that I think is novel, and what I think throws light on Ravetz, is the idea that *any* task can be over-learned in this fashion. Even tasks involving objects of inquiry that are usually operated on via computational processes, such as mathematics or scientific methods, can be learnt so well that the person no longer requires conscious effort (computational processing) to complete the task.

To return to my previous model, this is what suggested to me that concepts that are less well-known, more vague, are those that require more processing, and must be processed at greater levels of generalisation. As the concept acquires more instances and becomes more concrete, tasks requiring it sink

---

[129] Chapters 1, 2, 4, 5, 7, 8, and 13 of Waltz, D. & Feldman J.A. (Eds.) *Connectionist Models and their implications: Readings from Cognitive Science*, Ablex Publishing Co, Norwood, 1988, contain excellent examples of this.

towards the unconscious processing level. More correctly, the more coherence there is among the instances, the less complicated the generalisation process is, and the lower the concept can be dealt with in the mind. This idea is what allows for the phenomena of implicit knowledge, for the intimate, unspecifiable relationship between a master and his tools. Language is simply not able to carry enough information at a high enough speed to allow people to communicate their exact mental states. Thus justifying the the creation of folk psychology, and from there the creation of cognitive science.

Chapter 4 – Illustrative Examples

Why include examples?

This question needs to be addressed before we begin looking at some real-world cases. I have chosen the examples here because both are expert disciplines that are quite successful in their chosen domain, and thus have some commonality with Science. The various sciences are also expert disciplines, at the level of interacting with tools, data, and so on. Thus increasing our understanding of expert disciplines in general must increase our understanding of the sciences. This is an important point, and so I will reiterate it: I consider the sciences to be on a continuum with other expert disciplines, in that the same underlying processes are responsible for the similar effects found in other expert disciplines. To argue for this, I will be looking at two expert disciplines most definitely not traditionally related to science at all, to see some interesting properties that arose in the consideration of concept dynamics reflected in new ways. From this I hope to shine some new light on some of the more ‘scientific’ areas, such as cognitive science, and give them some ideas for ways to look at other expert disciplines.

Firstly, we will be looking at an ancient expert discipline, or more precisely a tradition of practices and methodologies for becoming an expert, namely Zen Buddhism. I am aware that a lot of questionable material has been written about Zen by its proponents and detractors (in the sixties especially), but I will be looking at a version of Zen espoused by Thomas Cleary – a version based on the more practical Zen philosophers of Japan.

Secondly, and somewhat related to the Zen example, I will be looking at what is called hacker culture. Now by this I do not mean ‘hacker’ in its vulgar, media-driven sense of someone who breaks into computer systems for fun and/or profit[130], but someone who enjoys learning systems in their most intimate detail, who delights in the discovery of minutiae, and who knows computers in fine detail. Throughout the example I will be using the word 'hacker' in this sense, and I will use the word 'cracker' to indicate the media-driven meaning.

Zen: The science of training the mind

Zen Buddhism is maligned in some quarters in the West, and rightly so in some senses, as the Zen espoused by Suzuki, Jung, and other twentieth century popularisers is obscurantist and is arguably not true to the spirit of the more practical, deeper writing. Instead, their version of Zen is the inheritor of the institutionalised Zen so repugnant to many practitioners, and to Thomas Cleary, in his book *The Japanese Art of War*.[131] Cleary is a scholar of Eastern philosophy (evidenced by the number of books in his bibliography about such matters[132]), and is able to understand Japanese directly (he has translated a number of texts). Thus he is amply qualified to give some counterpoint to the discussions of Zen popularised in the 1960s and 1970s by Suzuki and his followers[133].

Zen Buddhism is a Japanese branch of Buddhism descended from Chan Buddhism in China. It focuses on individual experience, introspection, and meditation to bring about 'enlightenment'.[134] In this respect it is interesting because, effectively, it is a research program for the perfection of the mind.

---

[130] For an illuminating statement by one of the truly great hackers, Richard M. Stallman of the Free Software Foundation, on the media's usage of 'hacker', see Appendix C in Eric S. Raymond, (Ed) *The Jargon File*, available at <http://www.tuxedo.org/~esr/jargon/>.

[131] Cleary, T. *The Japanese Art of War*, Shambhala, Boston & London, 1992.

[132] For example, numerous translations of works on the Tao, including *The Art of War* by Sun Tzu, *The Inner teachings of Taoism*, by Chang Po-tuan, and Cleary's own *Vitality, Energy, Spirit: A Taoist Sourcebook*. Some of his works on Buddhism include translations of *The Flower Ornament Scriptures*, and *Shobogenzo: Zen Essays by Dogen*. All of these works are published by Shambhala.

[133] Cleary is quite scathing in his treatment of Suzuki, especially with references to the idea of 'no thought', on p 47-48 in op. cit. 131. This continues throughout the book. (see also pp 49, 51, & 62)

[134] From ibid, Ch 1.

Not for the studying of the mind, but for its improvement, for the movement towards an ideal mind. This is a very important distinction to make, and one that will be explored more when I consider Zen and the removal of 'sickness' from the mind.

In this example I will quickly review some areas of Zen that have relevance to this thesis. Firstly, I will look at the importance of implicit knowledge in Zen, (with obvious echoes to Ravetz), followed by a look at what exactly the Zen program is all about: the removal of 'sicknesses' from the mind. I will finish with a closer look at the Zen views of training.

Zen and the importance of implicit knowledge

Zen emphasises the importance of implicit, unspecifiable knowledge:

> In contrast to the formal transmission of doctrine, Zen emphasized "mind-to-mind" communication of the ineffable. This does not mean something like mental telepathy, however, as it is ordinarily imagined. In early Chinese Zen lore the mind-to-mind acknowledgment is likened to two mirrors reflecting each other with no image interposed.[135]

That is, Zen emphasises the learning of knowledge that cannot be communicated without a deep knowledge of the other person, without a large amount of common knowledge. At the heart of Zen is the belief that some knowledge cannot be taught easily, that it can only come from long experience. Obviously this has many parallels to what Ravetz has to say about science being like craft-work, requiring a long apprenticeship, relying on implicit methods, and so on.

Because of this belief, classical Zen works[136] tend to use unusual means to

---

[135] Ibid, p 106

[136] As opposed to the modern expositions by Suzuki et al., although these also are laid in out in an unusual way, the problem with Suzuki is that the unusual means become the end, and the metaphors, koans and so on become the essence of Zen. Cleary is reading these classical texts in a very different way that is much more relevant to this

teach. They use tangential questions, metaphor, poems, and koans, explaining only in the most simple terms, in language everyone can understand, then leaving the teaching to come from the interaction of the person's own experience and the carefully chosen example. Tangential questions serve to connect things in the student's mind that were not previously connected, while koans are unsolvable riddles, or seemingly meaningless stories that seem to me to be intended to expose to the student the way that she thinks. That is, koans are carefully constructed so that, as the student thinks about them and speaks to the teacher about them, the student is brought to a greater understanding of how her own mind works.

Metaphor in particular is very important. Musashi Miyamoto, in *The Book of Five Rings*[137] used the metaphor of the master carpenter for leadership, following Taoist classics[138]. Miyamoto says that 'it is the duty of the master carpenter to understand the regulations of the country', and other high-level tasks. One of the more symbolically loaded passages is about the choosing of wood:

> When sorting out timber for building a house, that which is straight, free from knots, and of good appearance can be used for front pillars. That which has some knots but is straight and strong can be used for rear pillars. That which is somewhat weak yet has no knots and looks good is variously used for door sills, lintels, doors and screens. That which is knotted and crooked but nevertheless strong is used thoughtfully in consideration of the strength of the various members of the house. Then the house will last a long time.

---

study.

[137] Musashi, M, (Trans by Cleary, T.) *The Book Of Five Rings*, Shambhala Pocket Classics, Boston and London, 1994. Musashi lived from around 1583 until the latter part of the seventeenth century, and was renowned for being an amazing swordsman. Apparently he killed his first man in a duel at thirteen, and later in life, after mastering his particular style of two-sword fighting, began using wooden swords against his opponents' metal ones. The exceptionally amazing thing is, he never lost a duel. It is also said that during the last three decades of his life, he 'never combed his hair, never took a bath, never married, never made a home, and never fathered children.' (ibid, p xviii).

[138] op. Cit 131, p 25

> Even knotted, crooked and weak timber can be made into scaffolding, and later used for firewood.[139]

The meaning of this metaphor is never given explicitly, forcing the reader to do their own hermeneutics, which allows the reader to bring to the structure whatever context they wish.

In this, Zen shows an understanding that people do not only learn what is taught on the surface, but also what follows from the context, from the norms of the culture they are being taught in, and so on. Again, this mirrors what Barnes, Barsalou, and myself previously said about the importance of contextual learning, of learning the implicit rules at the same time as the explicit ones.

The removal of 'sickness'

Another of Zen's ideas is the seeking of enlightenment via the removal of 'sickness'. In Zen terms, 'sickness means obsessive thought'.[140] Sicknesses are fixations and biases that serve to distort your view of the world. 'This last function of the Zen technique depends on sharpening discernment of the fine web of subtle causal relations by removing the veil of mental preoccupations'[141] says Cleary. In this respect it sounds somewhat similar to Bacon's various Idols (of the cave, of the tribe, of the market, etc.) in that both are biases to be removed.

However, Zen also has an interesting caveat. It is also a sickness to concentrate on the removing of sicknesses. That is, if you spend too much time trying to remove your sicknesses, you will neglect other things and that otherwise worthy behaviour will become a sickness. 'To think of riddance is

[139] Musashi (op. cit. 137), p 11-12

[140] Cleary (Op. Cit 131), quoting Zen Master Yagyuu Munenori.

[141] ibid, p 33

itself sickness'[142] One could suggest that this is good medicine for those sciences obsessed with 'method'. If too much time is spent debating about what constitutes the 'correct method', then no time will be spent doing the research, developing the objects of inquiry to enable people to know what the best methods for dealing with such objects are, and to enable people to set criteria of adequacy and value.

However, by far the most important point for this example is the Zen view of learning.

Zen and Learning

One of the hardest things to comprehend about Zen is the Zen view of learning. There is a lot of talk about 'mindlessness' and 'no thought', and it is difficult to see how this can have any meaning.[143] There are also a lot of strange sounding notions about the circularity of learning that are very hard to understand. However, I will attempt to 'enlighten' you on this front.

To begin, a quote from Zen Master Takuan, written to Yagyuu Munenori[144]:

> You need to realize that when you practice from the state of the beginner all the way to the stage of immutable wisdom, then you must go back to the status of beginner again.
>
> Let me explain in terms of your martial arts. As a beginner you know nothing of stance or sword position, so you have nothing in yourself to dwell on mentally. If someone strikes at you, you just fight, without thinking of anything.[145]

---

[142] ibid, p 37, quoting Yagyuu.

[143] See footnote 133 for some references Cleary expounding his view and attacking Suzuki on this.

[144] Takuan was 'an elder contemporary of Musashi and teacher of the Shogun and the master swordsman Yagyu Munenori'. (Cleary, op. cit. 131, p 25) Munenori was more like Musashi in that he was also a warriour who took an interest in Zen. All of these people lived during the early Tokugawa era of Japanese History in the early seventeenth century.

[145] ibid, p 29.

To relate this to physics, when you begin learning, you have only your intuitive knowledge of the world to guide you. When you make judgements about the world, you use these well-learned ideas to make very quick decisions about the world. An example would be catching a ball.

> Then when you learn various things like stance, how to wield a sword, where to place the attention, and so on, your mind lingers on various points, so you find yourself all tangled up when you try to strike.[146]

When you begin learning physics, you have to learn many unfamiliar things. Because these things are new and you do not fully understand them, it can be difficult to work with them.

> But if you practice day after day and month after month, eventually stance and swordplay don't hang on your mind anymore, and you are like a beginner who knows nothing.[147]

But if you practice doing physics, doing experiments, using lab equipment, eventually all of your worries about doing things correctly will fall away, and you are then able to do things that seemed difficult or impossible when you were learning them with ease.

This is exactly what I am saying about processing without generalisation. To begin with, you have few instances, and must generalise from them concept tokens in order to work with them. However, as you learn more and more, the similarity relations between instances become better established, allowing you to more easily classify new instances, and eventually allowing you to work with some concepts without much cognitive effort at all.[148]

[146] Ibid, p 29

[147] ibid. p 29

[148] I should remind the reader of two points here to help this to fit into the larger picture. Firstly, Barnes and Ravetz, in the Kuhnian tradition, were both arguing against some formalised, bottom-up learning of a science and were in favour of a more similarity-based, craft-like method. These Zen ideas amply demonstrate these properties. Secondly, that expert negotiation in research in continuous with the processes of training, that is, the researcher uses the same processes to learn about phenomena as does the student, but the researcher is using these skills to produce results that must be evaluated by the community as a whole rather than by a teacher, textbook, or other simple authority figure.

This is perhaps one of the more appealing aspects of Zen to other expert disciplines. Although they may not have many people studying them, other expert disciplines often seem to find a lot that is meaningful in the Zen ideas of learning and training the mind, and in its ideas of what the mind should seek to be. Musashi was a very good example of this, applying Zen teachings to the art of sword-fighting. A modern day example is the hacker culture that began at the Massachusetts Institute of Technology's (MIT) Artificial Intelligence Lab.

Hacker Culture: From MIT to the world.

Before we begin looking at hacker culture, I should make readers aware that I am a member of the hacker culture to some extent. I would not classify myself as a hacker, but probably as a proto-hacker[149], one who is on the path to being a hacker. This is important for this example because it enables me to see the pattern of the culture from my own experience as well as from the references provided.

Hacker culture, as the technical discipline of using computers (and indeed any other system) to its maximum potential, is an interesting study. To begin with, it is interesting to look at what Barnes, Ravetz and myself have to say, taking this culture as background instead of science. The features described by Barnes, Ravetz and myself are there, but can be intriguingly different in some ways. Another thing that is very interesting about looking at hacker culture is that here we have a chance to look at the foundation and development of a technical discipline and its tools, *and* access to primary sources about it. Thus, in looking at the hacker culture as an expert discipline, we are able to access an even larger volume of more accurate information than if we were looking at the beginnings of science as a technical discipline. Before looking at the properties of the discipline, however, we need a brief overview of the

[149] see 'proto-hacker' in the Jargon File (op. cit. 130)

culture's history.

A brief history of hacking

Hacker culture began at MIT in the late 1950s.[150] To begin with, the people who were later to become the first hackers were all members of the Tech Model Railroad Club (TMRC). This was a club at MIT whose members built and maintained an enormous model railroad layout. There were two factions in the club, the trainspotters, who were interested in trains, modelling, and keeping the layout close to reality, and the Signals and Power (S&P) subcommittee, who were interested in the vast system of relays and so on that controlled the trains. This all changed, however, when the TX-0 arrived. This was one of the first *interactive* computers, that allowed you to change how the computer was running while it was running. Once the S&P faction had access to the TX-0, the change was underway. They began developing programs to do a whole host of things, converting Arabic numerals to roman numerals, and creating human-readable languages that would then be interpreted by the computer into something it could read, so that programs could be written without using the actual ones and zeros that the computer needed to run.

As the years went on, the hackers gradually became a part of the influential MIT AI lab. It is here that the culture of hacking became joined with the culture of Cognitive Science. As the Golden Age of the MIT AI lab hacking culture began to fade, it was picked up by Stanford University, and again by the computer games companies of the late seventies and early eighties, and once more in the 1990s by the Linux movement, dedicated to the creation of a freely available and modifiable operating system for personal computers.

One thing that has developed over this history has been the word 'hack'. Levy

[150] The information in this section is summarised from Levy, S. *Hackers: Heroes of the Computer Revolution*, Penguin, London, 1994 (first published 1984)

notes that 'hack' had long been used 'to describe the elaborate college pranks that MIT students would regularly devise'[151], but the new programmers took the word in a new direction. It acquired a legion of subtle, context shaded meanings revolving around the application of ingenuity, and which can thus send some light on the workings of implicit knowledge and contextual effects.

The meaning of 'hack': implicit knowledge in the real world

Firstly, The Jargon File defines 'hack' as follows:

> 1. n. Originally, a quick job that produces what is needed, but not well.
>
> 2. n. An incredibly good, and perhaps very time-consuming, piece of work that produces exactly what is needed.
>
> 3. vt. To bear emotionally or physically. "I can't hack this heat!"
>
> 4. vt. To work on something (typically a program). In an immediate sense: "What are you doing?" "I'm hacking TECO." In a general (time-extended) sense: "What do you do around here?" "I hack TECO." More generally, "I hack `foo'" is roughly equivalent to "`foo' is my major interest (or project)". "I hack solid-state physics." See Hacking X for Y.
>
> 5. vt. To pull a prank on. See sense 2 and hacker (sense 5)[152].
>
> 6. vi. To interact with a computer in a playful and exploratory rather than goal-directed way. "Whatcha up to?" "Oh, just hacking."
>
> 7. n. Short for hacker.[153]

[151] ibid, p 23

[152] Here the entry is referring to another entry in the Jargon File, 'hacker', and to the fifth sense in that entry.

Even in this short definition, a number of points are obvious. Firstly, there are a large number of meanings, and secondly that some of the meanings are directly contradictory. The way this ambiguity is resolved in practice is through the use of contextual information. For example, the use of descriptive words: ‘sublime hack’ and ‘dodgy hack’ are good examples. Hacking is thus an extremely good example of a context-dependent concept. Another quote from later in the Jargon File:

> “The word hack doesn't really have 69 different meanings", according to MIT hacker Phil Agre. "In fact, hack has only one meaning, an extremely subtle and profound one which defies articulation. Which connotation is implied by a given use of the word depends in similarly profound ways on the context. Similar remarks apply to a couple of other hacker words, most notably random.”[154]

I completely agree. Writing as a member of the culture, I find that on reflection I can recall an astonishing number of uses for the word ‘hack’, but the differences between them are exceedingly hard to explain. When thinking on this issue, I attempted to create an explanation of how a ‘dodgy hack’ differs from a ‘sublime hack’, and without showing actual programming code, it is very difficult. That is, without being able to provide a person with examples, saying *this* is a sublime hack and *this* is a dodgy hack, it is impossible to describe what they are. This sort of effect is *exactly* what Barnes, Ravetz, and myself mean when we write of implicit, unspecifiable knowledge, knowledge encoded as similarity relations.

Indeed, this type of phenomenon really does illustrate a problem with Classical Cognitive Science type symbolic theories of the mind. Without a real appreciation of the issues I have discussed in this work, explaining how undefinable knowledge can exist is problematic. Firstly, because there is no place in a symbolic theory of mind for symbols that cannot be articulated and

[153] ‘hack’ entry in the Jargon File (op. cit. 130)

[154] ibid. Appendix C.

defined, and secondly, even if one does allow for some sort of symbol being indefinable in some way, there is no way that such a symbol could be passed from one person to another without the use of similarity relations. That is, if it is impossible to communicate *exactly* what is meant by a concept, then there is no way that those attempting to learn the concept would be able to. Again, in less formal terms: in a symbolic theory of concepts, if you cannot say something, it cannot be learned, and indeed cannot exist, because something that is not learned must be innate to exist, and these concepts are plainly not innate.

However, the fact remains that these effects do happen. And one of the more interesting properties of hacker culture is the awareness that the exact meaning of the word ‘hack’ cannot be communicated, the awareness that some knowledge *is* implicit and requires direct experience to learn.

Hacking and the awareness of implicit knowledge

One of the characteristics of hacker culture is the awareness that some things are difficult or impossible to talk about. In particular, there is a real appreciation for words that have a large number of subtle shadings of meaning, (such as ‘hack’). This is arguably one of the reasons that so many hackers see such a connection between hacking and Eastern philosophies such as Zen, or the Tao.

A quote from the essay *How to be a Hacker*:

> But if you think of cultivating hacker attitudes as just a way to gain acceptance in the culture, you'll miss the point. Becoming the kind of person who believes these things is important for you -- for helping you learn and keeping you motivated. As with all creative arts, the most effective way to become a

> master is to imitate the mind-set of masters -- not just intellectually but emotionally as well. [155]

Hackers realise the importance in training of seeking a certain sort of experience for a certain reason with a certain frame of mind. To put it in the terms I used earlier, hackers have learned implicitly that one must acquire sets of instances representing the important cultural concepts that are similar to others in the culture in order to be understood as a member.

The point is illustrated again in another essay, The *Loginataka* (this is intented to be humorous in some ways):

> O Nobly Born: learn, and seek within thyself. Cultivate the cunning of the Serpent and the courage of the Tiger; sup deeply from the Wisdom of those who came before thee. Hack, and hack again; grow, by trial and by error. Post thy best hacks to the Net and gain in Repute thereby. Also, O Nobly Born, be thou grave and courteous in thy speech; be helpful to those less than thee, quick to succour and slow to flame. If thou dost these things faithfully, if thou travellest with high heart and pure intention, soon shall thy callow Newbiehood be shed. By degrees imperceptible to thyself shalt thou gain Power and Wisdom, Striving and Doing all the while. Gradually shall thy Puissance unfold and deepen. O Nobly Born, if thou dost all these things, thy Wizardhood shall surely come upon thee; but not of a sudden, and not until after thy arrogant Mind hath more than half Forgotten that such was its Aim. For know this --- you may not by thyself in Pride claim the Mantle of Wizardry; that way lies only Bogosity without End. Rather must you Become, and Become, and Become, until Hackers respect thy Power, and other Wizards hail thee as a Brother or Sister in Wisdom, and you wake up and realize that the Mantle hath lain unknown upon thy Shoulders since you knew not when.[156]

This passage is written to resemble Eastern philosophical texts, but it is

---

[155] Raymond, E.S. *How to be a Hacker*, <http://www.tuxedo.org/~esr/faqs/hacker-howto.html>.

steeped in hacker jargon.[157] However, the sense of the passage is readily apparent. That is, only by participating in the culture can you become a fully-fledged member.

This recognition of the importance of implicit knowledge is missing from modern science. Indeed, the emphasis placed on formal logical thought in this century by philosophers of science has arguably alienated them from the practicing scientists in many ways. Any theory of science that has any pretensions at all to being a descriptive rather than normative theory must acknowledge that implicit knowledge effects exist. This does not mean that they must abandon logical thought or some such nonsense, it simply means that those teaching students need to realise the full import of what they are teaching. Teachers, and more importantly, those who construct curricula for teachers, need to realise the importance of implicit knowledge in learning important technical standards such as criteria of adequacy and value.

Some conclusions from the examples

One can see even from these two very brief examples that expert disciplines in general have much in common with the workings of science. It would seem that one of the major factors that sets the sciences apart for other expert disciplines is the methodological rhetoric of Science (capital S). Of course, since Kuhn this has been depreceated in some quarters, but the general perception of science is of a monolithic discipline whose goal is the ultimate explanation of *everything*. Under this thinking, one of the reasons science has been so spectacularly successful as an expert discipline is that its chosen domain is the very large set of physical phenomena, that is, a scientist can legitimately study almost anything, provided they can win the support of the community. Contrast this with other expert disciplines, that acknowledge that they are expert in a much smaller domain, and thus restrict their work to within that domain.

---

[156] Raymond, E.S. *The Loginataka* <http://www.tuxedo.org/~esr/faqs/loginataka.html>

[157] All of the terms used here are defined in the File, but in particular, Wizardry and Wizard are used here to indicate

a master capable of feats far greater than lesser mortals.

Chapter 5 – Conclusion

This project has looked at a large number of areas. cognitive science, philosophy, the sociology of scientific knowledge, artificial intelligence, Zen, and hacker culture. What it has shown is that:

- concepts must be thought of as *dynamic*. Only then can the fact of the changing face of science, and any other discipline, be rendered intelligible.

- science is an expert discipline, albeit a very refined one, and exists on a continuum of expert disciplines, some successful, others not. Because of this, other expert disciplines have much in common with science and may even be able to show those who study science new ways of looking at it that may help them better understand exactly what is happening.

Of course, one cannot simply claim that concepts are not what we thought they were without some replacement available[158], so I have provided a sketch theory of concepts that is reliant on a number of ideas:

Firstly, concepts are dynamic.

Secondly, concepts can be thought of as abstractions generalised from sets of instances, that can then be operated on by symbolic operations similar to those in classical cognitive science.

However, the latter point only applies when the concepts are not known very well. As a concept is mastered, the way in which it is processed will sink from the symbolic level to the connectionist level, with the person being able to respond from experience rather than by performing symbolic operations on

---

[158] Well, I could, but it wouldn't work. That's the whole point of allowing for negotiation in the sciences.

concept tokens.

Therefore the symbolic models of cognition apply in novel cases, while connectionist models apply in familiar ones.

Next, extension of concepts is based on analogy, that is, based on the formation of similarity relations between the new events and known ones.

Thus, the purpose of training becomes to give the student enough experience (viz. instances) to be able to make analogies to new instances with a minimum of effort. However, because of the sinking of concept processing towards the connectionist level, as the student learns more it can, without proper structuring of the curricula, lead him to institutionalisation, dogmatism, in other words, difficulty with applying symbolic calculation to concepts that have sunk to the connectionist level.

This is why a discipline is formed, to create curricula of training that will enable students to learn enough to do further research without teaching them to be dogmatic. The community that forms the discipline maintains these curricula through the social negotiation of the objects of inquiry as they are modified and generated by new research.

For those of us who study science, this realisation means that we should be also looking at disciplines other than science, to compare and contrast, and see if we can negotiate in our discipline what seems to make disciplines in general work better. If we all remember that we are in this for the same reason, in the end, then perhaps we can avoid or minimise the impact of intradisciplinary ideological battling, and get on with our work. Hopefully this thesis can provide some openings for this to occur.

References

1. Barnes, B. *T S Kuhn and Social Science*, Macmillan, London, 1982

2. Barsalou, L.W., 'Intraconcept similarity and its implications for interconcept similarity', Ch 3 in Vosnaidou & Ortony (reference #20), p 76-121

3. Boden, M. A., 'Horses of a different color?', ch. 1 in Ramsey, W, Stich, S.P. & RumerlHart, D.E*, Philosophy and Connectionist Theory*, Lawrence Erlaum Associates, Hillsdale, 1991.

4. Cleary, T. *The Japanese Art of War*, Shambhala, Boston & London, 1992.

5. Dennett D. 'Mother Nature versus the Walking Encyclopedia: A Western Drama', ch. 2 in Ramsey, W, Stich, S.P. & RumerlHart, D.E, *Philosophy and Connectionist Theory*, Lawrence Erlaum Associates, Hillsdale, 1991.

6. Dennett, D.C. *Brainchildren: Essays on designing minds,* Penguin, St Ives, 1998

7. Fodor, J.A. *Concepts: Where Cognitive Science went wrong*, Clarendon Press, Oxford, 1998

8. Fuller, R. B. *Synergetics: Explorations in the geometry of thinking*, Macmillan, New York, 1975.

9. Hesse, M, *The structure of scientific inference*,London, Macmillan, 1974.

10. Hesse, M. B. *Models and analogies in science*, London, Sheed & Ward, 1963

11. Leatherdale, W. H*. The role of analogy, model, and metaphor in science*, Amsterdam : New York : North-Holland Pub. Co. , American

Elsevier Pub. Co., 1974.

12. Levy, S. *Hackers: Heroes of the Computer Revolution*, Penguin, London, 1994 (first published 1984)

13. Musashi, M, (Trans by Cleary, T.) *The Book Of Five Rings*, Shambhala Pocket Classics, Boston and London, 1994.

14. Popper, K. R. *The Open Universe: An argument for Indeterminism*, Hutchinson, London, 1982

15. Ravetz, J.R. *Scientific Knowledge and its Social Problems*, Second Edition, Transaction, New Brunswick, 1996 (originally published 1970)

16. Raymond E. S, (Ed) *The Jargon File*, available at <http://www.tuxedo.org/~esr/jargon/>.

17. Raymond, E.S. *How to be a Hacker*, <http://www.tuxedo.org/~esr/faqs/hacker-howto.html>.

18. Raymond, E.S. *The Loginataka* <http://www.tuxedo.org/~esr/faqs/loginataka.html>

19. Schuster, J.A.S., *The Scientific Revolution: An Introduction to the History and Philosophy of Science*, Schuster, Sydney, 1995, p 50-51

20. Vosnaidou, S and Ortony, A (Eds.) *Similarity and Analogical Reasoning*, Cambridge University Press, Cambridge, 1989.

21. Waltz, D. & Feldman J.A. (Eds.) *Connectionist Models and their implications: Readings from Cognitive Science*, Ablex Publishing Co, Norwood, 1988

22. Weitz, M. *Theories of Concepts: A history of the Major Philosophical Tradtition*. Routledge, London and New York, 1988.

23. Westbrook, R.F. ‘Lectures on the psychobiology of animal learning’, *Course Material, PSYC2081, Learning and Developmental Psychology*, UNSW