Collected essays and conference proceedings

Kelly Alexandra Roe

2003

Contents

Chapter 1

A realist aspect to the interpretation of selves

Presented to the Australasian Association of Philosophy, New Zealand Division, hosted by the University of Auckland.

Presented to the Australasian Association of Philosophy, Australian Division hosted by the University of Tasmania.

Abstract.

Dissociative Identity Disorder (formerly Multiple Personality Disorder) has been the subject of much controversy: Are alternative identities most plausibly viewed as alternative selves; fragments of selves; or adopted role-plays? The answer to this question will be shown to be a matter of interpretation; but this need not be taken to imply that selves are the entirely non-realist products of social and narrative construction. Although Dennett explicitly regards selves to be such 'creative fictions' I will attempt to show that it is more consistent with his overall line, and more plausible in general to grant a degree of realism to selves. There may be no fact of the matter as to how many selves there are associated with a single body – but this need not be taken to preclude realism. If we accept a 'multiple systems' version of the intentional stance we can perceive more pattern in these subject's behaviour, thus more charitably making the best sense of their lives.

Characterisations of alters and selves

The phenomenon that was until recently known as Multiple Personality Disorder has prompted a variety of accounts as to the metaphysical status of the alternative identities, or alters exhibited by subjects with the disorder. Some theorists have held that the phenomenon shows that we need to reconsider the status and nature of selves in general as it seems that it is possible for subjects to give rise to, or support more than one self e.g., Gillett, (1997); Hacking, (1991); Humphrey and Dennett, (1998); Kolak, (1993). This has led them to conclude that selves are fictions, whether in the 'multiple' individual or in those with a more conventional psychology.

An alternative to this account is to deny that alters are selves. A popular strategy amongst clinicians who treat the disorder is to consider them to be 'aspects', 'segments', or 'parts' of a self. On this view alters are to be seen as fragments that may be blended or fused together to add up to a single self of the sort exhibited by individuals without the disorder e.g., Gleaves, (1996); Kluft, (1988); Putnam, (1989). The other major alternative is to deny that alters are selves, and maintain that there is only one self even in the case of subjects who present with the disorder e.g., Brown, (2001) and Clarke, (1990). There has been a general tendency for theorists who maintain that alters are selves to emphasise that selves are fictions. While the theorists who maintain that alters cannot be regarded as selves often do not explicitly examine the nature of the self, there seems to be an implicit tendency towards realism. The metaphysical status of alters thus seems to be somewhat inversely related to the metaphysical status of selves[1].

The reaction to realism

Historically the main way of being a realist about the self was to take a Cartesian view of them. In recent times, however, the notion of a Cartesian Realist self has largely been discredited for two main reasons. The first problem arises with Descartes' dualism; his notion of the self as a non- physical, or immaterial thing. For Descartes the soul, mind, or self is a simple and unified entity that acts as a locus of control. Just how this non-physical

[1]I do think that this is a fair characterisation of the main positions although it seems possible to read 'parts' of a self as either not being real in their own right, or as real parts of a real whole. Spanos is one theorist who maintains that alters are role plays, yet would seem to consider a single self to be something of a role play also. I see nothing to prevent either of these theorists agreeing with me in that alters are as real as selves.

self can interact with the physical world is the notorious causal problem for dualists. While some (notably Chalmers, 1996) maintain that dualism is not incompatible with a scientific world view; so long as we broaden our notion of science to include qualitative experience as irreducible; it would be fair to say that the majority of theorists remain unconvinced. Most theorists claim to work within a broadly naturalistic or materialist framework, and by this they seem to mean that there is no immaterial mind or self that acts as a locus of control. The notion of this kind of self is (for the most part) rejected. Often theorists consider themselves non-realists in order to explicitly characterise themselves as rejecting the above variety of realism.

The second reason for the reaction to Cartesianism is Descartes notion of a localised region of the brain that is responsible for executive control. Descartes considered the pineal gland to be the seat of interaction and while this has been ruled out as implausible the majority of theorists draw the greater lesson that there is no localised region that fulfils the role of the mind, self, or 'place where it all comes together'. This realist notion of the self as an entity to be found within the subjects brain has largely fallen into disrepute as theorists are unable to pinpoint an area or pattern whose proper functioning is necessary and sufficient for self-hood. Feinberg, (2001) writes that there is no place in the brain in which 'all the brain's activity converges on "one pontifical cell" '. He goes on to propose a *nested hierarchy* theory of self where diverse areas of the brain contribute to the consciousness and selfhood that are seen as emergent properties of the normally functioning human brain. He shows us through a series of case studies that the self is not an all or none affair. It can break down or malfunction to differing degrees, and it cannot be the sole product of any one localised region. It seems that the homunculus that is the self is not to be found in a localised region of the brain at all. No neuron, mental module, or pattern of activation seems to constitute a self. Neuroscientists largely agree that there is no localised self to be found in the brain. These varieties of realism are thus held in disrepute.

The materialist intuition that there must be a physiological basis to the phenomena that give rise to talk the self is not disputed. It seems unanimous that the self is (somehow or other) a *product* of neural activity. That physiological changes are physical *causes* of behaviour is not disputed. Non-realists just want to deny that the self will turn out to be one and the same as a set of neural processes. Neuro-physiological accounts of how the behaviour arises that constitutes DID have been offered in terms of competing mental modules (physical structures within the brain), or patterns of activation with respect to firing rates or transmitter levels. What little empirical

evidence we have is mixed as to whether the neural activity of DID subjects when switching between alters is qualitatively different from controls (Adler, 1999). When the studies report that there are qualitative differences the interpretations of this finding are hotly disputed, as is the methodology of the study. Non realists deny that the self will turn out to be found in the brain and they emphasise social and narrative processes that they maintain lead to the construction of the self. These notions form the basis of their subsequent accounts. Non-realism about the self has emerged as the dominant position within psychology and indeed philosophy today.

The current emphasis on non-realism with respect to the self may be seen as a reaction to these discredited varieties of realism outlined above. Because the everyday term 'self' is taken to refer to one of these realist selves, and seeing as it turns out that these realist selves do not exist, some theorists have concluded that this shows us that selves do not exist at all. Most go on to construct a theory of what the self 'really' is, though regard themselves to be non-realists in that (a) they want to divorce themselves from the above varieties of realism, and (b) they believe that the self that they go on to talk about is contrary to our common-sense way of using the term.

We may instead take the line that the everyday term 'self' does not refer to a thing to be found within the brain (or even an immaterial thing not to be found within the brain). This seems the most plausible line to account for the everyday term 'self' as most of us do not have the opportunity to look inside the individual's brains that we attribute self-hood to. Yet we can and do consider individuals to have selves. What we do have access to and what therefore seems the most plausible to consider with respect to self-hood is the behaviour (especially the verbal behaviour) of subjects. This is not to say that the self just is a collection of behaviours, but if behaviour is necessary for our attributions of self-hood to others then it would seem that behaviour is a necessary part to self-hood. Realism may thus be able to get a toehold because the behaviour is real.

Given that we typically attribute one self to normal subjects, the following now becomes our question: Is it the case that individual's with DID have more than one self? Or alternatively, do they not even exhibit one until they are fused? Or, is there but one all along? Even if it turns out that different alters are correlated with different transmitter levels, or different mental modules that gain control of the motor cortex or language production areas, this cannot show us that alters are selves for the same reasons that data on brain activity cannot show us the one self in the brain. The data does not

carry with it instructions as to how we should interpret its significance and so it cannot show us the neurophysiological self or selves in the brain. The behaviour and the brain behaviour of subjects are not enough to tell us how the behaviour should be interpreted. Whether there are no such things as selves, whether there can only be one to a body, or whether there can be more than one is a conceptual or interpretive issue that needs to be decided on theoretical grounds.

The phenomenon usually seems to be interpreted as not implying that the subject has more than one self (usually because of some assumption that necessitates a one-one correlation between selves and brains / bodies), or as showing the fictional side to selves because there turns out to be more than one. I will go on to argue that whether there are one or many selves is largely a matter of interpretation, though in order to maximize rationality it may be more charitable to view these subjects as having more than one self - and to see alters as selves with equivalent metaphysical status to more typical selves. Just because it is indeterminate how many selves there 'really' are does not imply that selves / alters are purely fictional; there is a realist aspect to them that needs to be emphasized in this current climate where non- realism prevails.

Dennett's metaphysics of mind

The philosopher Daniel Dennett is best known for his account of intentional states such as belief and desire. He claims in 'Real Patterns', (1998) that he attempts to achieve 'the mid-point between realism and anti-realism' regarding the metaphysical status of intentional states (or alternatively of the mind). It is worthwhile at this point to take a detour into Dennett's metaphysics of mind, as I will go on to maintain that on Dennett's account of what selves are, the same metaphysical status must apply to them as well. Dennett, (1987) claims that intentional states are visible when we take the *intentional stance* towards a system's behaviour.

The intentional stance consists in a theorist viewing the behaviour of an object in a way so as to attribute intentional states to it. When the theorist views the object in this way 'real patterns' are said to emerge which provide a reality constraint on the theorist's attributions. The truth of the attributions of specific intentional states is subject to the constraints provided by the future behaviour. Future behaviour may lend support to the attributions, or

may serve to disconfirm them.

Dennett, (1987) teases apart a realist and fictionalist component:

> While belief is a perfectly objective phenomenon (that apparently makes me a realist), it can be discerned only from the point of view of someone who adopts a certain predictive strategy, and its existence can be confirmed only by an assessment of the success of that strategy (that apparently makes me an interpretationist).

Many took Dennett's early account of intentional states to be instrumentalist, or fictionalist in flavour because of his focus on interpretive attributions over facts of the matter, and his emphasis on behaviour rather than the brain. His response was 'Real Patterns', (1998) where he emphasized the realist aspect to intentional states. He claims that

> the success of folk-psychological prediction, like the success of any prediction, depends on there being some order or pattern in the world to exploit... The pattern is discernable in agent's (observable) behaviour when we subject it to 'radical interpretation' from the intentional stance.

He maintains that the pattern is objective; it is real because it gives us 'predictive leverage we can get from no other method', (1998).

Dennett's methodology is interesting in that he attempts to characterise both intentionality and self-hood from an objective, third person perspective. While there may or may not be a distinctive conscious experience that all selves have (the very suggestion is notoriously disputed), and while qualitative states, or qualia may or may not play an important part in determining beliefs or desires, we do not have this kind of access to another individual's mind. In the case of DID one might think that the question as to how many selves there are could be conclusively settled if only we could know how many centres of consciousness there were – but the search for even one of those in terms of neuroscience has proved vastly more difficult than the simple idea suggests. It has proven difficult to the point where based on the lack of scientific evidence many theorists want to deny that selves exist, or are real at

all.

We are unable to determine the qualia of another, and so if one considers that qualia determine whether a subject really is a believer or a self there would be no way we could ever know because we could not access the required evidence – we would remain in a solipsistic position with respect to others being conscious, being believers, or being selves. Dennett may be considered to be offering us an alternative way out. He steers clear of elusive qualia for determining evidence, instead redirecting the kind of evidence required by focusing on what we all do have access to – behaviour. Our everyday notion of the self cannot refer to private subjective experiences for the same reasons it cannot refer to goings-on in the brain. We lack the required access to either and so they cannot play a role in the way in which we choose to apply and withhold the concept of selfhood, (or mind) to others.

Dennett's account of the self

So, if the self is not a real thing to be found within the brain, or a distinctive qualitative experience what is it? Dennett, (1992) maintains that a self is an abstractum, or an abstract object, like a centre of gravity.

> The physicist does an interpretation, if you like, of the chair and its behaviour, and comes up with the theoretical abstraction of a centre of gravity... [we are] faced with a similar problem of interpretation. It turns out to be theoretically perspicuous to organize the interpretation around a central abstraction: each person has a self.

Being just any old 'intentional system' is not sufficient for selfhood however, as Dennett, (1987) allows that oil refineries and thermometers qualify as 'intentional systems', or 'true believers' (which some think shows that his account of intentional states is inadequate). Even if we were persuaded it is proper and attribute intentional states to these things, we most decidedly do not want to hold that they qualify as having selves, and (arguably) we would not expect animals to have selves – although some sort of 'rudimentary self' is plausible. Dennett considers

> Our fundamental tactic of self-protection, self-control, and self-definition is not building dams or spinning webs, but telling stories– and more particularly concocting and controlling the story we tell others–and ourselves–about who we are... unlike professional human storytellers [we[do not consciously and deliberately figure out what narratives to tell and how to tell them; like spider webs, our tales are spun by us; our human consciousness, and our narrative selfhood, is their product, not their source.

Language thus enables self-representation, and we may say that this is necessary for fully-fledged self-hood. We tell others, and of course ourselves about ourselves: our likes and dislikes, our plans and expectations, our explanations and memories. We thus take the intentional stance towards our own behaviour, as others take it towards ours, and in taking it towards ourselves we are creating ourselves.

Dennett, (1991) claims that our narratives, or the stories we tell others and ourselves about who we are are spun by us, or more specifically by our brains. He emphasises that he is not implying one conscious agent talking another one into existence - that would be begging the question and would lead to an infinite regress of selves / story tellers. He claims that our stories are instead spun by, or more properly are the products of 'unconscious' or unaware neurones in our brains. He uses the example of the termite colony to show that what can seem to be highly integrated and coherent group behaviour - even to the point where some theorists have posited a 'group soul', is in actuality the behaviour of many independent organisms 'largely doing their own thing'. They independently work towards a common end goal and thus we can legitimately predict the behaviour of them viewed as a group, or single system.

This parallels the human brain in that the brain, instead of being a simple, unified thing, or even an organised collection of simpler modules is simply (on the physical level) collections of individual neurones, each one doing its own thing. Dennett's 'Multiple Drafts' model of consciousness, (1991) has consciousness presented as emerging in the brain when patterns of activation persist for long enough. It is like several lines of sub-conscious thought run concurrently and what actually makes it to consciousness is the most persistent line. There are struggles for ascendency and the track, or stream is subject to change. This makes conscious thought (and indeed unconscious thought) subservient to brain processes - conscious thought depends on which

brain process is the 'most persistent' at any given time, but this account requires no conscious editor to chose between the lines of thought. The result is consciousness that has emerged from simpler organic processes that are not themselves intelligent conscious agents.

What is relevant about this account is that Dennett maintains that self narratives are not consciously spun by pre-existing selves. With respect to alters one criticism of them is that they must require a pre-existing self to strategically adopt the role, narrate them, and thus spin them into existence. On Dennett's account a conscious self is not a requisite for a self, and thus an alternative self to develop. It thus appears to at least be conceivable that alternative selves could develop independently of one another in that more than one could arise from a single body.

Indeterminacy: It's implications for realism

Dennett is very keen to emphasize the realist side to intentional states and he shies away from classification as an instrumentalist regarding the mind. It seems that one could argue (although Dennett does not) that we should not completely abandon realism with respect to the self on the same grounds that we should not completely abandon realism with respect to intentional states. The realist part to Dennett's account of intentionality is the objective behaviour, and objective patterns in behaviour. There is also a realist aspect in that future behaviour serves (in most instances) to either support the attributions of specific mental states, or disconfirm them. I think that the realist aspect to intentional states must also apply to selves, even though Dennett seems quite happy to regard them as fictions. It seems to be consistent with Dennett's overall line to likewise claim that the realist part to the self is, once again, the objective behaviour (or objective patterns in behaviour). To maintain consistency the stance he takes on one must apply to the other also, as he has pointed out that they are both theorists' constructs designed for and maintained by their utility in the prediction and explanation of behaviour from the intentional stance.

In 'The Self as a Centre of Narrative Gravity', (1992) he writes that selves may be indeterminate and that

> this indeterminacy is a fundamental property of fictional objects which strongly distinguishes them from another sort of object

> scientists talk about: theoretical entities, or what Reichenbach called illata–inferred entities, such as atoms, molecules and neutrinos. A logician might say that the 'principle of bivalence' does not hold for fictional objects.

A worry is that here he is attempting to distinguish between theoretical, or inferred entities (such as centres of gravity), and fictional entities (selves) on the grounds of indeterminacy. Concreta too can be indeterminate though, and we do not want to regard them any the less real for it. There may be no fact of the matter as to where one mountain ends and another begins if they are side by side, however this does not show us that mountains are fictions. Dennett also allows that intentional states can be indeterminate, yet he explicitly regards them to be theoretical entities, and he is keen to emphasise the realist aspect to them. Does the principle of bivalence hold for intentional states? What about questions regarding where one mountain ends and another begins? Indeterminacy need not imply non-realism.

The indeterminacy of attributions of intentional states and self-hood

Dennett provides an account of two rival theorists, Smith and Jones, who are attempting to predict the behaviour of the same subject. He claims that they 'agree on the general shape of this individual's collection of beliefs (and desires etc.) but because of their different idealizations of the pattern, they do not agree point for point'. While the different theorists might predict different behaviours, so it would seem that one interpretation might emerge as the better (that is to say more useful one), that might not be the case in principal. He notes:

> I see that there could be two different systems of belief attribution to an individual that differed substantially in what they attributed – even in yielding substantially different predictions of the individual's future behaviour – and yet where no deeper fact of the matter could establish that one was a description of the individual's real beliefs and the other not... The choice of a pattern would indeed be up to the observer, a matter to be decided on idiosyncratic pragmatic grounds.

Let us consider selves to be attributions that are made on the basis of a subject's behaviour (especially their verbal behaviour where they attribute a self to themselves). Because of the indeterminacy of attributions it follows that different theorists could have differing theories as to how many selves there are with respect to one body. One theorist could maintain that there was one self, and another that there was many. There may be no further fact of the matter that could decide between these two interpretations.

In 'Speaking for OurSelves', (1998) he writes:

> suppose, at different times, different subsystems within the brain produce 'clusters' of speech that simply cannot easily be interpreted as the output of a single self. Then – as a Bible scholar may discover when working on the authorship of what is putatively a single-authored text – it may turn out that the clusters make best sense when attributed to different selves.

Interpretation seems to be all-important, and the factor driving our interpretation is the desire to make the 'best sense' out of the behaviour (including verbal behaviour) of subjects. It seems that a major factor in deciding whether one or many selves are present is a matter of interpretation. Suppose a system's behaviour taken as a whole appears only minimally consistent - erratic, changeable, inconsistent, even unpredictable – a lot of noise must be budgeted for. But there may be more useful patterns there for the innovative observer to discern. Perhaps there may be discernable pattern in the noise, and the subject may be predictably unpredictable so to speak. A system, for example, may be better viewed as not just one system, but as several different systems. The unpredictability may lie more in uncertainty as to which system may be in control at a given time (although patterns may be discernable here too). The multiple systems view could have predictive leverage over the single systems version in that it budgets for considerably more pattern and less noise, that is to say it is more useful, not more real.

One thing that the intentional stance requires is a 'rational decider'. There needs to be a rational decider to co-ordinate the beliefs and desires and to act from there. With respect to a self we seem to be making the attribution that the rational decider is largely the same, and is largely consistent over time. That seems to amount to the content of the beliefs and goals being largely the same, or at least evolving in comprehensible ways. It also seems to involve the notion that the beliefs are largely non-contradictory. In terms

of the decision to view the subject as having one or many selves we may say that it is more charitable to maximize rationality, and rationality can be maximized by viewing the subjects as having separate centres of narrative gravity that are internally largely consistent and evolve comprehensibly but are mutually incompatible. The alternative is that rationality is not optimal and we have to budget for more noise. In the descriptive sense it may be practical to consider these subjects to have more than one self.

The attribution of one or many centres of narrative gravity may be indeterminate – there may be no further fact of the matter that could decide between the competing hypotheses of multiple centres of narrative gravity and an ardent refusal to describe the subject's behaviour in this manner. Even if one hypothesis explained and predicted more behaviour and less noise than the other both would 'get rich' as Dennett puts it, simply by one budgeting for more noise. The noisier pattern would make allowances for 'unpredictable' behaviour, whereas the other would see pattern in the noise. Even though one may emerge as having more pattern Dennett explicitly states that it does not follow that it is more real. So with respect to intentional states if there are two competing belief-desire ascriptions that predict behaviour then they must both be true (in that they are both derived from the theorist's valid interpretation of real pattern), and so as to the question as to whether the subject really has one or more selves there too there may be no fact of the matter.

There seems to be no reason why a single brain or body cannot give rise to more than one self in principal unless one adheres to some sort of biological notion of the self that restricts selves and bodies (or brains) to a one-one correlation necessarily. Brown, (2001) does this with his account of DID by holding an Animalist account of the self where there is, at maximum, one self per human animal. His motivation for this is to avoid 'moral problems' (unspecified) that he believes may result in taking alters to be selves, and he takes the 'reification' of alters to be 'metaphysically extravagant'[2]. Aside from these two motivations he does not provide an argument for animalism and a one-one correlation; rather this presupposition motivates his account. Biology has traditionally had less to do with the notion of a self, however,

[2]Although it does not seem to be metaphysically extravagant for the claim is that there are simply more tokens of selves than we previously realised – it is not another kind of 'stuff' different in kind that is being postulated. The criticism of metaphysical extravagance may be better applied to those who view alters to be different in kind from selves.

and more to do with the related notion of a person, which seems to also be the notion more tied up with moral rights and responsibilities. Those who are inclined to take some sort of biological line might find commissurotomy a more likely phenomenon with respect to challenging ones intuitions that selves and brains must enter into a strict one-one correlation.

If it turns out that different mental modules are responsible for the different patterns in behaviour that constitute the alters then this would seem to be a similar sort of case though. Or perhaps one specifically wants to attempt to draw a bodily criterion into the notion of self (as Brown does). The problem with this is that the body is obviously not sufficient as a criterion when one considers the possibility (in principal) of body swaps and the inclination most of us have to say that the self 'follows' the brain, and is no different for a different body. The bodily criterion here thus seems to be irrelevant – many claim that it is not necessary - with respect to the self. This is not, of course, support for the claim that disembodied selves are possible. Presumably it is necessary for there to be some sort of physical basis that gives rise to the self; it just does not seem to be terribly relevant which body it is so long as it can function in the world in similar ways, thus giving rise to similar patterns in behaviour.

So the three alternative positions that we began with were that (a) selves are fictions (as explicitly endorsed by Dennett); (b) alters are different in kind to selves but may be blended or fused together to result in a self; and; (c) alters are not selves, rather they are role-plays. While the latter claim seems to suggest that there is more to the self than an adopted role play which seems to implicitly require a degree of realism with respect to the self; non-realism about selves has emerged as the dominant line. Here I have attempted to show that it is a plausible alternative to disagree with all of the above and maintain that alters constitute selves and that there is a realist aspect to selves. Realism of the Cartesian variety may well be untenable, but behaviour must be the basis of our attributions of selfhood to others and so may be seen to provide an aspect of realism that constrains our attributions.

I think that there is a decision to be made as to whether we interpret the behaviour of these subjects as indicating that they have one self or many. We are understandably biased towards positing one centre of narrative gravity wherever possible, as one seems sufficient for explaining and predicting the behaviour of the majority of the population. In some cases, however, there may be a predictive and explanatory advantage to positing more than one self to a subject. This interpretation may be seen as more charitable with

respect to maximising the subjects' rationality, and explaining and predicting more of their behaviour. On Dennett's account this does not imply that this interpretation is 'more real', but I think that if more of the behaviour can be explained and predicted then this would seem to be a good reason for considering the multiple systems version of the intentional stance a better theory of these subjects' behaviour than the alternative.

References

Adler, Robert (1999). 'Crowded Minds', *New Scientist*.

Apter, Andrew (1991). 'The Problem of Who: Multiple Personality, Personal Identity, and the Double Brain', in *Philosophical psychology*, Vol. 4, Issue 2.

Brown, Mark (2001). 'Multiple Personality and Personal Identity' in *Philosophical Psychology*, Vol.14, No.4, 2001.

Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.

Clark, Stephen (1990). 'How Many Selves Make Me?', in *Royal Institute of Philosophy Conference on Human Beings*.

Dennett, Daniel (1998). *Brainchildren: Essays on Designing Minds*, Penguin Books.

Dennett, Daniel (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*, Harvester Press Limited.

Dennett, Daniel (1991). *Consciousness Explained*, Little, Brown & Company.

Dennett, Daniel (1987). *The Intentional Stance*, Massachusetts Institute of Technology

Dennett, Daniel (1996). *Kinds of Minds: Towards an Understanding of Consciousness*, Weidenfeld & Nicolson.

Dennett, Daniel (1989). *The Origins of Self*, Cogito, 3, 163-73.

Dennett, Daniel (1992). 'The Self as a Centre of Narrative Gravity' in F. Kessel, P. Cole and D. Johnson, (eds.), *Self and Consciousness: Multiple Perspectives*, Hillsdale.

Feinberg, Todd (2001). *Altered Egos: How the Brain Creates the Self*, Oxford University Press.

Gillett, Grant (1997). 'A Discursive Account of Multiple Personality Disorder', *Philosophy, Psychiatry & Psychology* 4(3).

Gleaves, David H (1996.) 'The Sociocognitive Model of Dissociative Identity Disorder: A Reexamination of the Evidence', *Psychological Bulletin*, 120(1).

Glover, Jonathan (1998). *I: The Philosophy and Psychology of Personal Identity*, the Penguin Group.

Graham, George (2002). 'Recent Work in Philosophical Psychopathology', *American Philosophical Quarterly*, Vol. 39, No. 2.

Hacking, Ian (1991). 'Two Souls in One Body', *Critical Inquiry*, 17.

Kluft, R (1988). 'The Phenomenology and Treatment of Extremely Complex Multiple Personality Disorder', *Dissociation*, 1.

Kolak, Daniel (1993). 'Finding Our Selves: Identification, Identity and Multiple Personality', *Philosophical Psychology* Vol.16 No.4.

Lilienfeld, Scott O; Lynn, Stephen Jay; Kirsch, Irving; Chaves, John F.; Sarbin, Theodore R.; Ganaway, George K.; Powell, Russell, A., (1999). 'Dissociative Identity Disorder and the Sociocognitive Model: Recalling the Lessons of the Past', *Psychological Bulletin*, 125(5).

McHugh, Paul, and Putnam, Frank (1995). 'Resolved: Multiple Personality Disorder Is an Individually and Socially Created Artefact (rebuttal)' *Journal of the American Academy of Child and Adolescent Psychiatry* 34 (7).

Merckelbach, Harald; Devilly, Grant J.; Rassin, Eric, (2002) 'Alters in Dissociative Identity Disorder Metaphors or Genuine Entities?' *Clinical Psychology Review*, 22.

Pitt, David (2001). 'Alter Egos and Their Names', *The Journal of Philosophy.*

Pope, Harrison G; Oliva, Paul S.; Hudson, James I.; Bodkin, Alexander J.; Gruber, Amanda J., (1999). 'Attitudes Towards DSM-IV Dissociative Disorders Diagnoses Among Board-Certified American Psychiatrists'

American Journal of Psychiatry, 156(2).

Putnam, F.W, (1989). *Diagnosis and Treatment of Multiple Personality Disorder*, Guilford Press.

Saks, Elyn (1994). 'Integrating Multiple Personalities, murder, and the Status of Alters as Persons', in *Public Affairs Quarterly*, vol. 8, No. 2.

Spanos, Nicholas P.; Weekes, John R.; Bertrand, Lorne D. (1985). 'Multiple Personality: A Social Psychological Perspective', *Journal of Abnormal Psychology*, 94(3).

Wilkes, Kathleen (1988). *Real People: Personal Identity Without Thought Experiments* Oxford University Press.

Chapter 2

System indeterminacy and interpretive constraints

Abstract. *The essential feature of Dissociative Identity Disorder is the presence of two or more alternative identities associated with a single body. One of the most controversial issues that arises from this is how we are best to conceive of alternative identities. Within contemporary psychology and psychiatry two rival models have emerged as dominant theoretical positions. While they are typically considered to be mutually exclusive, I shall attempt to recast the problem of alters in a way that is fairly neutral between them. This will involve a new application of a philosophical model of mind known as intentional systems theory. Intentional systems theory acknowledges that there may be a degree of indeterminacy with respect to what intentional state a system is in. If we conceive of alters or selves as a certain kind of complex intentional system, then it seems plausible that there may arise a similar phenomenon of system indeterminacy. The problem of alters may thus be re-conceptualised as the problem as whether to adopt a single or multiple systems interpretation of these subjects behaviour.*

Introduction

Dissociative Identity Disorder (DID) has been the subject of much controversy in psychological, psychiatric, and philosophical literature. Within mainstream psychology and psychiatry two rival accounts have emerged as dominant theoretical positions. Each theory offers an alternative account of the following three aspects of the disorder:

1. Aetiology

2. A conceptualisation of alters

3. A proposed course of treatment

Each theory is typically considered a package deal in that it offers a position that embraces each of these three aspects; and each aspect is considered to flow quite naturally into a position on the next. One may question the extent to which a stance on one aspect logically entails the rest of the theory. Theorists, however, have seemed to take a stand on the accounts considered as package deals. It will be useful to begin with an enumeration of these received views before I go on to offer an alternative conceptualisation of 2.; how we should conceive of the alters that are the distinctive feature of this diagnosis.

The post-traumatic model

The post-traumatic account originated from the work of theorists / clinicians in the 1980's. Braun, Kluft, Putnam, Coons, and Bliss are cited by Ross, (1989 p. 50) as important figures in re-establishing clinical interest in DID (formerly Multiple Personality Disorder) as a legitimate phenomenon. These theorists have gone on to write seminal work on (1), (2), and (3) in the form of papers and treatment manuals. While there are points of difference in emphasis and detail between supporters, there seems to be a general consensus on an overall view that has come to be known as the post-traumatic account. Gleaves, (1996 pp. 42-59) has recently written in defence of this view in response to Spanos offering the socio-cognitive model with the intention that it be accepted as a replacement conceptualisation (Spanos, 1994 p. 29). I will focus largely on Gleaves account as he clearly opposes the alternative model, and he seems to be fairly representative of the post-traumatic line.

According to the post-traumatic model alters originate in childhood when individuals with a diathesis for dissociation encounter severe, repeated trauma (Gleaves, 1996 p. 2). The child dissociates aspects of their experience from conscious awareness as a protective coping strategy. If the experiences were accessible to consciousness, or impinged on the child's consciousness the functioning of the child would be severely impaired. For example, an abused child may need to dissociate from abuse in order to behave trustingly to an abuser at other times in order to ensure that primary needs such as those for food and shelter are met (Gleaves, 1996 p. 2). Because the strategy is successful (in that it enables the child to cope) the dissociation is reinforced. Because

of the extreme and repetitive nature of the abuse the child comes to dissociate more often, and in these times their behaviour is governed by these alternative states.

It is a distinctive and defining feature of the disorder that these states develop an internal consistency and coherence of their own (DSM IV-TR, 2000 p. 529). Alters are thought to function to 'contain memories' of different kinds of experiences, and to act in ways believed to be required for the benefit of the child. One alter may be a passive and helpless recipient of abuse with access to distressing memories. Another may take responsibility for deriving pleasure from the abuse so as to behave in a manner that pleases abusers. Another may come out to 'fight back' by taking active steps to use force to protect the child's body. Because different alters have different protective functions they have access to different memories, emotions, and goals; and thus they behave in distinctively different ways.

On this account alters are conceptualised as dissociated aspects, fragments, or parts of the greater self that is their summation. Dissociation is thought to be a highly creative and adaptive strategy that enables a child to deal with child-hood abuse that they otherwise cannot escape. It is thought to become maladaptive when it continues once the abuse has stopped, and the behaviour of the alters causes distress to the 'main personality', or alter that presents for treatment. The goal of treatment is the integration or fusion of these dissociated aspects into one largely integrated and consistent self of the sort exhibited by individuals without the disorder.

The socio-cognitive model

Although it would be fair to say that the majority of clinicians are sceptical as to the legitimacy of the disorder (Pope *et al.,* 1999 pp. 321-323) it was not until fairly recently that an alternative to the post-traumatic account has been offered
[1]

. The number of cases diagnosed each year increases at exponential rates for a disorder that was once considered exceptionally rare. Lilenfeld *et al.,*

[1]While many used to voice their scepticism in the form of disbelief or outright denial of the phenomenon such a position is becoming increasingly hard to sustain. It also seems to have been long considered that subjects were play-acting, or making up stories but a sustained alternative account has not been forthcoming until the work of Spanos, (1994).

(1999 p. 508) report that while there were less than 80 cases reported worldwide prior to 1970, the figures at the close of the twentieth century, though difficult to estimate, appeared to be in the tens of thousands. While supporters maintain that these figures are more accurately reflective of true prevalence rates such a dramatic increase has led to increasing degrees of controversy, scepticism, and demand for an alternative explanation from other quarters.

While the post-traumatic account is psychodynamic in origin, Spanos, (1994) offers an alternative conceptualisation that is more consonant with behaviourist theory and practice. He emphasises the role of reinforcement contingencies in the creation, maintenance, and ultimate dissolution of the disorder (Spanos, 1994 pp.17-20). DID is conceptualised as a modern form, or variant of what he dubs 'multiple identity enactment'. Alters (as a phenomenon) are considered to function in a similar way to possessing sprits or demons reported in past eras. These phenomena are thought to be culture specific; they occur only where people 'believe in them' and thus their expressions are considered legitimate by the enacting subject and others.

Spanos, (1994 p. 20) considers that it may be reinforcing for subjects to strategically enact multiple identities, especially when they are allowed to avoid the consequences of their behaviour by interpreting it as the actions of other agents. He notes that Protestants who were treated with prayer and fasting for behaving possessed reported fewer cases of possession than Catholics who were treated with bed rest and elaborate exorcism rites (Spanos, 1994 p. 15). While he maintains that there is nothing pathological (or disease-like) about multiple identity enactments *per se,* he also considers that those who present with DID for psychological or psychiatric assistance in modern times do exhibit a greater pathology (Spanos, 1994 p. 28).

The Three Faces of Eve and *Sibyl* were bestseller biographies depicting subjects with DID. They were made into feature films which served to bring the disorder to the attention of the general public. The rise in the number of cases reported occurred shortly after the release of these films. The psychiatrists that treated 'Eve' (Thigpen and Cleckly, 1984 p. 64) reported being inundated with letters and phone calls from individuals who presented with different handwriting samples and different voices that claimed to be separate selves. While they concluded that they were (pathological) hoaxes they did not seem to investigate these claims in any great depth. Spanos considers, though, that this shows the impact that media attention has on subjects with certain pathologies. The disorder has been presented in such

a fashion that disturbed individuals are given an elaborate and glamorous explanation for their difficulties. The reinforcement provided by the media and greater society is thus the first factor that Spanos considers relevant to the dramatic increase in the number of subjects presenting with the disorder (Spanos, 1994 p. 20).

The second factor is considered to be the reinforcement contingencies provided by the clinicians that regularly diagnose and treat the disorder. Spanos considers that clinicians (perhaps unwittingly) provide cues by asking leading questions that educate and enable subjects to convincingly enact the multiple role. Some clinicians find the disorder intriguing and fascinating, and subjects with the disorder are thus given a great deal more attention and sympathy then they would otherwise obtain. For a subject with a history of severe abuse and / or a long history of worn out clinicians enacting multiple identities may be very reinforcing indeed (Spanos, 1994 p. 21).

He thus maintains that alters are artefacts, creations or roles that are produced and sustained in response to social reinforcement and the reinforcement provided by traditional forms of treatment. He proposes an alternative course of treatment, which involves altering reinforcement contingencies so as to extinguish the behaviours that constitute the disorder (Spanos, 1994 p. 20).

Intentional systems theory

Now that we have seen something of the mainstream views it will be useful for me to provide an outline of a philosophical doctrine in philosophy of mind known as intentional systems theory. An account that I will offer of (2) may be considered something of an extension of intentional systems theory and I will use it to recast the problem of (2) in a way that is fairly abstract thus neutral between the above accounts. While there would seem to be much empirical work to be done with respect to (1) and (3), the issue of how we should conceive of alters is more a conceptual or theoretical issue than an empirical one. While there is a tendency for theorists within psychology and psychiatry to consider (2) to be an issue of construct validity to be determined by the remainder of (1), and (3); construct validity is not the subject of this paper. I am interested in providing an account of the phenomenon of subjects presenting / living their lives with multiple identities. How we are best to conceive of their behaviour and the presence of alters would seem

to me to be an issue that can be teased out from how they came to be that way and how clinicians can successfully treat them. This approach may also cast a new light on (1) and (3) though it is logically separable from them.

Intentional systems theory is often taken to be an explicit rendering, or extension of what is known as 'folk-psychology'. Despite the behaviourists' success in the laboratory it would seem that we cannot function effectively in society without making use of such attributions as 'believes', 'desires', 'hopes', 'wants', 'fears' etc. The fact is we attribute these mental (or intentional) states to ourselves and others; and we use these attributed states to predict, explain, and thus make sense of our own and others behaviour. Whether these notions can be reductively explained in terms of neurological states, or whether they are merely fictions (so strictly speaking do not exist) is controversial. I will have more to say about the issues of reduction and fictionalism in subsequent sections.

Intentional systems theory notes that sometimes we regard an object as an intentional system. An intentional system is an object (or system) with mental states that interact so as to produce behaviour. When we want to predict, explain, or make sense of the behaviour of a system we can adopt the intentional stance towards the system, which consists in the following:

1. The attribution of particular beliefs to the system.
2. The attribution of specific desires to the system.
3. The attribution of practical rationality to the system.

The notion is that we attribute beliefs on the basis that a system 'believes what it ought to believe, given the situation they are in' (Braddon-Mitchell & Jackson, 1996 p. 146). We thus consider that an intentional system has beliefs regarding its environment. For example, we would consider that an intentional system sitting on a chair would believe that it was sitting on a chair (unusual circumstances aside). We attribute desires on much the same grounds. Living intentional systems are attributed desires for biological needs such as food and shelter at the appropriate times, and so forth. We also attribute all sorts of other beliefs and desires to intentional systems that are hard to specify but come quite naturally to us in our daily lives when we are employing folk psychology. Practical rationality is the ability to 'act to satisfy ones desires were ones beliefs true', or the ability to coordinate

beliefs and desires in such a way as to produce the relevant action (Braddon-Mitchell & Jackson, 1996 p. 145). For example, one may have the ability to coordinate ones belief that there is food in the fridge with ones desire for food in order to produce the relevant action of going to the fridge in order to get some food[2].

Dennett, (1987, 1988) and Davidson, (1980) consider that there are patterns that emerge when one adopts the intentional stance towards the behaviour of a system. Although this does not seem to be explicit in the literature it seems that by 'pattern' intentional systems theorists are primarily concerned with patterns discernable from something approaching a snapshot view as opposed to over significant periods of time[3]. As an example of what is meant by a snapshot view, we could briefly view a scene where someone is walking and there is a hole in the ground in front of them. By adopting the intentional stance we could attribute that the system believes that there is a hole in front of them and that if they continue walking they will fall into it. They desire not to fall into it and they are rational enough to realise (and act from the understanding that) they thus should walk around the hole. The 'patterns' would seem to be kinds of events or objects that are multiply realised on the physical level and thus are irreducible to it[4]. Intentional systems theorists are not committed to the view that we go around explicitly considering others to be intentional systems by running through these little hypotheses sub-vocally all the time. But they do consider that if we are asked to provide an explanation or make a prediction regarding behaviour

[2]Beliefs and desires may be conscious or unconscious. Someone who says that they are hungry but do not eat despite opportunity leaves themself open to the charge that they are not really hungry; thus there is no privileged first person access to intentional states. Intentional systems theory focuses on the role of behavior with respect to intentional states (and so offers an account of how we can learn to attribute them which is a major problem for introspectionist accounts). It is also thought that we can have many beliefs and desires that interact with one another. The strongest beliefs and desires are the ones that determine the relevant action. Thus, it is plausible that a hungry person may not eat because their belief that the food is poisoned and their desire not to eat poisoned food is stronger than their desire to eat.

[3]While there may not be such a clear distinction between a 'snapshot view' and 'patterns over time' in the literature I make this distinction so as to later go on to show how intentional systems theory may be used to not only provide an account of attributions of specific beliefs and desires, but also attributions of self-hood. Typically, the examples in the literature have to do with tiger or hole avoidance and food seeking and I want to distinguish between these relatively simple intentional states and the more complex character traits that emerge over time and prompt our attributions of self-hood.

[4]I will consider the issue of reduction further in a subsequent section.

then belief-desire explanations are cited as to what makes the behaviour, or our predictions of it rational[5]. (Braddon-Mitchell & Jackson, 1996 pp. 144-158).

Controversial issues within intentional systems theory include specifying in greater detail how we form hypotheses regarding what an intentional system believes and desires. The theory requires a specification of a criterion by which we accept or reject candidate hypotheses for belief and desire attribution. While intentional systems theorists may consider that this is the full story to be told about intentional states, other theorists consider that it needs to be supplemented with an account of corresponding brain states. It is also a matter of controversy as to whether it is plausible or legitimate to attribute optimal rationality to intentional systems. We seem prone to a variety of cognitive biases and heuristics that show us that the rationality exhibited by an intentional system is limited. While it is indeed an interesting research program to attempt to specify this in enough detail so that a computer could be programmed to formulate acceptable predictions and explanations from the intentional stance, I am happy to run with it at a fairly superficial level. While much clarification needs to be done, intentional systems theory seems to offer a plausible picture-view of how we go about attributing intentional states both to ourselves and others in our daily lives.

An intentional system as a self

Intentional systems theory is primarily a theory as to when we are entitled to say that a system is a true believer. Our attributions of specific intentional states are held to be true in virtue of their predictive success. Dennett, (1998) considers that the intentional stance gives us 'predictive leverage that we can get by no other method'. While this is controversial, we may consider that the prediction that Sally will go to a shop because she wants to buy a puppy cannot be translated into a prediction from the level of physics. Firstly, we may consider that there is nothing on the physical level that corresponds to

[5]'Rationality' is being used in the fairly technical sense detailed above. We can use intentional systems theory to predict rational behavior that results from irrational beliefs. For example, we could predict that a subject will avoid people because she (irrationally) believes that they are trying to kill her and she desires not to die. We can thus predict, explain, and understand behavior even when irrational beliefs are involved though it might be harder to hit upon the appropriate beliefs to attribute in these circumstances.

''a shop' either in the subject's brain or in the external world[6]. We may thus consider that *kinds* of behaviours that are crucial to intentional explanations (e.g., going to a shop) are multiply realisable both in the brain and in the external world, and thus they are irreducible to the physical level. The same could be said for the notion of a 'puppy' as an object[7], and (arguably) for the notion of 'belief' itself. While some (notably the Churchlands) consider that intentional states are irreducible and thus illegitimate and should be abolished, the fact is that the intentional stance is legitimated and sustained by its utility.

If we were to opt out of intentional psychology, we would not be able to function in our everyday lives, and it would be us that would become extinct and not the theory of intentional psychology. We cannot refrain from interpreting the behaviour of others from the intentional stance, and we cannot refrain from interpreting our own behaviour from the intentional stance. Inability to use the intentional stance adequately would appear to be a feature of pathology, such as when someone is unable to attribute appropriate emotional states to themselves (or to label them), or is unable to form adaptive beliefs regarding themselves or others that serve to facilitate their needs being met.

We may thus consider that the intentional stance is predictive in virtue of capturing real patterns or kinds of behaviour that are not visible from a lower level (physical) stance. It is in virtue of this predictive success that we

[6]While shops-in-the-world must indeed be physically instantiated, precisely what constitutes a shop would seem to be inexorably tied up with social and legal practices that are emergent to the intentional level. While there may be a class or set of shops on the physical level such a set would seem to be a disjunctive set of shop A, shop B etc. The kind 'shop' is thus multiply realized on, and thus irreducible to the physical level. The concept 'shop' would also seem to be multiply realized both in different individual's brains, and within the brain of a single individual. One lesson that might be taken from Lashley's infamous 'search for the engram' is that memories (and perhaps even the concepts involved in them) do not reside in any particular region of the cerebral cortex, rather they only become inaccessible when enough of the cortex's overall area is destroyed. Plasticity of function within the same individual also shows that concepts can indeed be multiply realized on the physical level.

[7]Most seem to consider that biological kinds are not natural kinds because there is no one property (on the physical level) running through many instances. For example, if we consider DNA to be relevant with respect to determining biological kind membership then it is irrelevant which physical bits of matter instantiate this, it is only relevant that they are in fact instantiated. Biological kinds thus seem to be emergent kinds, and here I am attempting to argue that intentional kinds are emergent in just the same way.

are entitled to use the intentional stance to explain and describe behaviour as well. While Dennett, (1987) considers that a variety of objects behaviour can be predicted by the intentional stance e.g., oil refineries and thermometers it would seem that adopting the intentional stance towards these objects does not buy us 'predictive leverage we can get from no other method'. We could equally well predict their behaviour from the design stance (where they behave as they are designed to behave other things being equal) and thus I consider that viewing these objects as intentional systems is to attribute a greater mental capacity than is needed to explain the phenomenon. These systems thus do not count as 'true believers'.

While intentional systems theory focuses on attributions of particular mental states, I think that it can be extended so as to provide a similarly rough picture-view account of our attributions of self-hood. While intentional stance theorists typically consider beliefs and desires that would be attributed fairly uniformly to any intentional system (e.g., that a system believes relevant things about the environment and has fairly standard desires for biological needs etc.,) sometimes the attributions that interest us the most are those that are fairly idiosyncratic to particular people or personalities. We can consider that when different people are in the same circumstances they often behave in different ways and when we know something of the particular people involved, we can often predict how they will behave compared to one another.

To consider the notion of a self or personality we need to look not only at the patterns that emerge from a snap-shot view when we view the subject as an intentional system; we also need to consider patterns that emerge as frequencies of these emergent kinds of behaviours when we view them over time. So, the picture we have is as follows:

1. When we consider an object as an intentional system 'real patterns' emerge that legitimate our attributions of specific beliefs and specific desires so as to predict and explain the systems behaviour.

2. When we view the patterns in the behaviour of an intentional system over time further patterns emerge in the frequency of kinds of behaviours that an intentional system exhibits. These patterns have to do with attributions of preferences and consistent character, or personality traits etc, and they serve to legitimate our attributions of selfhood.

For example, some intentional systems frequently respond to certain kinds of events by feeling stressed. Some intentional systems frequently deal with stress by exhibiting avoidance behaviours, and others work pro-actively to alleviate the stress. We often use these patterns (that emerge as frequencies) to predict how that system will behave in the future. We attribute personality traits such as 'avoidant' or 'pro- active' on the basis of many specific attributions that are made from the intentional stance. It thus seems reasonable to consider that the concept that we have of a unique individual, personality, character, or self is a more general attribution or inference that is built out of the specific intentional states that we attribute. It is a result of considering frequencies in our attributions, or the patterns that emerge in the behaviours that prompt our attributions when we consider either the behaviour of the system, or the frequency of our attributions to it over time[8].

While intentional systems theory considers that beliefs and desires ought to 'evolve in right and proper ways' it seems that by this they are primarily concerned with beliefs evolving in light of changes in the immediate environment and desires growing until they are satisfied (Braddon-Mitchell & Jackson, 1996 p. 148). When we consider the notion of an intentional system as a self, we may consider not only immediate or fairly immediate desires for biological needs, but also further reaching goals or plans, memories and preferences. We expect that an intentional system, or a self is largely consistent or continuous through time as the beliefs and desires evolve in right and proper ways, and do not alter abruptly for seemingly no good reason. Sometimes people do experience neurological damage which results in behavioural changes that has others conclude that they are not the same 'person' any more. We may consider that here the self has altered so abruptly, or has degenerated to the point that it is hard to see how the beliefs and desires could have rationally evolved from the earlier intentional system[9].

[8]While I have not emphasized the role of verbal behaviour in this paper it would seem to play an important role in our attributions of both specific intentional states and in our attributions of self-hood. Often a fairly good predictor of what intentional state a subject is in, or what they are going to do next is simply to ask them. While some consider that there is a first-person privileged access that is associated with a phenomenal feel one might also consider that we are typically better at predicting our own behaviour over other systems because we observe our own behaviour all the time whereas we have only intermittent access to other systems. Verbal reports are also frequently not highly predictive of behaviour. We might consider that some people don't 'know their own mind', or indeed know themselves very well at all.

[9]Of course, the alterations in their behaviour can be *reductively* explained in terms of neurological damage but this explanation is an explanation as to why the beliefs, desires,

The phenomenon of dissociative identity disorder

While the distinctive and defining feature of dissociative identity disorder is the presence of alters it is acknowledged by sceptics and supporters both that only 20% of DID patients exhibit clear-cut indications of this condition at the beginning of treatment. The remaining 80% exhibit only specific 'windows of diagnosability', namely transient periods during which the classic features of DID are evident (Kluft, 1991). Although there is disagreement concerning the exact percentages, 'virtually all authors in this literature have concurred that a large proportion – perhaps a majority – of DID patients in their samples exhibit few or no unambiguous signs of this condition prior to therapy' (Kluft, 1991).

When we consider the typical presentation of potential DID subjects, we are left with a more general picture of overall muddle. Often subjects with 'transient windows of diagnosability' may be considered to present as something of an unintegrated, fairly incoherent intentional system. Over time the intentional system varies radically in its beliefs and desires. It may profess one thing and act in accordance with it, but at other times it may disavow actions, memories, or past utterances. The behaviour of such a system would be lacking in integration and coherence, they would exhibit, contradictory beliefs and conflicting goals. The natural interpretation would seem to be that a system such as this is impulsive or unpredictable, contradictory, and perhaps with diminished rational capacity.

The amnesia requirement that was dropped from the DSM III was restored, partly as an attempt to curb the dramatic increase in prevalence rates. Subjects often meet this requirement by claiming that they find new possessions that they do not know how they acquired. They find their belongings moved around to a degree that cannot be explained by ordinary forgetfulness. They may claim that they are approached by people who claim to know them well but they cannot recall meeting them. They also claim that they have amnesiatic episodes where they cannot recall their behaviour. This seems to further illustrate that these subjects present as fairly

---

and behaviours have *not* evolved in comprehensible, rational ways.

disorganised intentional systems.

Some theorists have considered DID to be a variant of Borderline Personality Disorder (BPD) and as many as 70-80% of subjects with DID also meet the criteria for a diagnosis of BPD (Ross, 1996). If we ignore the issue of alters and consider the behavioural presentation of subjects with DID there is a large overlap of symptoms[10]. While supporters consider that BPD symptoms are best explained by the presence of alters; sceptics maintain that the presence of alters is best explained in terms of BPD symptoms with the addition of alters as a treatment induced artifact. The emotional 'instability' and impulsivity that could be interpreted as variability between alters is covered by criteria (2), (4), (5), and (8). (3), (7), and (9) relate to identity disturbance, dissociative symptoms, and subjects that report being afraid of the actions or voices of persecutory alters may be considered delusional or paranoid.

Alters also may be considered 'responsible' for the self damaging behaviours reported by criteria (4), (5), and (8). Subjects with DID are typically considered to have at least one hostile or persecutory alter who engages in damaging behaviours to the subject's body and / or other people. While it is considered that not all DID subjects meet the criteria for BPD, some clinicians consider that DID takes precedence and so would not list BPD as an additional diagnosis (Ross, 1989 p. 143). Not all BPD subjects present with alters, and so some theorists consider that DID is a form of, or severe variation of BPD. Ross (1996, p. 149) states that

> Looking at MPD patients from a borderline vantage point, they hold that MPD is an epiphenomenon of borderline personality. Basically, the argument is that MPD specialists create an MPD artefact in borderlines. Such clinicians rarely diagnose MPD because they deal with the "real" disorder, borderline personality.

Because the *Diagnostic and Statistical Manual of Mental Disorders* aspires to establish psychiatry with the same empirical grounding and treatment success as enjoyed by the rest of medicine, disorders are considered disease entities that are to be differentiated by unique aetiology (including age of onset), behavioural presentation (offered as a set of symptoms or syndrome), and effective course of treatment (course of illness and predicted

[10]See appendix for criteria.

treatment outcomes). Psychiatric disorders are thus conceptualised and presented as discrete, distinct, and all or none in that one either meets the criteria for the disorder or one does not. While this discrete disease entity conceptualisation works well for some illnesses (e.g., Alzheimer's), there is controversy as to whether the disease conceptualisation is appropriate for all of the listed pathologies (Davidson & Neale, 2001 pp. 69-71).

Dissociative disorders, post-traumatic stress, somatoform disorders, histrionic and borderline personality disorder, substance abuse, eating disorders, anxiety, and depression (that does not respond as effectively as clinical depression when treated) seem to co-occur in a number of subjects. The DSM is structured in such a way that there seems to be little natural relation between these disorders, whereas some clinicians recognise that they frequently occur together and they maintain that future structuring of the DSM should reflect this. These disorders also may be better conceptualised as lying along a continuum where symptoms are ranked for severity from normal to abnormal to severe. This would reflect the notion that many of the symptoms do appear in the normal population and it is the degree to which the behaviour is present that is of concern. This is currently debated and may result in a restructuring of the DSM in subsequent editions (Davidson & Neale, 2001 pp. 69-71).

Because there is overlap in content (with respect to symptoms) for diagnosing this cluster of disorders many individuals meet the criteria for more than one of these and some meet the criteria for different disorders at different times. While some individuals present fairly clearly with one or two (or three) of the above, others seem to be diagnosed with a variety of these over a 7-10 year period before a diagnosis of DID is made (Ross, 1993; Gleaves, 1996). Medication assists with symptoms in a limited way but does not seem to control the disorder the way it does with the model diseases such as schizophrenia, bi-polar, and true clinical depression. These subjects are the ones that seem to show that diagnosis can often be a somewhat arbitrary matter that is indeed, to a very large degree, a matter of interpretation.

Multiple systems theory

After considering the above three theories I am now in a position to outline an alternative position on (2), which I will call multiple systems theory. According to multiple systems theory (or a multiple systems version of the

intentional stance) it may be legitimate in some cases to interpret or view the behaviour of one subject as being best predicted and thus explained by multiple intentional systems being associated with a single body.

Different alters (intentional systems) are observed to behave in distinctively different ways. They would thus seem to have different sets of beliefs and desires that function to produce the behaviour of the body when that system is in control. The behaviour, and the beliefs and desires that are attributed in order to predict and explain the behaviour are largely incompatible between systems - which is why there is an advantage to postulating more than one such system. Internally the systems (as sets of beliefs and desires) are largely non-contradictory, and evolve in comprehensible ways. This is not a feature of episodes of psychosis, or psychotic voices. The sets of beliefs and desires thus constitute distinct intentional systems, or selves. So what does the multiple systems view buy us? I maintain that in some cases the multiple systems view buys us predictive and explanatory leverage that we cannot obtain from the single system view. In the 20% of subjects whose presentation is blatant and in the majority of diagnosed cases, it would appear that multiple systems theory has predictive leverage over the single systems view.

Where the single system view had to allow for unpredictable and inconsistent, irrational behaviour the multiple systems view buys us an account with greater predictive and explanatory power. I maintain that given the predictive advantage of the multiple systems view we may consider that in virtue of this it gives us a greater explanatory advantage as well. This being so the multiple systems view is the most descriptively adequate account that we have of these subjects behaviour. It would also seem to be the most charitable view with respect to making the best sense that we can of these subject's behaviour, as we no longer have to attribute defects in impulse control, rationality, consistency, or coherence.

Reductionism, fictionalism, and facts of the matter

It is controversial as to whether beliefs and desires can be reductively explained in terms of levels of activity or activation of certain neurons or groups of neurons. While it is typically considered that beliefs and desires must be

realised by neural activity plasticity of function and the fact that different people have different neural pathways challenge the notion that there may be such a thing as a 'grandmother' neuron (or group of neurons) that fire at a specific frequency when and only when one is thinking of ones grandmother. While this is controversial I think that neural activity will not assist us in getting any further ahead with respect to what specific belief and desire produces behaviour.

In the spirit of reductionism neuro-scientists attempt to find the correlates of intentional states in brain behaviour. In order to do this, we must already have some way to determine whether the subject really was in a particular intentional state or not. If (a) we could determine what intentional state a subject is in, and (b) we found that it was correlated with something distinctive in the brain, then (and only then) could we use brain behaviour to correct our attributions of specific intentional states to assist us in determining what intentional state a given subject is in. The problem is that (a) is often indeterminate, in that multiple interpretations are possible, and there is also a problem in how we choose to operationalise intentional state terms (which would seem to me to be further grounds for indeterminacy). That makes (b) highly unlikely and (b) would always seem to be moderated by correlating brain behaviour with the bodily behaviour that we had to start with.

While there have been studies on the brain behaviour of DID subjects the data is hotly disputed. We have the bodily behaviour of systems and we are starting to look at brain behaviour of systems in order to assist us in explaining the behaviour of the system as a whole. A study was done where an fMRI scan was performed on a subject with DID when she switched between alters, and when she role-played switching to an 'imaginary' alter (Adler, 1999). While there were distinctive brain changes that were correlated with the 'genuine' as opposed to 'fictional' switch the significance of this finding is hotly disputed.

Suppose we grant that there were significant differences when the subject switched between alters. This finding still needs to be interpreted in order for us to decide on its significance. Most seem to agree that memories are contained within located modules in the cortex for the findings to have achieved such notoriety. If we grant this then we can argue about whether it shows us that some alters *cannot* access those memories, or whether they *choose not* to access those memories. All it shows is that some alters *do not* access those memories. Brain behaviour still needs to be interpreted and so it is hard to see how brain states can assist us in getting further ahead with

respect to either specific intentional states, or the number of intentional systems. No collections of behaviour (bodily behaviour, physiological responses, or brain behaviour) will help us explain or interpret the significance of the phenomenon. But it is the significance or interpretation of the phenomenon that interests us the most and is the main subject of controversy. Typically, what determines the issue still further and facilitates the discovery of 'obvious', 'crucial' data that decides which of the alternative theories is correct is the phenomena of theorists converging on a single theory. This seems to put the issue not so much in the 'to be determined by science' basket, but as well within the realm of conceptual analysis.

Despite my pessimism regarding the reduction of folk psychology I do not think that it is entirely accurate to write off intentional psychology as a mere fiction either. While it is indeed a matter of interpretation, the space of possible interpretations is restricted quite severely by reality constraints. There is the reality of the behaviour that we are seeking to explain and predict. There is also the reality of subsequent behaviour that may support or dis-confirm our attributions. These reality constraints are obviously enough to render our predictions and explanations indispensable to us in our daily lives, but there is still a space of indeterminacy where multiple interpretations are possible.

I do not think that this indeterminacy counts against intentional psychology, particularly as one might consider that our rendering of essential properties and laws of nature is in the same boat. It seems plausible that there could be an indefinite number of 'final', or complete sciences that predict and explain all the past, present, and future nerve hits of mankind (Quine, 1960 p. 23); and it seems equally plausible that there should be an indefinite number of intentional state attributions that could predict and explain all the past, present, and future behaviours of any given intentional system. Such a consequence is not fatal to intentional psychology as it is not fatal to physics; such indeterminacy would seem to be inherent in any attempt that we make to *predict*, *explain*, *interpret*, and otherwise *make sense* of any given phenomena. This aspect of indeterminism, rather than being unreliable is what makes life interesting. It is the scope within which we carve out our own explanations, interpretations, and meanings, of ourselves, others, and the rest of the natural world.

Gestalt switches and change in perspective

The problem of (2) can thus be recast as the problem as to whether we should adopt a single or multiple systems interpretation of a given subject's behaviour. While the reality constraints would seem to dictate that the single systems stance is appropriate for the prediction and explanation of the behaviour of the majority of intentional systems; other systems seem to exhibit behaviour that is clearly more amenable to multiple systems theory. In the majority of 'potential' cases it would seem that there is a genuine indeterminacy between a multiple or single systems view.

While the diagnosing clinician clearly believes that what I have called the multiple systems interpretation is appropriate, other clinicians show a clear preference for insisting on the single systems view. The post-traumatic model considers that alters are not distinct selves, rather the self is the summation or fusion of all the alters. Such supporters might consider that alters do not constitute distinct selves, and might be hesitant to even consider them to be distinct intentional systems. In order to work with the alters to access their memories for the treatment goal of integration or fusion, however, it would seem that the clinician is required to 'get to know' them as distinct intentional systems. Spanos, (1994) maintains that the very act of listening to alters' pronouncements of separateness and continuing the charade by using alternative names etc., is what serves to reinforce the disorder. The main criticism from sceptics is that while supporters maintain that in theory alters are not separate or distinctive selves they treat them as such in practice.

It would thus seem that there is a decision to be made as to whether one adopts a single or multiple systems interpretation of these subjects' behaviour. While one might consider that such a decision need not be made, I think that in practice it must as we unavoidably interact with one another on the intentional level. Every modern theorist that I have encountered seems to consider it an absurdity to consider that selves are real and that alters are selves. The main argument against supporters is that this is what they are doing in practice. Multiple systems theory, however, considers that there is a realist aspect to the self (the behaviours that legitimate our attributions), and that alters, as intentional systems are indeed as real as any self could be[11]. It would seem that there is no way around making a decision;

[11]The main argument against the 'reification' of alters seems to be an aversion to the legal consequence that we could not hold one alter responsible for another in a court of law. I do not think that considering alters to be selves logically entails this, however.

we are required to do so in our interacting with others on the intentional level.

This being said, the intentional stance demystifies the notion of a self or personality as an intentional system; and we need no longer make room in our explanations for a fixed and immutable Cartesian soul. Adopting an interpretation at one time would not seem to preclude adopting a different interpretation at another time. Such a change in interpretation could be considered something of a gestalt switch that is facilitated by a shaping in behaviours so where a duck may once have been legitimate a rabbit is more appropriate now. Supporters consider that working with alters is the best way to facilitate behaviours more amenable to what I have called a single systems view. It would seem that supporters and sceptics both are united in a common goal – altering these subjects' behaviour so that the single systems interpretation is the most natural, plausible, predictively adequate account of these subjects' behaviour. The disagreement would seem to be over the best way to achieve that.

Descriptive adequacy, aetiology, and treatment success

While Dennett, (1998 p.51) maintains that

> Charcot himself demonstrated only too convincingly, a woman who feels no pain when a pin is stuck into her arm *feels no pain* – and calling her lack of reaction an "hysterical symptom" does not make it any the less remarkable. Likewise, a woman who at the age of thirty is now living the life of several different selves is now *living the life of several different selves* – and any doubts we might have about how she came to be that way should not blind

Perhaps a distinction could be drawn between selves (to do with a psychological criterion) and persons (to do with a bodily criterion). I am grateful to Tery Hardwicke for the suggestion that the subject be considered a corporation for legal purposes. Corporations are (sometimes) considered legal persons and thus the corporation as a whole can be held accountable despite the innocence or otherwise of particular employees (selves) that constitute the corporation. While I am just providing a hint of a response here, I suggest it so as to illustrate that considering alters to be selves does not entail legal immunity. Though I suppose the main reason to regard corporations to be persons is actually to try and smear responsibility from individuals within the company who are making the legally problematic decisions.

us to the fact that such is now the way she is.

Dennett is primarily considering the florid cases that have been diagnosed. Sceptics maintain that the predictive success that is gained by the adoption of what I have called multiple systems theory is one that is a matter of self fulfilling prophecy, as clinicians legitimate and sustain the behaviours they have predicted as a confirmation bias. Dennett's emphasis, though, would seem to suggest that in the florid cases there is predictive leverage to be had, and I have considered that it is not only the sceptics who resist the multiple systems version of the intentional stance *in theory*.

While treatment outcomes are obviously an empirical matter it seems plausible to me at least that those with merely a 'window of diagnosability' may be more amenable to alterations in reinforcement contingencies which serve to shape behaviours towards an unambiguous, single systems view. The more florid cases would seem to result from the subject having adopted multiple systems theory regarding their own behaviours. Shaping such behaviours away would seem to lapse into 'punishment' both in the technical, and non-technical sense. Spanos considers a case where a hospitalised subject was ignored and placed in isolation when he switched into alters that the staff had decided to 'shape away'. Over time he did indeed switch less frequently and this is considered a prime example of how such behaviours may be 'shaped away' by sceptics. Such 'shaping' would seem to me to be questionable on ethical grounds – who gets to decide which alter will be reinforced, and which should be 'punished for existing'? Supporters maintain that they treat many such subjects who have been punished in the above fashion for 7-10 years and the subjects came to maintain that the alters did not disappear, they just felt unwanted and chose to come out at different times or mimic more closely the behaviour of the 'acceptable' personality. Such 'shaping' would also seem to be counter-productive with respect to establishing and maintaining a healthy rapport and therapeutic relationship.

With respect to the question of when the disorder emerged one might consider that alters emerged at the point where the multiple systems interpretation of their behaviour became a viable option. While the disagreement seems to centre on whether they were present from childhood or not, we may consider that alters emerged whenever the stance was adopted. If alters are best construed as intentional systems, as I have maintained, then clinicians can expect to find 'windows of diagnosability' in children should they seek them out with the multiple systems interpretation in mind. Whether alters

have been present since childhood or not would thus not seem to be either confirmed or disconfirmed by finding it in children despite some theorists considering this to be crucial data. Dennett, (1998) considered a subject who claimed that her alters originated in childhood when her father would call her by a different name and pretend to abuse someone else. He considers that whether this interpretation is offered by an abuser when the subject was a child, or years later when the subject is an adult and the interpretation is offered by a clinician would seem to be fairly arbitrary. To consider that the case of childhood origin was somehow legitimate, while the case of adult origin was an artefact of treatment would also seem somewhat arbitrary.

The other point of controversy is something that I will just touch on briefly. There is dispute as to whether the disorder is necessarily traumatic in origin, or whether Spanos account of multiple identity enactment shows us that trauma need not be a requisite for alters. There is a danger in considering a history of severe abuse to be a causative factor in the development of any disorder lest clinicians and clients both consider that it is the only rationally acceptable explanation for their behaviour. Hopefully we have learned something about memory as a constructive process so that the Freudian error is not repeated;

> I no longer accepted her declaration that nothing had occurred to her, but assured her that something *must* have occurred to her... Finally I declared that I knew very well that something *had* occurred to her and that she was concealing it from me; but that she would never be free of her pains so long as she concealed anything. By thus insisting I brought it about that from that time forward my pressure on her head never failed in its effect (Freud, 1953-74 p. 154 in Webster, 2003 p. 11).

It may turn out that the majority of subjects with the disorder do indeed have a history of severe child-hood abuse. With respect to explanation, however it would seem to me that diathesis could go a long way. Surely all that is required for the post- traumatic account is that the child *perceived* a great trauma. For an extremely sensitive child (or indeed an adult) circumstances may not have to be considered as objectively of 'sickening severity' for the individual to feel traumatised. Perhaps trauma is not a requisite and there may be other explanations for the emergence of alters, as Spanos has indicated.

Spanos maintains that the issue is not the existence of the phenomena, rather it is the origin and maintenance of the phenomena (thus the controversy is over (1) and (3)). I think, though, that by recasting the problem of (2) as to whether one adopts a single or multiple systems theory to explain and predict these subjects behaviour a new light is cast on (1) and (3). If there is a degree of indeterminacy as to whether the single or multiple systems stance is appropriate, then perhaps it is too much to expect empirical facts of the matter to determine which interpretation we should adopt. While there may be facts of the matter with respect to subjects' histories (which are inaccessible) and treatment outcomes there would seem to still be a genuine indeterminacy as to whether some subjects are best predicted and explained by multiple systems theory or single systems theory.

These subjects present with unintegrated memories, desires, beliefs, and goals and thus treatment consists in integrating them. The role of reinforcement contingencies clearly plays an important role in the establishment and maintenance of any intentional system, no matter how many we have associated with a single body. A re-conceptualisation of (2) may thus be able to cut through both of the extreme views on offer. The views reflect quite distinct treatment approaches and theoretical frameworks in that the socio-cognitive model is fairly behaviourist and the post- traumatic model is fairly psychodynamic. While cognitive-behaviour theorists seem to have largely side-stepped the disorder, trusting its conceptualisation to the behaviourists, perhaps a middle ground could be reached by a tradition that in practice seems to take from both psychodynamic and behaviourist traditions. Perhaps there could be a re-conceptualisation of the disorder in a way that is moderate and demystifying; though it would seem that any theorist needs to take a stance on whether the subject is best viewed as a multiple or single system. To realise that this is a matter of interpretation (and thus is amenable to reinterpretation) is to demystify the decision either way while taking seriously the phenomenon of the alters that are the distinctive feature of this diagnosis.

References

Adler, Robert (1999). 'Crowded Minds', *New Scientist*.

Apter, Andrew (1991). 'The Problem of Who: Multiple Personality, Personal Identity, and the Double Brain', in *Philosophical psychology*, Vol. 4, Issue 2.

Braddon-Mitchell, D & Jackson, (1996). *Philosophy of Mind and Cognition*. Blackwell Publishers.

Brown, Mark (2001). 'Multiple Personality and Personal Identity' in *Philosophical Psychology*, Vol.14, No.4, 2001.

Chalmers, David (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.

Clark, Stephen (1990). 'How Many Selves Make Me?', in *Royal Institute of Philosophy Conference on Human Beings*.

Davidson, C & Neale, J (2001). *Abnormal Psychology, 8th Edition*. John Wiley & Sons.

Davidson, D (1980). *Essays on Actions and Events*. Oxford University Press.

Dennett, Daniel (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*, Harvester Press Limited.

Dennett, Daniel (1987). *The Intentional Stance*, Massachusetts Institute of Technology

Dennett, Daniel (1988). *Brainchildren: Essays on Designing Minds*. Penguin Books.

Dennett, Daniel (1989). *The Origins of Self*, Cogito, 3, 163-73.

Dennett, Daniel (1991). *Consciousness Explained*, Little, Brown & Company.

Dennett, Daniel (1992). 'The Self as a Centre of Narrative Gravity' in F. Kessel, P. Cole and D. Johnson, (eds.), *Self and Consciousness: Multiple Perspectives*, Hillsdale.

Dennett, Daniel (1996). *Kinds of Minds: Towards an Understanding of Consciousness*. Weidenfeld & Nicolson.

Dennett, Daniel (1998). *Brainchildren: Essays on Designing Minds*, Penguin Books.

Feinberg, Todd (2001). *Altered Egos: How the Brain Creates the Self*, Oxford University Press.

Gillett, Grant (1997). 'A Discursive Account of Multiple Personality Disorder', *Philosophy, Psychiatry & Psychology* 4(3).

Gleaves, David H (1996.) 'The Sociocognitive Model of Dissociative Identity Disorder: A Reexamination of the Evidence', *Psychological Bulletin*, 120(1).

Glover, Jonathan (1998). *I: The Philosophy and Psychology of Personal Identity*, the Penguin Group.

Graham, George (2002). 'Recent Work in Philosophical Psychopathology', *American Philosophical Quarterly*, Vol. 39, No. 2.

Hacking, I (1991). 'Two Souls in One Body', *Critical Inquiry*, 17.

Hacking, I (1995). *Rewriting the soul: Multiple personality and the sciences of memory*. Princeton University Press.

Kluft, R (1988). 'The Phenomenology and Treatment of Extremely Complex Multiple Personality Disorder', *Dissociation*, 1.

Kluft, R (1991). 'Multiple Personality Disorder' in Tasman, A & Goldfinger S (Eds.,) *American Psychiatric Press Review of Psychiatry*. 10.

Kolak, Daniel (1993). 'Finding Our Selves: Identification, Identity and Multiple Personality', *Philosophical Psychology* Vol.16 No.4.

Lilienfeld, Scott O; Lynn, Stephen Jay; Kirsch, Irving; Chaves, John F.; Sarbin, Theodore R.; Ganaway, George K.; Powell, Russell, A., (1999). 'Dissociative Identity Disorder and the Sociocognitive Model: Recalling the Lessons of the Past', *Psychological Bulletin*, 125(5).

McHugh, Paul, and Putnam, Frank (1995). 'Resolved: Multiple Personality Disorder Is an Individually and Socially Created Artefact (rebuttal)' *Journal of the American Academy of Child and Adolescent Psychiatry* 34 (7).

Merckelbach, Harald; Devilly, Grant J.; Rassin, Eric, (2002) 'Alters in Dissociative Identity Disorder Metaphors or Genuine Entities?' *Clinical Psychology Review*, 22.

Pitt, David (2001). 'Alter Egos and Their Names', *The Journal of Philosophy.*

Pope, Harrison G; Oliva, Paul S.; Hudson, James I.; Bodkin, Alexander J.; Gruber, Amanda J., (1999). 'Attitudes Towards DSM-IV Dissociative Disorders Diagnoses Among Board-Certified American Psychiatrists' *American Journal of Psychiatry*, 156(2).

Putnam, F.W, (1989). *Diagnosis and Treatment of Multiple Personality Disorder*, Guilford Press.

Quine, W (1960). *Word and Object.* MIT Press.

Saks, Elyn (1994). 'Integrating Multiple Personalities, murder, and the Status of Alters as Persons', in *Public Affairs Quarterly*, vol. 8, No. 2.

Spanos, N; Weekes, J; Bertrand, L. (1985). 'Multiple Personality: A Social Psychological Perspective', *Journal of Abnormal Psychology*, 94(3).

Spanos, N (1994). 'Multiple Identity Enactments and Multiple Personality Disorder: A Sociocognitive Perspective' *Psychological Bulletin.* 1. 116.

Thigpen C & Cleckley, H (1984). On the Incidence of Multiple Personality Disorder. *International Journal of Clinical and Experimental Hypnosis*, 32.

Webster, R (2003). *Freud.* Weidenfeld & Nicolson.

Wilkes, Kathleen (1988). *Real People: Personal Identity Without Thought Experiments* Oxford University Press.

Chapter 3

Direct reference: A route to dualism?

Introduction

Descriptivist theories of reference consider that cognitively accessible descriptions provide a successful criterion for fixing the reference of expressions. The arguments against the psychological reality of descriptions that are necessary or sufficient for determining reference are often considered decisive. The modal arguments lead in to the notion that what is necessary and sufficient to determine reference are 'real nature' essential properties to be determined by science. These essential properties are not required to be cognitively accessible to speakers; but they determine the nature of the object/s that can be legitimately referred to by way of that expression. While the descriptivists focused on appearances, or qualitative aspects of the phenomena with respect to the appropriate place to look for reference determination; direct reference theorists consider this level to be irrelevant. Putnam, Kripke, Chalmers, and Braddon-Mitchell & Jackson consider that it is metaphysically possible for qualitative properties and 'real nature properties to vary independently of one another. 'Real nature' properties determine the referents of expressions that denote objects / kinds of stuff in the world, and qualitative properties determine the reference of mental state terms.

Instead of haggling over the primacy of whether 'real nature' essential properties or qualitative essential properties determine the reference of various expressions I will argue that such a stipulated dichotomy between qualitative properties and 'real nature' properties is misguided. It is typically accepted that the properties coincide in the actual world, and I think that

if we accept that they may be teased apart in various possible worlds then dualism is an inevitable consequence of our stipulation. We are left with the difficulty of deciding whether the qualitative or the 'real nature' criterion is relevant for various expressions, and we may need to resort to there being two different criterion, or 'senses'. I take the dualistic consequence to be a reductio ad absurdum of such a stipulated dichotomy between qualitative and real essences. While the world in itself cannot determine whether qualitative and real essences are essentially connected or not there are advantages to stipulating that they are essentially connected rather than stipulating that they are essentially different. Stipulating that they are essentially connected has the desirable consequence that our inter-subjective experiences of the world are essentially related to, and thus may plausibly be explained by, our scientific accounts of the natural world. When dualism about 'water' and 'heat' are disallowed, we may plausibly avert dualism about the mind as well.

I shall begin with a characterisation of the descriptivist account of reference that was widely accepted preceding the rise of direct reference theories. This is required in order for me to illustrate the difference between the 'qualitative' properties that interested the descriptivists, and the 'real nature' properties that interested the direct reference theorists. We also need to be aware of some of the problems that the descriptivist account faced so that we can ensure that any modifications made to the direct reference theory does not have it fall victim to the same problems and objections that were ultimately fatal to descriptivism.

Descriptivism: The word–world link by way of description

Frege and Russell offered descriptivist accounts of the way in which speakers succeed in referring to objects by using language (in Baillie, 1997 pp.1-69). Frege maintained that speakers have a cognitively accessible description, or sense, that they associate with an expression. Russell held that names, as we usually know them, were equivalent to definite descriptions. These senses or descriptions were thought to provide necessary and jointly sufficient conditions which provide a criterion that a single object may meet. If there happens to be an object in the world that meets the criterion that the description provides then speakers can succeed in referring to that object by way of that expression. The expression actually chosen may be elliptical in

that it abbreviates a longer description, and it is in virtue of an object's meeting that longer description that reference is achieved.

Frege required senses to be inter-subjective so that different speakers could pick out the same object by uttering different instances, or tokens of the same expression. Frege considered that the sense of a sentence was part of a 'common store of thoughts which is transmitted from one generation to another' (in Baillie, 1997 p.26). The descriptions that we associate with expressions are passed on to new speakers and it is grasping the sense that enables them to achieve linguistic competence. He maintains that 'the sense of a proper name is grasped by everybody who is sufficiently familiar with the language or totality of designations to which it belongs' (in Baillie, 1997 p.24). On this account one could not succeed in using language to refer unless one grasped the criterion that the referent must fill.

Russell, in reducing proper names to descriptions which themselves are reduced to expressions containing logically proper names may be thought to have altogether missed the point of attempting to construe either description or reference as inter- subjective. Logically proper names were supposed to pick out sense data that the speaker is acquainted with. Because one cannot access another's sense data in principle Russell seems hard pressed to account for language as an inter-subjective, social phenomenon. On Russell's account it is hard to see how there can be shared reference as individuals live in worlds with different objects that are composed of our subjective, idiosyncratic conjunctions of sense data. Russellian analysis does not seem to help us understand how language can function to secure either an inter- subjective or objective referent. On his account it is hard to see how language, as an essentially social phenomenon, is possible at all[1].

Fregean and Russellian accounts are thought to be similar in that they largely agree on the role that descriptions play with respect to linguistic phenomena. The following three theses are usually considered to characterise the descriptivist view of the role of descriptions.

- (i) *Cognitive Accessibility*; Descriptions are held to be cognitively accessible to the speaker.

[1]Although this may be averted if one could provide an account of individual sense data as tokens of an inter-subjective type. Russell considered that we could be acquainted with universals and thus have knowledge of them, so perhaps it is unfair to Russell to say that he missed the point of inter-subjectivity.

- (ii) *Inter-subjectivity*; For all members of the speech community the same description fixes the significance of a name[2].
- (iii) *Reference Determining*; Descriptions are necessary and sufficient to determine unique (individuated) reference[3].

One might reasonably expect cognitively accessible descriptions to be those that speakers would report when questioned as what they meant. Typically reported descriptions, however, are idiosyncratic and show considerable variation between speakers. Cognitively accessible descriptions thus do not seem to be inter- subjective. Frege concedes that different speakers may associate different descriptions with the same expression and he gives the following example of two different senses that speakers may associate with the name 'Aristotle' (in Baillie,1997 p24).

- (iv) The Pupil of Plato and the teacher of Alexander the Great.
- (v) The teacher of Alexander the Great that was born in Stagira.

Frege notes that 'So long as the thing meant remains the same, such variations of sense may be tolerated, although they are to be avoided in the theoretical structure of a demonstrative science and ought not to occur in a perfect language' (in Baillie, 1997 p.24). It is hard to see from the above account, though, how there is any assurance that the 'thing meant' remains the same as Alexander may have had many teachers that were born in Stagira. Descriptions like (v) are not generally sufficient to individuate a referent, and it seems that most of the cognitively accessible descriptions that speakers report are like this. Descriptions may also be seen to be unnecessary for reference if we consider the expression 'I wonder who Mary Elizabeth Baxter is?' to refer to Mary Elizabeth Baxter despite the speaker having no

[2]Searle may be thought to offer an account of inter-subjectivity that is looser than this with his cluster theory of descriptions which allows for some variation in the descriptions that speakers associate with expressions.

[3]I am not sure that either Frege or Russell would accept all elements of the above characterisation. Russell wouldn't seem to require inter-subjectivity, and we will go on to consider an example where Frege allows inter-subjectivity to be breeched. A conjunction of these three claims seems to be the standard characterisation of the descriptivist thesis, however. Even if this characterisation is something of a straw man it can still be useful in showing us what cannot be required for reference.

associated description. These versions of descriptivism thus do not seem to provide an adequate account of how expressions succeed in denoting objects in the world.

Direct reference: An unmediated word–world link

Although some have attempted to modify the descriptivist account (e.g., Searle) many have abandoned the descriptivist paradigm in favour of an alternative initiated by the work of Kripke, (1972) and Putnam, (1975). While Descriptivists maintained that reference was achieved by way of a cognitively accessible sense, meaning, or description; Kripke and Putnam offer us an alternative account. According to the direct reference account some of our expressions are thought to hook on to the world directly in a way that is not mediated by a description. Kripke initially addressed the issue of the reference of proper names of individuals, though he also considers natural kind terms such as 'water' and 'gold'. Putnam independently worked on natural kind terms and he attempted to extend the account to expressions that refer to artefacts. Kaplan, (1989) considered indexical expressions, and introduced an operator allowing some expressions denoting objects to be analysed in a directly referring way[4]. Much work has been done on attempting to extend the account to plausibly embrace different kinds of expressions. There has also been much debate as to whether an inevitable consequence of direct reference is dualism with respect to mind and body as Kripke takes it to be; or whether this may somehow be averted.

Modal contexts and rigid designation

The differences between descriptivism and direct reference are most striking when considering the way in which they handle modal contexts.

- (vi) 'Aristotle could have been born in Greece and become a vase painter rather than turning to philosophy'.

[4]I will consider Kaplan's 'dthat' operator further in the subsequent section on modal contexts.

Kripke considers that the intuitive analysis of such an expression would be that the man denoted in the actual world by the expression 'Aristotle' could have done such and such. Kripke thus considers names to function as rigid designators in that they refer to the same individual across possible worlds (Kripke, 1972 pp.4-15). A descriptivist analysis of (vi) however, would produce a different outcome depending on which description we take to be the sense, or which description we take to be deductively implied by the expression. If we consider a conjunction of (iv) and (v) to be the correct analysis of 'Aristotle' then the object referred to must have done those things (or meet that description) as a matter of analytic necessity, and thus (vi) is (implausibly) not only false; but it is false as a matter of logical necessity. Metaphysical possibility would thus seem to be constrained in a highly counter- intuitive analytic, or logical (as opposed to empirical or a-posteriori) way[5]. If (iv) and (v) are taken to be an adequate analysis of 'Aristotle' then, Kripke maintains, the name would serve to pick out whoever happened to do those things in any given counter-factual situation. Descriptions would thus function to pick out different individuals who happen to answer to the description across different possible worlds (Kripke, 1972, pp. 6-7).

One could talk about 'this essay in front of you now' and consider that it is quite possible that another essay (e.g., the next one in the pile) could have been the one that is in front of you now. The descriptivist account would thus be a suitable analysis if the intended meaning of the expression was something along the lines of 'whichever essay was in front of you at time t'. Kaplan, (1989) considers, though, that we may use a description solely in order to get to a particular individual or thing. We could thus consider that it is possible that ''dthat' essay in front of you now might not have been in front of you now' if, for example it had been sent to someone else. Kaplan's 'dthat' operator thus allows one to rigidify a description.

Kripke considers that an appropriate analysis of names is that they function as rigid designators. If we consider a name in some kind of modal sentence such as a counter-factual conditional then an appropriate analysis is that the name picks out the very individual in the actual world and describes that individual in different (possible) circumstances. Kripke con-

[5]We could not alter the description without picking out a different individual with different essential properties. The distinction between logical and metaphysical necessity and possibility is a problem within philosophy of science, however it would seem that there can be no difference between them on the traditional descriptivist account, which some at least find counter-intuitive.

siders that descriptions function as flaccid, or non-rigid designators in that they designate different individuals across different possible worlds as different individuals happen to meet a given description. Since names are rigid and descriptions are non-rigid Kripke concludes that names cannot be equivalent to descriptions.

Essential Properties and Metaphysical Possibility and Necessity

Kripke introduces the notions of Metaphysical possibility and necessity as a 'surprising consequence' of direct reference. He maintains that the objects in the actual world that are directly referred to have essential properties that must be retained in order for the object to remain the same object. This has consequences for the counter-factual situations that we can legitimately describe with respect to any given object. There is no possible world in which 'dthat' object can appear without the essential feature required for the objects identity as that object. Kripke considers that science will empirically discover the relevant essential properties[6]. He is thus a scientific essentialist about essential properties, which seems a plausible view on the reference of natural kind terms.

He maintains that this notion of necessity is not a matter of analytic or conceptual truth, rather it falls out as a consequence of the way the world is. The objects in the world have 'real nature' essential properties that may or may not be known by us. We do not need to know these properties in order to refer successfully to the objects but the properties do determine what is objectively possible and impossible with regards to those objects. They also serve to show how we can be mistaken in our categorisations when we go by observed features instead of the reference determining 'real nature' ones. Kripke maintains that his scientific essentialism is something of an 'aside' and not a requirement of his thesis. There does need to be something to distinguish between what is and is not metaphysically possible, however, for the notion of metaphysical possibility and necessity to be meaningful. Kripke presupposes a realist metaphysics where:

- (a) Objects in the world have essential properties intrinsically.

[6]Mental states are exempt.

- (b) When we refer to these objects (by demonstration) we are referring to their 'real nature' essential properties whether or not we know what these are[7].

- (c) Part of the scientific enterprise is the discovery and characterisation of these 'real nature' essential properties. Direct reference theorists typically take the relevant essential properties to be those that the 'final science' would endorse.

The arguments against the role of descriptions for reference are often considered decisive, especially when compared to the relative success that the direct reference theory has had while to a large extent by-passing them[8]. What the greater moral of this story is, though, remains controversial. Direct reference theorists typically take the moral of the story to extend beyond the fact that cognitively accessible descriptions cannot determine reference. They consider that it demonstrates the irrelevance of qualitative or experiential properties with respect to determining reference altogether[9].

Searle is often considered an advocate, or defender of the descriptivist approach. He attempted to alter the theory to render it more plausible by maintaining that, instead of providing strict necessary and sufficient conditions for reference, descriptions functioned as cluster concepts (in Lycan, 2000 pp.42-43). Many, or most but not all of the description needs to apply to an object in order for it to be referred to by way of the expression. A certain amount of variation with respect to different speakers associating different descriptive features with an expression could thus be tolerated. Perhaps Searle and others who oppose the direct reference analysis are guided by intuitions that one cannot simply disregard qualitative or experiential properties with respect to reference determination.

---

[7]Salmon, (1981 p.42) considers that some direct reference theorists consider that the proper referent is the essential properties; though others maintain that the referent is the object whose nature is determined by its essential properties.

[8]Putnam does consider the role of descriptions, or stereotypes (qualitative features) with respect to how we fix reference or categorise in the actual world, and Kripke seems to make a similar concession. Descriptions and qualitative features are still held to be irrelevant, though for determining the essential properties that are relevant to the majority of our expressions, and thus irrelevant with respect to assessing counter-factual situations. I will go on to consider this further.

[9]Mental state terms are exempt.

Qualitative / experiential properties and 'real nature' properties

The deference to 'real nature' properties initially seems plausible because we typically do acknowledge that qualitative features, or appearances can be misleading, and we can be mistaken. Kripke considers a situation that Donnellan, (1966) describes where someone is at a party and asks 'who is the man drinking a Martini?', though it turns out that there was water in the Martini glass, and thus the term 'Martini' was incorrectly applied. Qualitative properties also vary as in the case of natural kinds when we see an object or a substance in new circumstances yet we want to maintain that it is the same object or the same stuff. It is also considered that we need 'real nature' essential properties to fix the reference so that different scientific theories can be about the same object, or the same thing. Presumably the referent does not change as we accumulate experiences with the referent, and we refer to the same substance with the expression 'water' as those who picked out the 'same kind of stuff' in lieu of a scientific theory of it's nature. If we want to say that we are referring to the same kind of stuff, the metaphysical realist considers that we require objectively existing 'real nature' properties to fix or determine the real nature of the referent.

While Kripke, Putnam, Chalmers, and Braddon-Mitchell & Jackson consider observable properties to be something of a ladder that allow us to ascend to 'real' essences and then may be tossed aside as irrelevant I wish to consider whether this disregard for qualitative properties is a wise move.

Firstly, in Donnellan's example the speaker would, presumably, acknowledge their error if they tasted the 'Martini', or smelt it, or had drunk enough of it. I do not see that this example entails a fault with qualitative properties in general; rather the categorisation error would seem to result from inadequate qualitative information or experience. We sometimes consider that we are mistaken because later experience provides further information that shows us we were in error. We would like to say that we would not have judged or categorised as we did if we had access to that further information at the time of categorisation. One thus does not need to descend to the 'real nature' level of analysis in order to provide an account of how it is possible that we can be mistaken. In the literature qualitative properties seems to have been equated with 'appearances', which are often associated with a superficial glance. Appearances are thought to be deceptive much of the time, and by equating qualitative properties with appearances qualitative proper-

ties have been viewed as unreliable and have thus fallen into disrepute by association.

I think that we are wise to consider the role of appearances or qualitative properties with respect to the scientific enterprise itself. It is probably safe to say that most people conceive of science as the enterprise of predicting (and thus explaining) qualitative or experiential features of the world[10]. It should be noted that I am using 'qualitative' or 'experiential' in the sense of the kinds of features that interested the descriptivists, the observable or 'watery-stuff' kinds of features. The posited ontology of science consists in entities with theoretical essential features that are defined in such a way that they functionally interact according to laws of nature. This functional interaction produces what we perceive, observe, or experience as the behaviour of middle sized objects. Or what we would observe of these objects if the appropriate boundary conditions were met. This may be seen when one considers that the success or otherwise of scientific theory is judged by it's *applications* to the phenomenon we observe. The properties relevant to Newtonian objects would seem to be inertia, mass, velocity etc, and these are defined according to how they interact according to Newton's laws in order to produce the phenomenon that we observe. At this theoretic level of analysis essential properties may thus be described as *functional* properties that get us from a law to a phenomenon. We may likewise consider laws to be a function from essential properties to a phenomenon. Essential properties and laws of nature may thus be seen to be inter-defined.

To clarify the sense in which I am using the notion of 'function' it may be useful to consider a couple of examples of a similar notion of function that I have encountered in the literature. Frege considered concepts to be functions that get us from an object to a truth-value. Concepts are thus defined functionally according to how they interact with other notions such as 'object' and 'truth value'. We can thus consider a concept to be a 'black box' or placeholder where the properties that are essential to its nature are 'black box' properties that determine how it interacts with objects and truth-

[10]While it remains controversial just what the enterprise of science is, it does seem clear that the mark of a 'good' (or indeed 'bad') theory is whether it can adequately predict and explain phenomena that we encounter in the world. While 'functional laws permit of values which a real parameter could not attain, for instance very high temperatures, as in the ideal gas law: PV = nRT' (Weinert, 1995); it would seem that its adequacy is assessed by the predictions it makes with respect to phenomena we may observe (within the relevant boundary conditions), and the explanation (as part of a greater theory) that can be offered as to *why* we will never observe the phenomena that the law rules out.

values. Ramsey sentences use a similar notion of function to show us how we can functionally define mental state terms / mental states so as to avoid circularity. We can treat beliefs as a 'black box' or placeholder that interacts with desires to produce behaviours. We can also treat desires as placeholders that interact with beliefs to produce behaviours. Beliefs and desires thus functionally interact with each other to produce the behaviour that we observe. While this account is clearly insufficient to differentiate beliefs from desires, it seems plausible that they engage in other functional relationships that are different (belief forming mechanisms are different from desire producing mechanisms).

I would say that terms such as 'mass' and 'velocity' are functional terms in the same way. A scientific theory of 'real nature' essential properties has them interacting with the laws of nature in order to produce the phenomena we observe. 'Light' for Einstein was a constant (by definition if you like), whereas this was not explicitly so for Newton. The terms would seem to be defined differently with different essential properties interacting with different laws of nature in order to produce (upon careful observation) slightly different phenomena. While scientific realists may consider that Newton was wrong about the essence of light (and thus 'light' has the same referent in both theories); perhaps the reference is the same in virtue of 'light' being a term for (very nearly) the same *phenomenon* for both theorists (with respect to the two theories being theories of fairly much the same sets of experiences). If we want to say that Newton was wrong about the essence of light it would seem to me that this is because he could not predict (thus explain) the phenomenon as adequately as scientists observed anomalies when they were more thorough with their observations. The role of qualitative properties thus seems to play the same role for scientific investigation as it does for the speaker in the gin / water example. In both cases our judgements are revised in the face of later experience.

Essential properties may thus be best thought of as functional, rather than intrinsic. If an object can be found that does not obey Newton's laws then the object would not be a Newtonian object. Newton would have been wrong about the essential properties and laws of nature *of that phenomenon*. If we are entitled to say that Newton was wrong about laws or essential properties this is in virtue of our observing that some objects do not in fact behave as his theory requires[11].

[11]Indeed we have restricted the Newtonian theory as adequate provided that certain boundary conditions obtain. Theories with boundary conditions are the ones that explain

I am aware that this characterisation of essential properties may be controversial (especially to scientific realists). I do not have the space here to provide a thorough and sustained argument for it. I introduce it merely to outline an alternative route to enable different theories or accounts to be of the 'same thing'. They refer to the 'same thing' in virtue of accounting for (to a very large extent) the same phenomena, or the same experiences that we encounter. The difficulty of incommensurability, or reference change with every new theory and it's postulated objects / substances with essential properties does not require a scientific realist metaphysical system for its solution.

While some consider realism with respect to essential properties and laws of nature to be the only way in which to answer these difficulties I think that we can find a satisfactory answer from the level of observation, appearances, and experience. Different theories can be assessed with respect to the degree of adequacy that they have in the prediction and thus explanation of the behaviour of objects that we observe. The alternative is to posit reference determining essential properties that do not co-vary with our experiences and observations of the world and this has the consequence that they are beyond our grasp in principle.

Kripke (1972, pp.135-137) considers that we initially baptise a sample of water by ostensive definition. The term is then passed on from speaker to speaker in virtue of a causal-historical chain extending back to the baptism ceremony. New speakers succeed in referring to the same kind of stuff (or indeed an individual) in virtue of intending to use the expression to refer to the same thing as the person that they heard the name off. In accepting Kripke's account of a legitimate baptism of a sufficient sample of 'water' the problem becomes the issue of how we determine what is relevant for fixing the reference determining essential properties.

Kripke considers it is the real nature to be determined by science and that the appearances are irrelevant, but he does not give adequate consideration to the way in which scientists 'discover' or 'create' their essential properties

---

and predict the phenomena that we observe. It is sometimes considered that the genuine laws (that is to say the universal laws) of nature will explain and predict the boundary conditions that the boundaried theories presume. These genuine laws thus will not predict and explain phenomena directly. I would think, though that whether they are acceptable or not will have quite a lot to do with their utility in predicting and explaining experience indirectly, by way of the boundaried (and thus not truly universal) theories or laws of nature which directly explain and predict observed phenomena.

in the first place. I propose that what is relevant (once we have a legitimate initial baptism) is that the other instances are always observed to behave the same (in relevant respects) to the initial sample[12]. We may consider how scientists actually go about categorising samples of various substances in practice. They make observations of the objects behaviour and they perform experiments and observe the resulting behaviour. If something behaves differently from what would be expected from the sample then this is the evidence that enables them to infer that it cannot have the relevant essential property and thus is a different kind of stuff. This is why science is correctly considered to be an *empirical* endeavour, because it is attempting to explain, describe, and predict the phenomena we encounter.

While the above may be too sketchy to win converts I will consider three cases in which my approach leads to conclusions that are different from the ones typically reached by direct reference theorists. There are practical advantages to accepting the thesis that the 'real nature' level of analysis is determined by how things appear to be, and our posited essential properties and laws of nature which are supposed to predict and explain our observations. If we consider that it is illegitimate to divorce observable properties from 'real nature' properties we may be able to avoid Kripke, Chalmer's, and (arguably) Searle's dualism about mind and body. We also do not have to concern ourselves with the arguments as to whether qualitative or real properties are relevant for determining reference, especially for expressions that pre- date science. Before I do this though, I will deal with an objection that qualitative properties entail an unsatisfactory reversion to descriptivism.

Salmon, (1981, pp.22-23) makes a point about the possibility of essential properties being described. He considers that the descriptivist might object that if essential properties can be described then the descriptivist account was a correct analysis all along (though the requirement of cognitive accessibility would have to go). Salmon states that even if 'real nature' essential properties can be described that does not mean that they are essentially descriptions. It is to no cognitive advantage that they can be described and descriptions don't seem to be required in order for speakers to successfully refer. On the typical account of descriptivism [(i), (ii), (iii)] cognitive accessibility is often held to be fairly analytically a major thesis in the descriptivist account. It would indeed seem to be vital as without it (and the implausible cognitive accessibility claim) descriptivism is in danger of collapsing into direct reference.

[12]'In relevant respects' is a vague notion. I suspect that relevant respects have a lot to do with our purposes but will not pursue this further.

It seems to me, though, that Salmon's point could equally be made regarding qualitative properties. Although the descriptivists' focus on qualitative properties seems to be run together with their focus on descriptions in the literature, I think that the two notions should be separated out. I don't think that the moral of the failure of descriptivism was its focus on qualitative properties; rather it was due to its focus on the role of descriptions. The direct reference account does seem to be the most plausible account of reference that we have but I think that the role of qualitative properties with respect to reference determination needs to be restored. 'Real essences' may be described in the same way that observable essences can be, but neither are essentially descriptions. Real essences may be known or not known by various speakers, and some observers may make more thorough observations than others but 'real essences' cannot be the reality that we want to capture if they do not essentially predict and explain our observations and experiences of the world.

Putnam's 'Twin Earth' and 'water' as essentially H_2O

Putnam, (1975) describes what he takes to be a metaphysically possible world in which there is a substance XYZ that is qualitatively identical to H_2O. He maintains that it is not metaphysically possible that XYZ is water, as the essential property of water is H_2O. The claim that "'water' is H_2O" is considered to be provisional in that we have yet to see whether it will be endorsed by the final science. The point, though, is that whatever the final science endorses as the relevant essential property is essential to the substance. I do not take issue with this, I just take issue with the notion that it is metaphysically possible to separate qualitative properties from essential properties the way that Putnam has[13].

Chalmers, (1996, p.57) considers that we can separate 'A' and 'B' intensions which seems to be just another way of saying that it is legitimate to

[13]Perhaps Putnam wouldn't really want to say that earth and twin earth are qualitatively identical. If he does not want to say this then the Twin Earth thought experiment result seems hardly surprising at all. People often mis-categorise when they only have a quick glance. Perhaps intuitions as to the plausibility of Putnam's result co vary with the degree to which one equates qualitative properties with cognitively accessible descriptions and / or the superficial.

separate qualitative properties from 'real nature' properties. He maintains that the 'A' intentions enable us to fix the referent in the actual world. The 'A' intensions, or the qualitative properties of water are that it is watery-stuff. 'Watery-stuff' seems to be taken as shorthand for the observable qualitative properties of water; e.g., that it is odourless, colourless, falls from the sky etc. Chalmers maintains that it just happens that in the actual world watery-stuff turned out to be H_2O. Because of the direct reference take on qualitative properties as superficial it turns out that the correlation isn't perfect, but it is good enough to fix H_2O as the relevant essential property, or the 'B' intension of the term 'water'[14]. One might consider that the 'direct' route to reference has turned out to be an indirect route to essential properties (they are reached by way of watery- stuff) but it is indeed hard to see how it could be otherwise. Chalmers considers that although watery-stuff and H_2O are correlated fairly often in the actual world (and it is in virtue of this correlation that we were able to identify 'water' as H_2O) once we have determined the essential nature of 'water' e.g., that it is H_2O then this is what is relevant to determining metaphysical possibility. The qualitative properties turned out to be something of a ladder that enabled us to get to the real, and once the real was reached the qualitative then falls out as irrelevant (Chalmers, 1996, pp.57-59).

Braddon-Mitchell and Jackson, (1996, p.71) maintain that it is possible that H_2O appear black and tarry and it is possible that watery-stuff be XYZ (as on twin earth). They thus consider that it is metaphysically possible for qualitative and 'real nature' properties to vary independently of one another[15]. Kripke and Putnam also both state that it is metaphysically possible for the real nature and qualitative properties to vary independently of one another. One may consider how the scientists know that the substance on twin earth is not water. The only way they could know this is if they observed it to behave differently from water. It is literally inconceivable that

[14]I maintain that if one distinguishes the superficial from the qualitative then the qualitative is correlated perfectly with the real (in the relevant respects). I think that this is as perfect a law as any other to be found within science.

[15]They even go so far as to say that this could occur without the scientists even needing to explain why the H_2O appeared black and tarry. This seems to me quite absurd as without a good explanation (as to the breech of boundary conditions) we would have a counter-example to the claim that H_2O is water. While a defence may be that this does not occur in an actual world, but in a counter-factual situation I think we should allow the qualitative and real coincidence in the actual world to dictate metaphysical possibility. Especially considering the way in which these real essences are (and always must be) arrived at.

something appear exactly the same yet differ in internal structure. This is multiplying entities beyond necessity and the 'real essences' in such a case would be idle and empty, they would do no work[16].

There is a line which we could use to challenge whether the substance on twin earth was water or not. One could consider that our pre-scientific term 'water' should have it's reference fixed by the pre-scientific qualitative level of analysis. While I have some sympathy with this approach it once again rests on a difference in kind between qualitative and 'real nature' properties. I think that instead of debating whether qualitative or 'real nature' is primary, indeed instead of allowing them to vary independently across metaphysically possible worlds we are best to see them as lying on the same (empirical) continuum. Chalmers acknowledges that the qualitative and 'real' coincide in the actual world. It seems to be just in thought experiments that are supposed to determine metaphysical possibility that they vary. Is this a discovery about the nature of necessity though, or a stipulation?

If they are found to co-vary in the actual world and it is only in virtue of this correspondence that we ever had a notion of the 'real essence' then why don't we consider that the qualitative properties and 'real nature' properties must co-vary of metaphysical necessity? This acknowledges the way in which science actually does operate to discover (or create) 'real essences'. I cannot think of any reason why we should not stipulate this way. Stipulating in this direction may also lead to a more satisfactory analysis of the following case, which may indicate a way in which we can avoid dualism about the mind.

---

[16]On Dupre 'Natural Kinds and Biological Taxa', Geoffrey Reid writes: 'Maybe there is also a logical point... Suppose the molecular structure H_2O is both necessary and sufficient to explain the phenomena...Then XYZ cannot be both necessary and sufficient to explain the same phenomena. For the sufficiency of the first denies the necessity of the second. If H_2O will explain the phenomena, then we do not need XYZ to explain the phenomena. XYZ is not (and cannot be) a necessary condition of the phenomena (personal correspondence). If XYZ is always observed to behave the same as H_2O it would seem idle, empty, and pointless to multiply structures beyond necessity. It would seem to me that the scientists would be best to consider that H_2O = XYZ. The difference between H_2O and XYZ must ultimately be detectable from the qualitative or experiential level to be meaningful.

Heat: Qualitative or 'real'?

Kripke acknowledges that there is an ambiguity of reference for the expression 'heat'. He considers that 'what seems hot to me' or the sensation of heat essentially has a qualitative feel and thus has an essentially qualitative referent Kripke, 1972 pp.148- 153). He thus treats it as a mental state term as he gives the same analysis to 'pain'[17]. He considers that our pre-scientific term 'heat' has a 'real nature' referent as we took heat to be a property of external objects rather than qualitative sensations whereas he goes the other way with mental state terms. The expression 'heat' thus refers (except when it is a-typically used in the first sense) to whatever the 'real nature' of heat turns out to essentially be (provisionally, mean kinetic energy).

Searle takes the opposite reading. He considers that our pre-scientific term 'heat' referred to a sensation and once science came along we changed the meaning of the term, or the criteria for determining what essential properties were relevant. He takes the qualitative reading to be primary and the 'real' reading to be derivative; the result of a change in meaning / referent (Searle, 1992, p.119). He maintains that we have changed our criterion as to where to look for (or what is relevant for) reference determining essential properties. He thus considers it an analytic or stipulative matter as to whether we consider the qualitative or the 'real' to be the relevant place to look for essential properties. The world in itself would not seem to be enough to distinguish which is relevant for the reference of our expressions; stipulation is needed. I think that this is especially true in that qualitative and 'real' properties are found together in the actual world. The distinction between the qualitative and the 'real' may thus be a formal distinction rather than a real one. It is hard to see how formal distinctions determine metaphysical possibility rather than analyticity. We do not need to choose between Searle's deference to qualitative properties and Kripke's deference to 'real' properties if we consider that both criteria are correlated in the actual world and thus perhaps we should more appropriately stipulate that both criteria provide essential properties that determine the nature of the referent.

[17]It is interesting to consider whether 'pain' can be essentially private for Kripke, or whether he requires a notion of inter-subjective types of experiences in order to avoid private language difficulties. If he requires inter-subjective types (as I think he does) then perhaps the most plausible account of them is that they are functionally defined with private qualia filling something of a 'black box' whose 'essentially private nature' may be just as implausible as 'essentially real nature' reference determining properties when divorced from some notion of inter-subjective appearances / phenomena that are subject to public observation.

We seem to be forced to choose between qualitative properties and real nature properties for essential properties that are supposed to fix the referent for various expressions. We could bypass this problem by defining heat as a function of mean kinetic energy producing characteristic sensations in an observer. If there is no observer and we want to specifically talk about the object then we more properly have mean kinetic energy with a power to produce the sensation (if an observer were present). If we just have a hot sensation without the mean kinetic energy then we have a 'heat-sensation' which is a 'heat'-sensation in virtue of someone considering it to be the sort of sensation typically produced by mean kinetic energy. These both seem (to me) to be slight deviations from the more properly both-aspect-inclusive reference determining essential properties of the term 'heat'. They are passable when we speak loosely.

The moral of the story

While descriptivists faced insurmountable problems with cognitively accessible descriptions I think that they made no mistake in focusing on observable properties. I might have been a little unfair on direct reference theorists; Putnam might not want to maintain that Twin Earth is a qualitative duplicate of Earth (with respect to all the experiences that the citizens could encounter on that planet). I presented his case in this light, though, in order to convince the reader that it is metaphysically impossible that Twin Earth be a qualitative duplicate of Earth. My position can be summarised in the following claims:

- (i) Qualitative properties and 'real' properties do not vary independently of one another in the actual world: so long as we distinguish between superficial observation and qualitative experiences / observations.
- (ii) It is a matter of stipulation that the relevant essential properties that determine the referent of our expressions refer to either 'real nature' or qualitative properties or both.
- (iii) I propose that since 'real' and qualitative properties co-vary in the actual world, we should stipulate that they co-vary as a matter of metaphysical necessity.

- (iv) The advantage of this is that science can thus plausibly be construed as informing us of the essential properties of the world in which we inhabit and experience. If the qualitative and 'real' realms vary independently then the 'real' realm is always beyond us in principle. The notion of metaphysic necessity and possibility thus lapses into the notion of epistemic necessity and possibility; in that we could never tell which was which in principle. 'Real nature' essential properties would also be beyond the accessibility of scientists forever; as a matter of stipulated principle.

There are greater issues to do with whether direct reference provides an adequate account of metaphysical necessity or not, but I am running with Kripke's account which does indeed seem to collapse the distinction between logical (thus analytic) and metaphysical possibility and necessity. While some consider this to be acceptable others do not. While it might be possible to supplement the notion of metaphysical necessity where it is relative to a theory and the structure of the theory rules out certain experiences from occurring, I cannot argue this here. A consequence of stipulating that the 'real' and qualitative properties are essentially connected is that it is not metaphysically possible that two worlds have the same qualitative properties (observed to behave the same as the sample), but different real properties, and vice versa. Twin Earth (as a qualitative duplicate) is not metaphysically possible. This would seem relevant to an analysis of mental state terms as if it is found that mental states = a functional state of ones brain, then this would be metaphysically necessary. If it is metaphysically impossible that a real duplicate not be a qualitative duplicate then Chalmers type Zombie thought experiments would thus be ruled out as metaphysically impossible (if we accept that there are no zombies in the actual world). While I do not have the space to explore this issue further it does point to a way forward from an acceptance of dualism. While it may be objected that I have merely stipulated that zombies are disallowed I think that this stipulation is a better route towards an understanding of our world than stipulating the alternative.

References

Braddon-Mitchell, David; Jackson, Frank, (1996). *Philosophy of Mind and Cognition*, Blackwell Publishers Inc.

Burge, Tyler, (1973). 'Reference and Proper names', in *Journal of Philosophy*, 70, pp. 425-439.

Chalmers, David J., (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.

Donnellan, Keith., (1966). 'Reference and Definite Descriptions' in *Philosophical Review*, LXXV, pp. 281-304.

Dupre, J., (1981). 'Natural Kinds and Biological Taxa', in *Philosophical Review*, XC.

Frege, Gottlob, (1997). 'On Sense and Meaning' in Baillie, James, *Contemporary Analytic Philosophy*, Prentice Hall, 23-40.

Frege, Gottlob, (1997). 'Function and Concept' in Baillie, James, *Contemporary Analytic Philosophy*, Prentice Hall, 7-23.

Kaplan, David, (1989). 'Demonstratives', in Almog, J.; Perry, John; Wettstein, Howard (eds.) *Themes From Kaplan*, Oxford University Press

Kripke, Saul A., (1972). *Naming and Necessity*, Harvard University Press.

Kripke, Saul A., (1977). 'Speakers Reference and Semantic Reference', in *Midwest Studies in Philosophy*, II, 255-275.

Locke, John (1690). *An Essay Concerning Human Understanding*, reprinted in (ed.) Nidditch, P.H., Oxford University Press, book III Chapters III-VI. Book II, Chapter XXXI.

Lycan, William G., (2001). *Philosophy of Language*, Routledge.

Putnam, Hillary, (1997). 'The Meaning of "Meaning"' in *Mind, Language and Reality*, (1975), Cambridge University Press, 215-271.

Russell, Bertrand, (1997). 'Descriptions' in Baillie, James, *Contemporary*

Analytic Philosophy, Prentice Hall, 48-56.

Russell, Bertrand, (1997). 'The Philosophy of Logical Atomism' in Baillie, James, *Contemporary Analytic Philosophy*, Prentice Hall, 56-67.

Russell, Bertrand, (1905). 'On Denoting', in *Mind*, 14, pp. 479-493.

Salmon, Nathan U., (1981). *Reference and Essence*, Princeton University Press.

Searle, John R., (1992). *The Rediscovery of the Mind*, MIT Press.

Strawson, P.F., (1950). 'On Referring', in *Mind*, 59, 320-344.

Weinert, Friedel (ed.), (1995). *Laws of Nature Essays on the Philosophical, Scientific, and Historical Dimensions*, Walter de Gruyter & Co.

Quine, W.V.O., (1960). *Word and Object*, MIT Press.

Chapter 4

Armstrong's scientific realism about universals

Abstract

Armstrong argues for a scientific realist notion of universals. He maintains that 'The Final Science' will tell us what universals and particulars there are. Armstrong thus relies on science construed as the activity of discovering and reporting on the underlying reality that explains the similarity or sameness at the level of appearances. He takes the existence of universals and particulars to be an empirical matter. The objective facts are supposed to determine such things. The kind of reality that Armstrong requires is one in which (a) there is a fact of the matter as to whether universals and particulars exist and (b) there is a fact of the matter as to how many of each there are. I will attempt to show that neither claim is plausible as there are no facts of the matter that could determine the answers to these questions. Our decision rests not with facts of the matter but with considerations such as adequacy, simplicity, coherence, etc. As such the answer to the question of why things appear to be the same cannot plausibly be given a strictly realist empirical interpretation. Realism may be seen to either put the relevant facts of the matter and thus the explanation for the similarity forever beyond our grasp in principle, or to collapse back into conceptualism.

Explaining similarity or resemblance in appearances

One way in to the problem of universals is to consider that often times we judge that different things appear to be the same. While this may have a paradoxical ring to it we consider that many different things are white, and / or sweet, and / or cats[1]. The question, or problem that arises from this is: 'In virtue of what do these different things appear to be the same?'. It is useful to distinguish exact resemblance from partial similarity in that universals are called in to explain exact resemblance in the first instance (Loux, 1970, p.3). Sometimes we have exact resemblance as when we have two different things that are the exact same shade of red[2]. Most often what we have, however, is partial similarity as when we have two instances that are similar shades of red though they are not exactly the same. Partial resemblance may be explained derivatively as when instances have some of the same universals (that is to say they are exactly the same in some respects); and some different universals (which is why they are not the same in all respects). We can thus account for degrees of similarity and difference on the level of appearances.

Universals are thus distinguished from the many superficial similarities we observe. The notion seems to be that there will be relatively few universals that will explain the variety of similarities and differences observable in appearances. There has been an ongoing debate as to the ontological status of universals, most especially with respect to the question of whether there actually are any. We will now turn to a brief enumeration of trope theory, model copy realism, and Armstrongian realism in order to differentiate Armstrong's view from these alternatives.

---

[1]Throughout this essay I shall focus on properties that are most plausibly viewed as being explainable by empirical investigation in order to present Armstrong's case for scientific realism in its most plausible light. I shall thus not focus on numbers or shapes; and I will also leave relations to one side.

[2]Armstrong doesn't give us any examples of universals but I shall use the notion of the 'exact same shade of red' so as to introduce the difference between exact similarity and resemblance in a comprehensible way. We shall return to his lack of examples in a later section.

The ontological status of universals: tropes, and model / copy realism

Trope theorists consider the relevant fact to be that in the world you never see a bare particular or an uninstantiated universal (Campbell, 1990 p.479). They conclude from this that the world is composed of tropes, or states of affairs. The distinction between universals (or whatever sameness it is that is captured by our general terms) and particulars is purely formal (Campbell, 1990 p.3). It is the result of our ability to abstract one away from the other in thought. Universals and particulars, for the trope theorist are both considered to be abstractions. Neither have independent existence in the world, though they may both be considered abstract building blocks of tropes in a similar way to Wittgenstein's simples were thought to be the building blocks of states of affairs. The trope theorist maintains that while properties and relations may repeat, tropes do not it and so it is a brute fact that properties and relations repeat and things appear to be the same (Campbell, 1990 p.484).

Platonic realism distinguishes between universals and particulars and maintains that both have the same ontological status: Both exist. Particulars are to be found in the world and universals are to be found in the realm of forms where the universals that our general terms correspond to are thought to exist in their ideal or perfect state. Resemblance or similarity is construed as something more or less resembling the ideal universal in this other realm. The main problem with this model / copy realist construal of universals is that they are not to be found in the other realm; but also that they are not not to be found in this other realm.

There is also the problem as to how to judge the relation between the universal and the dim copy. In virtue of what is the instance a dim copy of that universal? If the dim copy is construed as being related to the imperfect instance then we seem to need a relation to relate the relation to the relata, and thus we seem to have launched Bradley's regress. In order to judge similarity we also seem to require some sort of access to the ideal so as to compare the instance to the ideal. Plato's doctrine of the soul's remembering them from a time before one was born seems every bit as problematic as the other realm business. Armstrong may be thought to be offering realists a way forward from realist views that are construed in such a way that universals lie beyond the scope of scientific confirmation or dis-confirmation. Or so it would seem...

Armstrong's scientific realism

Armstrong is a realist about universals (like Plato) in that he maintains that things appear to be the same because there is something about them that is, in actual fact the same (Armstrong, 1978, p.108). This seems to have a fair bit of intuitive plausibility in that when questioned this is the kind of response a typical speaker may be expected to give. This typical way of speaking commits one to the existence of both particulars and universals, and in the name of common sense this is the direction in which Armstrong travels. Armstrong maintains that both universals and particulars exist in the world and they are within the scope of scientific discovery (Armstrong, 1978, p.126). The empirical facts are thus sufficient to determine both that there are universals and particulars, and also how many there are of each. Armstrong does not give any examples of universals (or particulars) because he maintains that they are to be determined by, or read off from 'The Final Science'.

Armstrong takes care to distinguish his variety of realism from Platonic model or copy realism. The Platonic notion is that universals exist 'over and above' particulars. Armstrong maintains that universals are not over and above particulars; rather they are exhausted by their instantiations (Armstrong, 1978 p.112-113). Universals are held to 'inhere in' particulars and Armstrong maintains that inherence should not be construed as a relation between particulars and universals. Construing inherence as a relation would see realism subject to Bradley's regress again and so Armstrong may be thought to have prevented that problem.

A motivation for realism is that when we ask why things appear to be the same we can give this intuitively plausible, common-sense answer: It is because things really are the same. Armstrong considers that there is a numeric identity between different particulars instantiation of universals (Armstrong, 1978, p.111). If we consider whiteness to be a universal then the whiteness of this page of my essay is numerically identical to the whiteness on the next page of my essay. They both partake of the same whiteness and it is in virtue of this that both pages appear to be the same. Armstrong maintains that universals are exhausted by their instantiations which is the main distinction between Armstrong and Plato. Whereas Plato thought that instances partook in a dim copy of the universal that existed in its ideal in the realm of forms Armstrong maintains that there is no more to the universal than is exhausted by its instantiations.

Armstrong acknowledges that universals and particulars do not occur in the world in isolation from each other. We always find them in combination. Particulars have spatial temporal location and it is impossible for more than one particular to occupy the same position in space-time whereas universals are different in that they can occur in many places at once (e.g. on this page and on the next page) and also in that more than one universal can inhere in the same particular. (This page can have odour and taste and whiteness all together.)

Scientific realism and 'the final science'

Armstrong does not give examples of universals because he maintains that we don't know what the universals are until the final science is in. While these two pages of my essay may appear exactly the same with respect to whiteness it is possible that on some level they are not exactly (or numerically the same) because of microscopic differences in their texture which results in slightly different frequencies of light being reflected (Armstrong, 1978, p.135). Most of our predicates that refer to properties are like this (Armstrong, 1978, p 134). Our terms may apply to similar things but Armstrongian universals are required to be identical, that is to say exactly or numerically the same. The notion is that most of our common-sense terms that refer to properties don't really refer to universals because we are grouping with respect to similarity rather than them being exactly the same.

Armstrong maintains that this superficial similarity though is to be explained with respect to universals. Things appear to be the same because at some level of analysis they really are (numerically) exactly the same. It is the universality that explains the similarity. While it is hard to see why Armstrong couldn't give provisional examples of universals (on the provision that current science is the final science) so it would be easier to get at what he is saying perhaps Campbell could be thought to provide such an account. Campbell isn't a realist about universals (as such). Rather he maintains that all that exist in the world is states of affairs (or tropes) and universals and particulars are formal distinctions both (Campbell, 1990, p.535). We distinguish between them in our minds whereas in the world they always are found in concatenation. He maintains (similarly to Armstrong) that the final science will determine what tropes there are and he gives an example of five tropes (provisionally – should they make it to the final science.) He maintains that the five field forces are examples of tropes, the electro-magnetic

force, the strong nuclear force, the weak nuclear force, etc (Campbell, 1990, p.146). Perhaps this is the way that Armstrong is going but it is hard to be sure. Perhaps whiteness can't be construed as a universal because of the different shades of whiteness, but if we have one instance of whiteness and then cut that instance in half it would appear that we have two instances that are exactly the same as Armstrong would require them.

Perhaps with respect to counting universals we will end up with a very small number (which is the way Campbell does), or perhaps we will go the other way where we end up with many universals where one is distinguished from the next by a just noticeable difference. In lieu of examples from Armstrong it is hard to know which way he was thinking. Armstrong thinks that the final science will determine this issue for us however this may not actually be the case. When we consider the state of science currently there seems no way to say that science lends greater support to Campbell or to Armstrong so perhaps the final science will be the same. It is hard to see what empirical evidence could decide between Campbell and Armstrong – perhaps the issue isn't an empirical one (as they take it to be) but rather we have a formal distinction to make. Do we want to say that the difference between universals and particulars is a formal distinction or that it is given by the world? Given Armstrong's definitions of universals and particulars the evidence can be construed as showing us what particulars and universals there are. Given a Campbell definition of tropes the same evidence could also be construed as telling us what tropes there are. Which way we go seems to be not determinable by the empirical data and rather should be decided on grounds such as simplicity, coherence, etc. The answer to the problem of universals may thus be a conceptual issue rather than an empirical one as Campbell and Armstrong take it to be.

Scientific realism and conceptualist collapse

This construal of the problem seems to place the problem of universals and particulars back within the realm of human decision or choice. If what universals and particulars there are (even whether there are any universals and particulars) is construed as a theoretical decision rather than a discovery then it seems to shift the focus back from mind-independent reality to a mind dependent or conceptual reality.

Quine considers that science is about posits though 'to call a posit a posit

is not to patronise it' (Quine, 1960, p.22). Science begins with observation on the appearance level (or at the level of surface irritations for Quine) and then posits entities or processes in order to explain perceived regularities in the surface phenomenon. Armstrong and Campbell both seems to construe science as a strictly realist enterprise where science is construed as the cumulative discovery of objective truth. Though perhaps Armstrong sees that it is not cumulative (in that he isn't willing to give examples of universals based on what science has discovered so far) he relies on a strictly realist view of science to support his strictly realist view of universals.

Quine suggests that there could be an indefinite number of final sciences and there may well be no further grounds to decide between two competing theories (Quine, 1960, p.23). They may be perfectly matched for simplicity, coherence etc, or where one scores higher on one consideration it might also score lower on another and there may be no non-arbitrary way to decide between them. Armstrong's realism seems to rely on a correspondence notion of truth where what the scientists say can be either a true or false description of reality. Although Armstrong does not have to be committed to there actually being a final science (after all how would we know when it had arrived) he seems to require there being an actual fact of the matter that is enough to determine what universals and particulars there are in the world. If we cannot access that reality directly then it seems that if we maintain that reality it self decides what universals and particulars there are then we can never know whether we have access to that final science or not because we cannot access the reality side of the correspondence relation.

Because of this impossibility in principle of our accessing the required information to determine what universals and particulars there are it seems that we have a notion of science that it cannot possibly hope to meet in principle. If we are concerned with the reality behind the appearances and we cannot hope to access that reality directly then what we are left with is our theories or conceptions of that reality. Our theory posits an underlying reality in order to explain the surface appearances and regularities. It would seem that Armstrongian universals are such posits they are posited to explain the appearances and the similarity we find in the appearances. There may well be more than one satisfactory explanation for the surface regularities though, and in this case either the answer to the problem of universals is beyond our reach in principle or we will have to consider that the answer can only be decided on conceptual and theoretic grounds.

Kuhn, (1970) by way of examples in the history of science showed us

that science is not a cumulative endeavour rather there is a theoretic shift in world view, our theories radically change and we perceive this new theoretic ontology in a gestalt fashion. Though it seems that we can never decide between the truth of two competing paradigms with the world directly and thus determine which is true or false with respect to which one corresponds with reality we may be able to compare them with respect to simplicity, coherence, adequacy (for explaining the most superficial regularities and differences) etc.

Quine maintains that despite these we may still end up with two different theories that were different though there were no further grounds for deciding between the two theories (Quine, 1960, p.23). Reality cannot determine it because we can never access that reality and what we are trying to do, at any rate, is to construct that reality as an adequate (simple etc) explanation of the surface phenomenon. As such science can be construed not as a realist enterprise that is concerned with a reality that it can never access in principle, rather it can be construed as a search for the best postulated underlying reality to help us make sense of, explain and predict surface regularities including why it is that things appear to be the same.

Determining universals from no universals from tropes

Campbell descends to the level of physics in search of tropes and thus Campbell also shows that he has a reductionist view of science where the physical level determines all the rest. We can see that universals are called in to explain as they arise in answer to the question 'in virtue of what do things appear to be the same' and so we may also see that it is this similarity or sameness on the appearance level that drives us to search for something to explain the sameness.

Even if we take the state of science now the empirical evidence does not determine between Campbell and Armstrong. Even if the final science arrived (though we could never know that it had in principle) it wouldn't seem to be capable of distinguishing between Campbell and Armstrong in principle. It does seem likely though that 'The final science' is a realist pre-Kuhn myth and that there could be an indefinite number of final sciences that take into account all the nerve hits of mankind in the past present and future. In this case it seems that the different final sciences could differ in that they posit different ontologies. In so far as science can distinguish what universals

and particulars there are (and it seems that it cannot) we may end up with radically different particulars and universals depending on which version of the final science we accept.

Realism thus seems to collapse back into conceptualism where we cannot say that things appear to be the same because they are the same, but instead we must say that we construe things really being the same in virtue of them appearing to be the same. Our scientific theories are constructed to explain and predict on the observation level and the theoretic entities that science provides are posits postulated in order to predict and thus explain the appearances. It thus seems backwards to explain the appearances by the reality when the reality is reached by being inferred from the appearances however this is an issue with science in general and is not restricted to the problem of universals and particulars. Armstrong succeeds in answering the question in virtue of what do things appear to be the same but he uses an inference from the superficial sameness in order to explain it.

Grounds for deciding

Despite this Armstrong's ontology is no larger than a theorist who maintains that there exist universals but no particulars. Because he maintains that the universals exist in the particulars and there is nothing over and above about them he is not multiplying entities in maintaining that universals are real. We may well wonder then what the ontological difference is between Armstrong and the trope theorists and it seems that the distinction may be more verbal or formal than real (so to speak). It is hard to try to map universals and particulars onto tropes to know who ends up with more stuff at the end of the day but if each theory can use the same empirical evidence to support its view and state what universals, particulars, or tropes there are then at the end of the day they would seem to empirically say the same thing. If the distinction between tropes and universals and particulars is not determinable by all available empirical data then this seems to point back to universals, particulars, and tropes being squarely within the realm of human conceptual decision. We must decide which way we want to describe reality, and both theories would seem to be fairly equal descriptions of the same empirical reality. Both fall down on the same point, however; they fail to acknowledge that the problem of universals is a conceptual rather than empirical issue.

An initial intuitive appeal of realism is that we do seem to want to say

on a first pass that things appear to be the same because they are the same and this fact about the way the world really is seems to explain this adequately. Unfortunately this intuitive credibility is severely challenged when we consider whether science is best construed as a strictly realist enterprise or whether we actually are entitled to say anything about that reality. I am reminded of Wittgenstein when he said 'whereof one cannot speak thereof one must be silent" and think that this might well be the problem that Armstrong implicitly acknowledges in his refusal to give examples of instances of universals.

We can insist on maintaining that there is a fact of the matter as to what the real world is like. Perhaps Armstrong is silent on universals because he realises that we are never entitled to say what that world is really like (which may be why he may in fact not hold out much hope that the final science will arrive). However, we may consider that even if the final science arrived and we were assured by God himself that our scientific posits were true as they corresponded to reality then we would still be no further ahead as to whether there are universals and particulars or tropes; whether there should be one universal or trope for every just noticeable difference that is repeated; or whether we can go more global and maintain that one field of force is as far as we need to distinguish.

So what universals and particulars there are may well be indeterminate. Whether there are universals or not may be indeterminate. There seems to be no fact of the matter that could decide between them. The issue seems to me not to be one that is undecided because of the lack of empirical evidence but rather because of a decision we must make on grounds of simplicity, adequacy, coherence, etc. Aesthetics may also feature where Armstrong initially appears to be at an advantage over the other theories because of the intuitive beauty that we find in 'things appear to be the same because they are the same'. Realism seems to come out slightly ahead as a realist view of science is popular and we like to think of science discovering the facts about the real world. We have seen thought that Campbell is really on a par with Armstrong here regarding an intuitively pleasing realist view of science but perhaps his notion of tropes or states of affairs is less pleasing to Armstrong's. The adequacy of either view, however, seems threatened by their construal of the problem of universals as an empirical rather than conceptual issue.

Wittgenstein states that 'explanation must stop somewhere' and whether it stops at the 'things appear to be the same level' or at the 'things are the same level' may be seen to be neither of them prior to the other because of the

way in which we come to 'know' of that reality. If we could access it directly that would of course be another matter – but perhaps the only reality that we have this kind of access to is what we have chosen to dub 'appearances' and thus that is what we try to seek to explain. They seem to be two sides to the same coin to me, and because we must accept that the only reality we can access, and thus the only reality we are capable of sustained debate and description of is rather an inter-subjective theoretical construction. This being said which is prior, the appearances or the reality are really both two sides of the same thing though we must take appearances as prior. Things appear to be the same is basic, and for us to ask why and thus postulate a 'reality and use it to explain' seems to be a further epicycle tagged onto the initial problem that doesn't help to explain it any more simply.

References

Armstrong, D.M., (1778). *Universals and Scientific Realism* (Volumes I and II), Cambridge University Press.

Armstrong, D.M., (1980). "Against "Ostrich Nominalism": A Reply to Michael Devitt', in *Pacific Philosophical Quarterly*, 61, 1980, pp. 440-449.

Armstrong, D.M., (1984). 'Replies to Aune', in Bogdan, Radu J, (eds.) *D.M. Armstrong*, Dordrecht, Holland; Boston: D. Reidel Pub. Co., pp. 250-256.

Aune, Bruce, (1984). 'Armstrong on Universals and Particulars', in Bogdan, Radu J, (eds.) *D.M. Armstrong*, Dordrecht, Holland; Boston: D. Reidel Pub. Co., pp. 161-169.

Baxter, Donald L.M., (2001). ' Instantiation as Partial Identity', in *Australasian Journal of Philosophy*, Vol. 79, No 4, pp. 449-464.

Bogdan, Radu J, (eds.) (1984). *D.M. Armstrong*, Dordrecht, Holland; Boston: D. Reidel Pub. Co.

Bradley, Michael, (1979). 'Critical Notice of Universals and Scientific Realism', in *Australasian Journal of Philosophy*, 57, pp. 350-358.

Campbell, Keith, (1990). 'The Problem of Universals', Chapter Two in, *Abstract Particulars*, Oxford, UK ; Cambridge, Mass., USA : B. Blackwell.

Campbell, Keith, (1990). 'Fields: Dealing with the Boundary Problem', Chapter Six in, *Abstract Particulars*, Oxford, UK; Cambridge, Mass., USA: B. Blackwell.

Devitt, Michael, (1980). '"Ostrich Nominalism" or "Mirage Realism"?', in *Pacific Philosophical Quarterly*, 61, pp. 433-439.

Kuhn, T. S., (1970). *The Structure of Scientific Revolutions*, University of Chicago Press.

Loux, Michael J., (1970). *Universals and Particulars: Readings in Ontology*,

Anchor Books.

Moreland, J.P., (2001). *Universals*, Acumen.

Williams, Donald, C., (1986). 'Universals and Existents', in, *Australasian Journal of Philosophy*, 64, No.1.

Quine, W.V.O., (1980). Soft Impeachment Disowned' in, *Pacific Philosophical Quarterly*, 61. pp. 450-451.

Chapter 5

Wilkerson on natural kinds

Abstract

Wilkerson attempts to distinguish between four kinds of kinds. While he maintains that metaphysical realism presupposes the existence of natural kinds, he considers that they are not to be found in the social sciences. This conclusion seems to rest on his construal of the essential properties that are supposed to determine membership in a natural kind. If the kinds of essential properties that Wilkerson requires are not forthcoming then it seems that we are faced with a decision: Either there are no natural kinds in the way that Wilkerson characterises them, or we must alter our conception of what is necessary and sufficient for natural kind membership. If we are led to such a predicament then it would seem that whether there are natural kinds or not is not only something to be determined by the world in itself as Wilkerson takes it to be, but is also the result of our analytic decision.

Wilkerson's characterisation of natural kinds

In the course of offering an account of natural kinds, Wilkerson distinguishes between four types (or kinds) of kinds:

- (1) **Natural kinds**, which are characterised by real essences, by intrinsic properties that make the individuals or stuffs the kinds of things they are, and which lend themselves to detailed scientific investigation, e.g., the kinds *electron, proton, neutron; carbon, water, cellulose; chimpanzee, stickleback, narcissus.*

- (2) **Dependent kinds**, whose members are what they are because of a relational dependence upon something else, e.g., the kinds *table, coin, lintel, threshold, cliff, glacier, north wind, perennial, annual.*

- (3) **Real but superficial kinds**, which are characterised by real, non-relational but comparatively superficial similarities and differences between things, similarities that do not lend themselves to detailed scientific investigation, e.g., the kinds *tree, shrub, cloud, pebble, honey water.*

- (4) **Hybrid kinds**, especially hybrids of (1) and (2), e.g., the kinds *vegetable, fruit, pot plant, cattle, medicine, ham, pork, bacon*; and hybrids of (2) and (3), e.g., the kinds *ski slope, surfing beach, gravelpit, oasis, biennial.* (Wilkerson, 1995 p.59)

His examples of natural kinds are those taken from the fields of physics (*electron, proton, neutron*); chemistry (*carbon, water, cellulose*); and biology (*chimpanzee, stickleback, narcissus*). Wilkerson, (1995 p.73-87) considers that while there is nothing that logically rules out the objects of the social sciences as natural kinds; they are best classified as (2) or (3). I will go on to consider the plausibility of there being a difference in kind between (1), (2), and (3).

I shall attempt to argue that while there may be a difference in degree between the 'hard' and social sciences, there is not the difference in kind that Wilkerson takes there to be. As such it would seem that natural kinds may not exist in quite the way that Wilkerson takes them to; or that, alternatively, there cannot be any natural kinds that we can have knowledge of. Either way requires a revision of Wilkerson's account. Whether there are natural kinds or not would seem to be not only determined by the world, but also a matter of analytic decision as to the precise nature of the essential properties that are required to determine natural kind membership.

Foundations of metaphysical realism

Firstly, we will need to begin by examining some of the foundations on which Wilkerson builds his account of natural kinds. He considers that he will presuppose metaphysical realism (Wilkerson, 1995 p.30). Wilkerson (1995, p.29)

takes metaphysical realism to be the position that 'there is a distinction between reality or 'nature' on the one hand, and our beliefs and theories about it on the other'. He considers that metaphysical realists are committed to the existence of natural kinds (1995, p.29).

While I shall not consider his two arguments for this, he does seem correct to note that metaphysical realists are committed to the existence of natural kinds[1]. If they maintain that objects, laws of nature, and metaphysical necessity and possibility exist mind independently, or objectively then on Wilkerson's account of what natural kinds are, they presuppose the existence of natural kinds. Essential properties are required in order for the object to count as the *same* object; they are required in order to delineate the class (or kind) of things relevant to the law; and they are required in order to determine metaphysical possibility and necessity.

Wilkerson makes the following claims regarding essential properties;

- (i) Essential properties exist mind-independently and intrinsically in objects. They are necessary and sufficient for membership in a natural kind. (Wilkerson, 1995 p.33.)
- (ii) Laws of nature exist mind-independently and govern objects behaviour in virtue of the objects having essential properties that determine the law (Wilkerson, 1995 p.62).
- (iii) Essential properties are responsible for the superficial similarities in appearances that we observe (Wilkerson, 1995 p.42, 55).

Kripke, (1972); Putnam, (1975); Braddon-Mitchell & Jackson, (1996); and Chalmers, (1996) may be considered metaphysical realists in that they maintain that the referent of a natural kind term is a natural kind that exists as such in virtue of one or more essential properties. Their variety of metaphysical realism is different to Wilkerson's in that they consider that it is metaphysically possible for a substance qualitatively identical to water to have a different real essence, and that it is metaphysically possible for H_2O

[1]I will not consider his arguments in any great depth because they seem to me to be flawed though it will not assist my argument to critique them. His observation that metaphysical realists presuppose the existence of essential properties seems correct in that they consider essential properties to exist in the world and acknowledge that these determine natural kind membership.

(as the real essence of water) to appear black and tarry (Braddon- Mitchell & Jackson, (1996). They consider that it is metaphysically possible for essential properties and appearances to vary independently of one another, and thus would not seem committed to (iii).

Locke maintained that knowledge of real essence is unattainable as we are 'destitute of the faculties to attain it' (Locke, 1993 II, xx iii). The usual Kripke / Putnam variety of metaphysical realism allows the nominal and real realms to vary independently of one another, and thus it seems hard to see how they can maintain that science can inform us of real essences[2]. It seems that the following claims are hard to reconcile;

- (a) Real essences may vary independently from nominal essence.
- (b) Real essences are knowable to us.

It also seems hard to credit real essences as being explanatory if they are unknowable. Wilkerson considers that 'anti-realists would be right to be sceptical about a conception of reality that was, in principle, wholly inaccessible to any kind of scientific investigation' (Wilkerson, 1995 p.66). The Kripke / Putnam view does consider that part of the scientific enterprise is the discovery of real essences, and thus Wilkerson offers us a plausible account of what would need to be required in order for scientists to have any chance of discovering them.

Wilkerson, (1995, p.55) avoids this unsatisfactory conclusion by maintaining that essential properties are of interest to us precisely because 'The real essence of a thing not only determines its proper de re classification... but also directly explains many of its properties... It is precisely because gold has the atomic number 79 that in normal atmospheric conditions it is malleable, fusible, soft, and heavy'. Locke maintained that real essences produced nominal essences and thus real essences explain the nominal essence; and so Wilkerson may be thought to be a Lockean regarding the explanatory power of real essences.

Wilkerson thus offers a rather plausible version of metaphysical realism. He maintains (1995, p.66) that metaphysical realism entails the existence of natural kinds, and that

[2]I may be extending how Locke's notion of the 'nominal' realm is usually interpreted in considering it to be the whole of appearances or everything that we can observe and experience of the world.

> the commitment to natural kinds is a commitment to certain real essences that permit scientific generalisation, and that scientific generalisation consists in the articulation of de re necessary truths about the causal powers of things... scientific generalisation is over causal powers, some of which are constituted or realised by the real essences that determine membership of natural kinds. So the properties that determine membership of natural kinds are precisely the properties that determine in general how individual members must behave in such and such circumstances (Wilkerson, 1995 p.62).

He thus considers that real essences determine nominal essences and that intrinsic essential properties determine the behaviour of the objects according to the laws of nature. Science is thus plausibly construed as the enterprise of discovering, and informing us about real essences (among other things).

Natural Kinds in the Social Sciences

While Wilkerson considers a variety of reasons why kinds in the social sciences do not constitute natural kinds I shall focus on three of his main objections. The first is that he considers that;

> many kinds in the social sciences are clearly dependent kinds, that is, membership of the kind is determined relationally by something else. In some cases it depends upon human laws, conventions, interests, moral attitudes, etc... Had the conventions and interests been different, there would not have been the same nations, clubs, banks etc (Wilkerson, 1995 p.79).

One could consider that biological kinds may be similar in that if their interests had been different; if, say they were not interested in surviving and reproducing, then they would not exist either. It would seem, though that Wilkerson's point is more that social kinds logically depend on interests whereas biological kinds do not. This point rests on the notion that the essential properties that determine natural kind membership are intrinsic and not relational.

He considers that 'since causal powers are constituted or realised by intrinsic properties, it follow[s] that the real essences of natural kinds would be intrinsic properties' (Wilkerson, 1995 p.61).

> The basic furniture of the world consists of 'powerful particulars', which, given their intrinsic features, necessarily change and develop in certain ways and not others, the first state of affairs must in context generate the second. Given the intrinsic features of hydrogen, any application of a naked flame to the hydrogen in the presence of oxygen must produce ignition and water vapour. Hydrogen must ignite and produce water under certain conditions; if, per impossible, it did not, it would not be hydrogen (Wilkerson, 1995, p.68).

He maintains that natural necessity is best construed as either (a) a property of essential properties, or (b) a property of true scientific generalisations at to how the object must behave (Wilkerson, 1995 p.72). The notion would thus seem to be that the intrinsic properties of the objects determine the laws of nature without remainder. It is not as though the world consists of objects with their essential properties and that the laws of nature exist as a superimposed extra; rather the laws are determined by the objects intrinsic essential properties.

He acknowledges that sometimes physics characterises objects in terms of relational properties; 'An electron is a particle with a negative charge, which orbits the nucleus of an atom; an acid is a proton donor; a gene is a complex molecule which governs the properties of the phenotype; and so on' (Wilkerson, 1995 p.32). But he then maintains that it should not do so as 'if we fail to produce a story about intrinsic properties, we are left with a mystery, with the unexplained brute fact that various objects are related in various ways' (Wilkerson, 1995, pp.32, 33).

Problems with the intrinsic / relational distinction

It seems to me that in considering intrinsic properties to be something over and above a complex bundle of relational properties Wilkerson is faced with the following, difficulties:

- (a) Considering that various objects have essentially non-relational properties that determine how the object will interact with the world and appear to us would seem to be a contradiction[3].
- (b) It is hard to see how we could discover or come to know of an intrinsic property that is not essentially relational.

If objects do not essentially have relational properties then relational properties would seem to be irrelevant for natural kind membership. It would thus seem that the essential properties that are relevant for determining natural kind membership are intrinsic, non-relational, thus can leave no mark on the world; and are thus unknowable in principle. Wilkerson would seem to be led to the same problem that Locke, Kripke, Chalmers, Braddon-Mitchell & Jackson, and Putnam's variety of metaphysical realism faced. He would need to sacrifice (iii) (p.3 above) and be left with the unsatisfactory conclusion that we cannot know real essences. I think that Wilkerson's story would be more plausible if he acknowledged that the essential properties that determine natural kind membership are essentially relational. Different natural kinds would thus be distinguished in virtue of engaging in distinctively different relations with other objects. This latter line has the consequence that it does not count against kinds in the social sciences that their essential properties are relational.

He considers that if we are externalists about beliefs and consider that beliefs represent in virtue of having the appropriate causal connections with the objects of the kind represented then 'It follows that psychological representations, such as beliefs, thoughts, and intensions, are determined, not by the intrinsic features of the people concerned, but by a complex relation between them and the rest of the world' (Wilkerson, 1995 p.79). If we consider Newtonian mechanics then the relevant essential properties are mass, force, velocity etc. While each term may be thought to pick out an intrinsic essential property of the object, the terms are defined relationally or functionally with respect to how each property interacts with another to produce behaviour that scientists may observe. I fail to see how this example from

[3]I can see that perhaps the main motivation for considering the relevant essential properties to be intrinsic and to unfold in such a way as to produce the behaviour and appearances that we observe is that an object could be the same object even if everything else in the universe ceased to exist. I still consider, though, that such a property is either relational (in the sense of being *causal* or correctly defined as *functional*) or else unknowable and as explanatory as Locke's 'pin cushion' model of substance.

physics is different in kind from Wilkerson's example of beliefs, thoughts, and intensions. While it may be possible to re-describe relevant properties in a way that is non-relational it is hard to see how essentially non-relational properties can produce effects on the world or be the objects of scientific investigation.

Multiple Realisability

Another of Wilkerson's objections to natural kinds in the social sciences is that kinds in social sciences are multiply realised.

> There can be no science of psychology, economics, politics or sociology, no confident generalisation or prediction at the high level of function... The possibility of multiple realisability is likely to undermine all but the roughest and least ambitious explanatory remarks. Success is possible only in cases where all the realisations are fundamentally rather similar... But then the success of the high level sciences (e.g., the social sciences) depends entirely on the explanatory success of those at the lowest levels, those dealing with the realisation (e.g., physics, chemistry, biology)' (Wilkerson, 1995, pp.85-86).

He also considers, though, that at one level of explanation cancer might be construed as a natural kind, whereas at a lower level there might be different kinds of cancer (Wilkerson, 1995, p.75). He considers, though, that the success of biology is that its kinds seem to be realised by kinds of chemicals, and that the success of chemistry sees its kinds realised as kinds of particles. He also maintains that biology cannot be reduced to chemistry and that chemistry cannot be reduced to physics 'for all sorts of familiar reasons' (Wilkerson, 1995, p.86).

Wilkerson considers chemical isotopes which have the same atomic number but a different atomic weight and he concludes that he has

> constantly referred to different explanatory levels, and have quite deliberately left open the possibility that two objects might belong to the same kind at a higher level and to different kinds at a lower level. Indeed, if the distinction between function and realisation is ever to have application to members of natural kinds,

> we would be foolish not to leave that possibility open (Wilkerson, 1995, p.110).

Wilkerson thus seems to vacillate between considering that biology and chemistry are successful only because they are realised by physics, and maintaining that each level supports legitimate natural kinds that cannot be reductively explained because they are multiply realised from the perspective of the lower level. He considers that multiple realisability on the chemical level does not count against chemical kinds as natural kinds. I do not see that there is a difference in kind between multiply realised chemical kinds and multiply realised psychological kinds.

It would seem that natural kinds are determined by real essences but we need to come to a decision as to what real essences are relevant for the kind that we are interested in. Chimpanzees, gold, water can all realise Newton's properties of mass and velocity but in a sense the essential properties are not multiply realised because the instantiation is irrelevant. Likewise, we have decided that isotopes are only irrelevantly different on the chemical level and thus the atomic weight is irrelevant with respect to chemical kinds. He considers that we can 'quite consistently lump with the chemists and split with the physicists (Wilkerson, 1995 p.110)' and I think that a similar case can be made for lumping with the psychologists or social scientists and splitting with the biologists, and / or chemists, and / or physicists.

Scientific generalisability and boundary conditions

Wilkerson considers that the social sciences to not support scientific generalisations. Interestingly, he also considers that agriculture, horticulture, geology, geography and meteorology are in the same boat.

> Indeed the conspicuous success of statistical methods in say, meteorology depends on the constancy and stability of terrestrial conditions. It would be impossible to use terrestrial agriculture, horticulture, geology, and geography and meteorology to explain and predict the behaviour of objects whose constitution and local conditions were very different from those on earth. (Wilkerson, 1995, p.84).

What he fails to acknowledge, though, is that the essential properties utilised in the hard sciences are also restricted to boundary conditions or qualified by a *ceteris paribus* clause. Newtonian mechanics is largely considered correct although the boundary conditions are more restrictive than Newton envisaged. Physics relativises its essential properties and laws to systems, which is a delineation of boundary conditions.

While Weinert distinguishes between *phenomenological laws* that apply to predict and explain observable features and are restricted to boundary conditions, and *fundamental laws* that are not restricted to boundary conditions but instead predict and explain the boundaried laws, Wilkerson makes no distinction (Weinert, 1995, 49- 51). The phenomenological laws are intrinsic to objects but will only produce their usual effects if certain external boundary conditions obtain. They thus would seem to depend on external conditions for their realisation. The fundamental laws that predict and explain the way in which phenomenological laws interact with boundary conditions are functional laws that describe the relationship between the 'intrinsic' properties / laws and extrinsic boundary conditions and are themselves not restricted to any particular boundary conditions, or particular laws.

While this account of physical laws / essential properties may be controversial I do think that there is more a difference in degree rather than kind between the restrictions that apply to, and the scope of the generalisations that are legitimately made from, the natural and social sciences. Wilkerson considers that

> I can presumably predict that trees will be blown over by gales, will die from drought, will undermine my neighbours foundations and will plunge his garden into shade without knowing their species (Wilkerson, 1995., p.56).

There do seem to be rational predictions and generalisations that we can make from all of the kinds of kinds that Wilkerson acknowledges. While he seems to consider that there is a sharp divide between the social and natural sciences, though, it would seem more plausible to consider that there is rather a difference in degree. While this is controversial, Dennett (1998) considers that use of the intentional stance (in this context considering beliefs and desires to be natural kinds), gives us 'predictive leverage we can get by no other method'. We can predict that Mary will go to a shop to get some toilet paper by noticing that she is out of toilet paper and thus will soon come to believe she is out and will desire to get some more. Such a

prediction would not seem to make sense at any lower level of explanation as no lower level can capture the relevant kinds of belief, desire, or indeed the notion of a shop, or toilet paper.

Wilkerson offers his position as an alternative to the explanatory liberal (who is not a metaphysical realist but considers that there are natural kinds).

> they were prepared to countenance any entity that fulfilled certain very general conditions, and any entity that appeared in a serious descriptive and explanatory discipline. The very general conditions were that the entity should have a clear criterion of identity, that it should have a certain structural unity, and that it should lend itself to description and explanation in terms of relevant scientific generalisations' (Wilkerson, 1995 p.29).

Wilkerson maintains that this is not acceptable, but his alternative account of when a scientist is entitled to say that they have discovered an essential property relevant for determining a kind seems quite similar;

> we always have some sort of a guarantee that our scientific theories are true, if they obey the general constraints on rational acceptability – if, for example, they are consistent with observation, are comparatively simple and mathematically elegant, yield true predictions, generally get us from truth to truth, and minimise inexplicable coincidence. Not only can we be confident that a theory that passes such tests records natural necessities, but it would be absurd for us to ask for any further guarantee... The possibility of scientific mistake is cause for congratulation, not complaint. (Wilkerson, 1995, p.69)

This, though does not seem to be an adequate account as to when a scientist may consider they have discovered the objective, mind independent essential property that determines natural kind membership. If there is always the possibility that the scientist is wrong then it would seem that we can never have knowledge of these essential properties. If essential properties are essentially intrinsic then it would seem that we cannot observe them and thus intrinsic essential properties are explanatory posits and we have no way of determining whether they mirror reality or not as we can never access the reality itself. The criterion that we have for determining whether the posits are acceptable or not can thus not be whether they correspond to reality or not; rather we are left with the explanatory liberals' account of adequacy,

simplicity, coherence etc.

Wilkerson's metaphysical realism thus seems to lapse back into explanatory liberalism with respect to the practice of scientific investigation and the criterion that we have for accepting essential properties that determine natural kind membership would seem to be the same for both Wilkerson and the explanatory liberal. Wilkerson cannot maintain that the essential properties that determine natural kind membership are both essentially intrinsic and accessible to scientific investigation. As such it would seem that we have an analytic decision to make with respect to whether there are natural kinds or not. Either this is a matter that is to be determined by the world in itself and thus is beyond us in principle; or essential properties are posits that explain and predict nominal similarities and natural kinds are determined by the success we have with predicting and explaining phenomena. If the latter is the case then it would seem that the differences between Wilkerson's kinds of kinds is a matter of degree. If we wish to distinguish between them categorically then the distinction is determined by a relational property, the *comparative* predictive and explanatory success of our posits.

References

Braddon-Mitchell, David; Jackson, Frank, (1996). *Philosophy of Mind and Cognition*, Blackwell Publishers Inc.

Chalmers, David J., (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.

Dennett, Daniel C., (1998). *Brainchildren: Essays on Designing Minds*, Penguin Books, Great Britain.

Dupre, J., (1981). 'Natural Kinds and Biological Taxa', *Philosophical Review*, vol. XC.

Elder, C.L., (1989). 'Realism, Naturalism, and Culturally Generated Kinds' *Philosophical Quarterly*, vol. 39.

Goodman, N., (1965). *Fact, Fiction, and Forecast* Bobbs-Merrill.

Kripke, Saul A., (1972). *Naming and Necessity*, Harvard University Press.

Locke, J., (1993). *An Essay Concerning Human Understanding*, The Guernsey Press.

Lycan, William G., (2001). *Philosophy of Language*, Routledge.

Putnam, H., 'Is Semantics Possible?' (1975). *Mind, Language, and Reality*, Cambridge University Press.

Putnam, H., (1975). 'The Meaning of Meaning' *Mind, Language, and Reality*, Cambridge University Press.

Quine, W.V.O., (1969). 'Natural Kinds' in *Ontological Relativity and Other Essays* Columbia University Press.

Salmon, Nathan U., (1981). *Reference and Essence*, Princeton University Press.

Weinert, Friedel (ed.), (1995). *Laws of Nature Essays on the Philosophical, Scientific, and Historical Dimensions*, Walter de Gruyter & Co.

Wilkerson, T.E., (1995). *Natural Kinds*, Avebury.

Wittgenstein, L., (2001). *Philosophical Investigations*, Blackwell Publishers Ltd.

Chapter 6

Theory structure and the explanation of natural necessity

Abstract

The kind of necessity that has been variously known as nomic, physical, and natural is a puzzling notion in the philosophy of science. While the empiricists considered that necessity was something that could be empirically discovered, Hume (1739-40) countered that we can never observe necessity or the causal nexus that compels one event to follow another. Since then there has been much debate and controversy as to whether necessity is a mind-independent (extensional) fact to be discovered by scientific observation and generalisation, or whether Kuhn (1962) and others have shown that observation is theory-laden and that science is governed by (intensional) constructs. Hung (forthcoming) has recently offered an explanation of physical necessity, where necessity is considered to be relative to a theory, or cross-theoretic. While conceptual spaces or theoretical structures are constructed by us, they are intended to be an adequate space for the modelling of nature. The structure of the theory is thought to restrict the range of possible experiences that we can encounter in the natural world. My interpretation of Hung is that Natural necessity is thus governed by the world, and explained by our representations of it.

Universal claims by way of generalisation

I will begin with a variation on a parable that Hung, (forthcoming, p.8) utilises to illustrate the appearances / reality distinction, and to introduce the

notion of conceptual shift. I shall use it to illustrate alternative conceptions of the scientific enterprise that have been held by various individuals and schools of thought through the history of science; and ultimately to demonstrate the cross-theoretic nature of natural necessity.

> There are a group of children (A) who inhabit a very large room. The room is divided in half by two rows of iron bars. The children are prevented from passing through the first row of bars. Each child discovers that behind the second row of bars there is a counter-part child (B) that looks just like him or her[1].

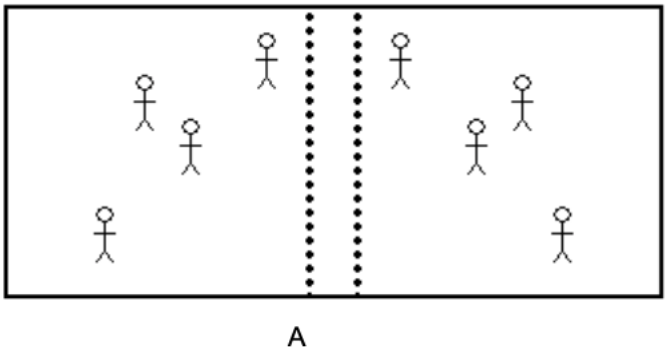

Figure 1.

Some of these children, being scientifically inclined, observe the movements of the counter-parts closely. They observe that counter-part n_1 mimics at t_1, t_2, t_3, t_4 ... and they generalise, or infer from these observations that a counter-part (counter-part n_1) always mimics. Counter-parts n_2, n_3, n_4 ... are also observed to mimic the movements of each relevant child, and so the children generalise from these observations to the following: all counter-parts mimic. By conjoining these observations they reach the following universal claim by way of generalisation:

- (a) All counter-parts always mimic.

[1]The term counter-part is not intended in David Lewis' sense. 'Counter-parts' are simply the children who are behind the second row of bars.

The empiricists: Necessity by way of universal generalisation

Empiricists such as Bacon (1561-1626) and Mill (1806-1873) considered science to progress from observation of phenomena to generalisations, as illustrated by the way in which the children arrive at (a)[2]. It was thought that further progress was made in virtue of the generalisation from some finite number of observed instances, capturing a law of nature. Hempel (1948, in Brody, 1970 p.11) and Popper (in Caws, 1965 p.180) considered laws of nature to play a special role in the logic of scientific explanations. The classical view of science presented the logic of explanations as being of Deductive-Nomological, or D-N form. The explanandum is explained in virtue of its being deductively implied by one or more initial conditions, together with one or more laws of nature. This is known as the *covering law thesis*.

Laws of nature come to play a special role in scientific explanation as if it is accepted that the initial condition(s) obtain and that the law of nature is naturally necessary,then it follows that the explanandum inherits the natural necessity of the law and thus could not have been otherwise. Let us suppose that one of the children asks why a particular counter-part mimics on a particular occasion, and the scientists offer the following explanation:

- (i) *initial condition* This is a counter-part.
- (ii) *law of nature* Counter-parts always mimic.
- (iii) *explanandum* This counter-part mimics on this occasion.

Given the initial condition and the necessity of the law the explanation is considered to be satisfactory. Laws of nature would thus seem to be required to compel the explanandum (in order to explain that it could have not been otherwise), and as such we have an explanation as to why the explanandum is naturally necessary. This characterisation of natural necessity would seem to leave us with the following question, however: Why do counter-parts always mimic?

[2]Bacon and Mill considered that science progressed by induction by analogy and simple enumeration; and they systematised five other methods of induction: agreement, difference, agreement and difference, concomitant variation, and residues (e.g., Mill, 1952, pp.253-264).

Hung, (forthcoming, pp.83-84) considers that a 'nominal' sort of necessity can be reached by arbitrary stipulation.

> Why is it that all P's are Q's? Because it can't be otherwise. Why can't it be otherwise? Because... If no answer can be provided for the second question, the explanation is only explanation in name. It is a nominal explanation. Criterion: An explanation of E is a nominal explanation if its explanans amounts to a statement of the form: 'Necessarily E,' where the claim of necessity is left unsupported with further reasons.

Hung (forthcoming, pp. 83-84) relates that some theorists (e.g., Burks) have considered that by prefacing the law with '©', or another operator to signify natural necessity the necessity is carried through to the explanandum. While some theorists (that we shall go on to consider) maintain that laws are nothing other then true generalisations (and thus there need be no *compulsion*) between associated phenomena this cannot be considered an explanation as to *why* the law and the conclusion are necessary. Could the children not stipulate the necessity of the law that counter-parts always mimic in the same fashion, by prefacing *(iii)* with '©'? If we do not have an explanation as to why some statements reached by generalisation are laws whereas others are not then how could we tell whether any given generalisation is naturally necessary or not[3]? While Wittgenstein (2001, p.35) considered that explanation has to stop somewhere, to stop the explanation at this point is to not even get an explanation up off the ground. To characterise natural necessity as nominal should be a last resort strategy, one whose adoption signifies that one has given up on an explanation of natural necessity.

A problem that would seem to arise with 'laws of nature' reached by way of empirical generalisation is that as explanations they are circular, or, *ad hoc* (Hempel, 1966, p.28). We observe instances and we abstract or generalise from these instances to general 'laws'. To then use this abstraction or generalisation to explain those very same instances would be explanatorily circular, as the law was inferred from those instances[4]. With respect to explanation we would need an account of the necessity of the law. In virtue of

[3]Hempel, (1966, p. 55) states that he needs to 'consider the explanation of laws by theories', which is something that I shall go on to do.

[4]The radical behaviourists consider that this is why mental states (such as mimicking dispositions) are unacceptable as explanations of behaviour (Baum, 1994 pp.33-35). They are considered to be inferred from instances of behaviour and so to use them to

what are such generalisations as 'copper conducts heat' naturally necessary, while 'counter-parts always mimic', is intuitively contingent? We shall come back to this.

Necessity: scientific realism and the empirical discovery of necessity

Direct reference has increased in popularity as an account of how some of our linguistic expressions succeed in denoting objects in the world. Kripke, (1972, pp.120-127), Putnam, (1975), and other direct reference theorists presuppose a realist view of science for their account of reference. It is considered that the reference of some of our expressions is determined by an object, substance, or kinds 'real nature'. The real nature consists in essential properties that are to be determined by a-posteriori scientific investigation (Kripke, 1972, pp.122-127). We may consider the view of science that this picture encourages:

- (a) There is a mind independent world consisting of objects with their essential properties that are governed by laws of nature.
- (b) The business of science is to a-posteriori (or empirically) discover these essential properties and laws of nature.
- (c) There are facts of the matter about essential properties and laws of nature. Scientists may be right or wrong about them (derived from Kripke, 1972 and Salmon, 1981).

Kripke and Putnam consider that natural kind terms (such as 'water' and 'gold') have their essential properties fixed by the world. Kripke considers that we have an initial baptism of a sufficient sample of water, and thereafter the sample is fixed by the essential properties of that sample (Kripke, pp. 135-140). Kripke and Putnam maintain that something that does not share the essential property does not count as water (or the same kind of stuff) even though they both consider it to be metaphysically possible for H_2O (or

explain those same instances is circular as an explanation. I shall consider Comte's call for the abolition of 'metaphysics' (theoretical terms) from science and the prospects for operationalising theoretical terms in a subsequent section.

whatever the final science endorses as the essential nature of water) to appear black and tarry, and another substance (XYZ) to appear watery.

This notion of metaphysical necessity seems to me to be puzzling, as it seems that they endorse three claims that cannot all be true:

- 1. It is metaphysically possible for experiential properties (watery-stuff) and real nature properties (H_2O) to vary independently of one another (e.g., on Twin Earth)[5].
- 2. The reality explains the appearances (which is the thesis we want to hold for the notion of necessity to be interesting to us)[6].
- 3. Scientists will discover this mind independent reality and reveal it to us.

The problems would seem to be;

- 1. How is it possible for scientists to discover this reality if it is not essentially related to appearances yet is more than a human construct?
- 2. How is it possible for the reality to explain appearances if it is metaphysically possible for them to vary independently of one another?

Hume (1739-40) may be considered to provide a sceptical challenge to the notion that necessity can be discovered a-posteriori by scientists, as the direct reference theorists require. Hume notoriously maintained that necessity could not be a- posteriori discovered, as we cannot observe it (1978 pp78-82). Although we may observe that n_1 types of events always seem to be followed by, or associated with $n2$ types of events, we cannot observe the causal nexus, the compulsion, and hence the necessity.

While many find Hume's analysis somewhat disconcerting a decisive refutation has not been forthcoming. While some have taken the point that

[5]Direct reference theorists shun observation in favour of unobservable 'real nature', which seems to be the converse of the logical positivists (who we shall come on to in a subsequent section).

[6]It is not only that we want to hold onto this thesis. If the observation theoretical distinction does indeed collapse (as Quine and Kuhn convincingly illustrate) then it may be untenable to drop this claim.

necessity is not a causal compulsion (e.g., the logical positivists considered this to express illegitimate belief in an occult power as we shall soon see); Hume's claim just seems to be discounted by realists on the grounds that it is too radically sceptical. The challenge remains, though, as to how we can provide an explanation of necessity. Ultimately what may be required is an abandonment of the notion of necessity as something to be observed or reached by a straightforward process of abstraction, or generalisation as the empiricists took it to be. The dangers of realism would seem to be that necessity is forever beyond us in principle because we lack the faculties by which to apprehend it and it is required to be independent of human construction.

Towards laws of functional relations

There is a story, often told, that Newton discovered the law of gravity (by empirical generalisation) when an apple hit him on the head. But this story would not seem to provide an explanation as to why the 'law' is necessary any more than the children were able to do, with their explanation by generalisation. We may also consider that there are counter-examples to such a 'law' as expressed by the statement 'apples (or other objects), when unsupported fall downwards'. Indeed there are not many universal generalisations that are without exceptions. A great wind or tornado, for example could have an apple blown side-ways. Likewise, it would seem to be conceivable (at this stage of the investigation) that a counter-part may cease to mimic. The very conceivability that there is a world of objects (including counter- parts) and these objects *prima facie* could move in a variety of ways, but they do not is the very thing that needs to be explained.

'Laws' reached by generalisation or association of observed phenomenon do not seem to be enough with respect to ruling out certain phenomena from occurring (and thus providing natural necessity that is comprehensible to us). Typically the most highly prized and revered laws, and those that are considered to provide strict physical necessity or compulsion are laws pertaining to the functional interactions of theoretic notions (such as Newtonian force, mass, density etc) to produce phenomenon that are naturally necessary given the necessity of the laws. Some of these laws of functional interactions are expressed as equations[7]. Newton provided the corpuscularian theory of

[7]E.g., Boyle's law and Hook's law.

light, where light corpuscles are thought to functionally interact or behave in accordance with the laws of motion that he enumerates.

Let us suppose that some of the children take (a) to be naturally necessary, though they consider that this is not to *explain* why it is necessary. They enumerate a theory of functional interactions in order to explain why it is that a counter-part must mimic. They consider that this 'mimicking disposition' is problematic as it is circular (or ad hoc) as an explanation. One of the children proclaims:

> 'We know that a counter-part possesses the mimicking disposition because we see it mimic, and if it didn't mimic then we would explain this by saying that it possessed the mimicking disposition no longer. But is it the counter-parts that possess the mimicking disposition, or is it us; and how could we decide? We cannot, and thus we should exorcise this superfluous 'mimicking disposition' from our explanations'.

The child goes on to elaborate her theory:

> "Our body and our environment cause our beliefs and desires. Our beliefs and desires cause our behaviour. Look again inside our room, Counter-parts have duplicate bodies and environments. They thus have duplicate mental states; which cause their behaviour to be duplicated as well. My theory is better because it explains how your movements are duplicated *when your backs are turned*".

Counter-parts, on this account would thus seem to live in something like Leibnizean pre-established harmony. There is no question as to who is mimicking who as the movements are synchronised in time and thus, according to this theory counter-parts do not mimic, they just appear to do so. Let us attempt to render this explanation in D-N form:

- (i) *initial condition* Counter-parts have duplicate bodies and live in duplicate environments
- (ii) *laws of nature* Body and environment cause beliefs and desires. Beliefs and desires cause behaviour.

- (iii) *explanandum* The children will always conclude that Counter-parts always appear to mimic. (or, more perspicuously, the children will never encounter an experience that would falsify (a)[8]).

Explanation 1. Functional Psychology

Logical Positivism and the observation / theoretic collapse

Positivists such as Comte, (1788-1853) and Mach (1838-1916) considered that there was a sharp distinction between observational and theoretical terms. Comte characterised science as progressing through three stages: The theological stage, the metaphysical stage, and the positive stage (in Hung, 1997, p.320). According to the positivist's theoretical terms such as such as 'mass', 'force', 'gravity' and 'velocity' were metaphysical postulates called in to compel observable phenomena. They considered that the proper business of science was to exorcise these occult forces and proceed in the manner that the empiricists had enumerated, employing methods such as Mill's to associate observed phenomena and make generalisations from them.

We have already considered the prospects of explaining necessity when it is obtained by generalisation from observed phenomena. If we cannot observe necessity (as Hume maintained) then the necessity must be unobservable. The positivists considered that these metaphysical notions were explanatory fictions and that theoretical terms failed to refer and so were meaningless. They thus concurred with Hume that there isn't any necessity in the natural world.

The logical positivists (e.g., Carnap, 1937 in Danto & Morgenbesser (Eds), 1960 pp.150-158) Schlick, in Hung 1997 p.324) continued this theme but were faced with the successes of laws expressing functional relations between theoretic entities. They switched the focus from the theoretic terms to providing an analysis of statements employing theoretical terms in accordance with Frege's maxim that one should never ask for the meaning of a term in isolation from the context in which it occurs (Frege, in Baillie pp.

[8]We will come back to the issue of whether this explanation is of traditional D-N form or not. The reason why I have chosen to formulate the example in this way (and indeed the question as to whether I could have formulated it otherwise) will emerge in due course.

23-40). They thus introduced the famous verification principle of meaning so as to account for laws of functional relations expressing legitimate generalisations in science. It was thought that the meaning of a theoretical term could be given by an operational definition where they were thus translated into statements that secured reference to the natural world. By specifying an operation that could be performed to either support or falsify the statement it was thought that the statement referred to observable phenomena after all even though it used what *prima facie* seemed to be theoretical terms.

The logical positivists considered there to be a theoretically neutral observation language and that scientists should use this language to record and accumulate data. In this way they would be able to generalise to more accurate general claims. There were insurmountable problems with attempting to 'unearth' or construct a theory neutral language (out of sense data, or even 'object language'), and with the attempt to provide a single operation for each 'theoretic' term. As scientists discovered new operations the term would alter in meaning if the meaning was thought to be given by the operations.

This is in direct contrast to the direct reference theory where part of the motivation for there being an objective 'essential nature' is to ensure sameness of reference despite changes in the theories we have of it. While the direct reference theorists are concerned with 'observational terms' it would seem that the verificationists would have meaning vary as new operations are discovered and old ones fall into disuse. The notion of a non-arbitrary distinction between observation and theory also came under fire from theorists such as Quine (1953 in Baillie, 1997), and Kuhn (1962).

Kuhn and theory laden observation

Kuhn (1962) challenged the traditional notion of the scientific enterprise as progressing by accumulation of a-theoretic observations and data. As a historian of science he considered the way in which science has progressed through history. Instead of finding that science progresses cumulatively, Kuhn (1962) found that the history of science is characterised by the following stages:

- (1) *Pre-paradigm*, before the scientific community adopts a paradigm.
- (2) *The emergence of a paradigm*, several compete for the attention of the scientific community and eventually a paradigm is adopted.

- (3) *Normal science*, a period of productive science ensues where scientists construct a cumulative record of data and set about solving problems.
- (4) *Crisis*, anomalies arise that the paradigm cannot explain.
- (5) *Revolution*, a new paradigm is adopted, before the resumption of normal science.

Kuhn thus has a paradigm view of science. Although Kuhn is not clear on the distinction between alternative theories and alternative paradigms (which we will see to be important when we consider the claims that he makes about paradigms) it seems that the best examples of Kuhnian paradigms are Aristotelian, Newtonian, and Einsteinean mechanics; or the alchemists theory of matter, and the atomic theory of matter. These are general theories and may be contrasted with more specific ones such as theories of the chemical constitution of compounds that occur within the paradigm of Dalton's atomic theory.

Kuhn considers that all observations are theory laden and scientists working within a paradigm frame their questions and express their findings from within the paradigm (1962, pp. 16-17). As an example, we may consider that one scientist may record a certain amount of caloric fluid flowing from one substance to another, while another may record one objects molecular motions causing another object to start vibrating as well. Kuhn considers that scientists working within two different paradigms are thus living in (observing and experimenting on) two different worlds. He notoriously makes the following claims regarding paradigms:

- (1) Paradigms do not share any facts in common.
- (2) They do not share any of their problems or standards of solution.
- (3) They do not share any terms (with the same meaning).
- (4) They do not share statements or subject matter.

Kuhn considers that accumulation of data only occurs within a paradigm. Once a revolution has occurred the scientists have to start again (1962, p.13). Because there is (according to Kuhn) no theoretically neutral observation language, and the language of one paradigm is not translatable into another paradigm, he considers that science starts anew each time a new paradigm

is adopted by the scientific community (Kuhn, 1962 pp. 95-96). We may consider that scientists are not currently schooled in the findings of the alchemists, or Aristotelian mechanics and thus there would seem to be some truth in this notion.

Although it is widely acknowledged that Kuhn was a brilliant scientific historian with a tremendous knowledge of the history of science, philosophers have been puzzled by his philosophical remarks about incommensurability and the notion of science as being non-cumulative. If scientists are only productive when governed by a paradigm and if different paradigms influence our world-views so much then how can we hope to discover objective (existing in the world) natural necessity? It seems hard to see how we can consider that we are progressing towards an adequate model of reality if each paradigm needs to start afresh. Kuhn seems to embrace relativism at times and he considers that the notion of objective natural necessity is something that is beyond the reach of scientists.

Such an account has inspired the schools of conventionalism, which I shall just touch on briefly. According to conventionalists there is no objective necessity or causal connection to be found in the world (as Hume maintained). Our theories are constructs, and truth is relative to a theory (in Hung, 1997 ch.9). I shall not consider this further as it seems to me that this account of necessity is giving up on explaining why some phenomena can and cannot occur. While a strictly realist take on necessity (of the sort that the Empiricists or direct reference theorists adhere to) would seem not to be forthcoming, complete relativism or conventionalism should be saved as a last resort strategy with respect to attempting to explain necessity. It is not so much an explanation as an admission of failure.

Despite Kuhn's claim that different theories are different worldviews and thus cannot be compared, he also considers the grounds that we have for choosing one theory over another. He considers criteria such as predictive power, simplicity, and consistency however it seems contradictory for Kuhn to consider that we may need to choose between different theories if they are not even theories of the same thing. While the implications of Kuhn's incommensurability thesis are hotly debated, it seems unanimous that Kuhn's ability as a historian was remarkable and thus his views cannot lightly be dismissed even though he seems in danger of lapsing into relativism.

Let us now consider an alternative functional explanation of the counterparts behaviour.

- (i) *initial condition* counter-parts are made of light corpuscles
- (ii) *laws of nature* light corpuscles are governed by the 3 laws of motion
- (iii) *explanandum* The children will always conclude that Counter-parts always appear to mimic. (or, more perspicuously, the children will **never** encounter an experience that would falsify (a)).

Explanation 2. Newton's Corpuscular Theory of Light

Explanation 1 and 2 are two radically different theories of the children's observations of the counter-parts. Do the Corpuscular theory of light and the Functional psychology theories constitute different paradigms? It would seem to me that they are good candidates for paradigms or world-views: The psychologists consider counter-parts to be real people, with body and mind, whereas the Newtonians consider them to be light images. While the psychologist would record 'counter-part n_1 duplicating the movements of child n_1' the Newtonian would record 'the light image of child n_1 reflecting off the mirror'.

It is interesting to consider that these two alternative theories would indeed be rival explanations of the same *phenomena*. While the scientists recording their observations would record them in different terms, indeed they would not seem to see the same things in this sense, intuitively they seem to be two alternative explanations of the same phenomena. Both theories are attempts to explain why the children will always observe that a counter-part mimics. Hung, (personal communication) considers that while there may be no theoretically neutral language, there is the language of common sense. If we are attempting to explain our experiences in the natural world with a paradigm theory then the language of common sense (while not theoretically neutral) would seem to be a middle ground with which we may compare paradigms in some cases[9].

[9]While this would seem to me to apply to the case of the Newtonians and the functional psychologists in this case (given what they are seeking to explain) Hung considered that Newtonian and Einsteinean mechanics may not be so compared as there is no 'common sense' theory of the phenomena that Einstein was seeking to explain.

Mapmakers and conceptual spaces

Hung (forthcoming pp.12-14) considers another parable that is designed to illustrate the notion of a scientific theory as a conceptual space, and to show that some phenomena are naturally impossible because they are unable to be represented due to the structure of the scientific theory and the limits of the representational space that it provides. I will need to consider this example so as to assist us in making sense of the differences between Newton's theory of light corpuscles and the Folk- Psychologists theory of the functional interaction of mental states.

> Once upon a time four ET's landed on earth. They told the earthlings the distances between their homes, A, B, C, and D, satisfy the following equations:

1. (D1) AB=BC=CD=DA=2unitlengths.

Mapmaker one (MMI) came up with the Square Hypothesis and drew a 2 unit sided square marking each corner clockwise with A, B, C, and D to satisfy the hypothesis (Hung, forthcoming, p.12).

MMI was then informed that:

1. (D2) AC = AB

So he changed it to a rhombus in accordance with the Rhombus Hypothesis. MMI was then informed that:

1. (D3) BD = 2 units length.

Hung considers that 'to map four mutually equidistant points on a piece of paper seemed an impossibility'. Within the conceptual framework of MMI it would be naturally impossible for (D3) to occur. Mapmaker II (MMII) changed the flat (2D) medium of representation (or conceptual space) to a three-dimensional space, and came up with the Ellipsoid Hypothesis where the distance between each pole and the equator is one-third the length of the equator. One ET lived at a pole and the others were spaced out around the equator.

The notion is that a flat piece of paper would not seem to be particularly theory-laden, and yet it restricts the range of phenomena that can be represented by that medium. Hung considers that theories are Category

Systems (ch. 3), Representational Spaces (ch. 4), and Languages (to be distinguished from sets of statements (ch.5). Scientific theories are designed so as to represent the structure of the natural world. Structures rule out the possibility of certain phenomena occurring (Hung, forthcoming, p.31). If we want an explanation as to why a phenomena cannot occur ('why will we never observe a counter-part to not mimic?') then a structure, or a theory can provide limits as to what is possible and thus provide an explanation as to the natural necessity of the phenomena. If we take the structure to be an adequate representation of reality then we can understand why that phenomena cannot occur.

Hung (forthcoming) considers that natural necessity is relative to a theory. We start with the explanandum. The explanandum seems to be contingent, which is why we want it explained. The necessity does not come from the covering law in the sense that we stipulate that the law is necessary and use it to deduce the explanandum, rather we accept the framework that the theory provides and we thus understand why we will never have an experience that would have us conclude that the explanandum was false. If we take the logic of explanation to be D-N then the problem is pushed back one step to the problem of the necessity of the law. Hung considers Wittgenstein's distinction between saying and showing and considers that a conceptual space can show us why the phenomenon is necessary.

If we consider (iii) in both explanation 1 and explanation 2 it becomes apparent that they are not really of traditional D-N form. We have attempted to argue that traditional D-N form is not sufficient to explain the natural necessity of the explanandum because the problem is merely pushed back a single step. Hung considers that instead of explaining (a), scientific theory proceeds by denying the explanandum. The scientist does not seek to explain why the explanandum is necessary, rather the scientist proceeds to explain why it is necessary that the children will never encounter an experience in the natural world that would have them conclude that the explanandum is false (Hung, forthcoming, p.10).

Hung considers that the scientist proceeds by denying the ontology of the explanandum. Instead of seeking to explain why the children conclude that (a) the Newtonian's deny that there are such things as counter-parts. If we accept the Newtonian framework then what is necessary is not (a), rather it is naturally necessary that the children will never have an experience that would have them conclude that (a) is false.

I initially intended to extend the psychologists first attempt at an explanation (the mimicking disposition) to a functional explanation of interacting postulates such as 'belief' and 'desire' in a way that was comparable to Newton's interacting postulates such as 'force', 'velocity' etc for this example. Perhaps intentional psychology could be a real science *just like physics*. The functional psychologists do not deny the ontology of the explanandum in the sense that there are no counter-parts, but they do deny that they mimic (they only *appear* to mimic, but they do not really mimic, they duplicate). While it is clear that once the children find a way to get beyond the bars, or otherwise interfere with the mirror Newton wins with respect to explanation to draw no greater moral from this example may be to pass up an opportunity.

The functional psychologists do not really make a conceptual shift to a new space of possibilities; rather they attempt to reduce the space provided by common sense with the addition of their laws. A greater problem in this case would seem to me to be that there is no independent test of the mental state terms and thus of the laws of functional relations between mental states. While cashing out independent tests of belief and desire states may be problematic for intentional psychology in general, it would seem that in this example it is the crucial problem. If the counter-parts are considered to have duplicate bodies, which support their mental states, then this theory is simply wrong in that counter-parts have no bodies.

I am also led to consider the prospects for intentional psychology in general, as to whether there can be psycho-physical laws similar to those in the above example that are naturally necessary or not. It would seem that intentional psychology is a theory that stays on the level of appearances, however. In so far as we 'reductively explain' intentional phenomena by conceptual shift in terms of physiology there may be cross-theoretic natural necessity. If we stay on the intentional level, however, there would seem to be generalisations of the sort that the empiricists favoured, but no explanation of the natural necessity of the phenomena.

The shift to Newton's theory of light images, on the other hand does provide a radical change in the conceptual space. *There are no such things as counter-parts.* I think that what this shows us is that while it may be possible to make a verbal manoeuvre to render explanations in Hung's variation on the D-N form of explanation, it is indeed the conceptual space provided by the framework that renders the explanation satisfactory. While some consider laws of functional interactions to provide natural necessity it would seem that this is not so much a requisite for natural necessity as conceptual

shift. While in a sense the conceptual shift from MMI to MM2 could be considered a matter of degree (just the addition of another dimension) the significance of this shift is attested to by the alterations phenomena that each theory allows for and prohibits.

Intra-Theoretic laws

While some theorists consider that theories cannot be true or false as it is only statements that can be true or false it would seem that conceptual spaces or theories can be more or less adequate for the task we put them to. If the task is to provide an adequate space for the representation of our experiences in the natural world then it would seem that theories can be assessed by whether the 'impossibilities' ever occur, or whether what is supposed to be 'necessary' does not. MMI was faced with a phenomenon that was impossible according to her framework. The phenomenon was an anomaly for that theory which showed that the theory was inadequate for its purpose.

The question would seem to arise as to whether this notion of necessity is subject to the problem faced by the covering law thesis. If the necessity of the explanandum is due to the necessity of the laws then we would seem to need a further account of the necessity of the laws. Here, though it is not the laws that provide the necessity in an absolute fashion, rather it is that if we accept the framework (and the laws entailed by that framework) then it is inconceivable (from within that framework) that the experience that would falsify the explanandum could occur.

The intra-theoretic laws are thus not necessary in an absolute fashion. They can be more or less adequate, more or less simple etc, but not absolutely true. What is meant by this absolute notion of necessity, though, would seem to be mind- independent reality that is beyond us in principle. We cannot discover necessity in the world, but we can provide a cross-theoretic notion of natural necessity that is subject to the reality constraints of the experiences that we have in the natural world. Hung considers framework truths (or laws of nature considered from within the framework) in much the way that there are framework truths to common sense; such as nothing can be both red and green all over at the same time.

Hung likens this to Wittgenstein's notion of the limits of sense. It would thus seem that framework laws are rather similar to analytic truths, or as

Hung maintains they are 'true by convention'[10]. This is why necessity cannot be provided within a framework, but instead is cross-theoretic, the result of a conceptual shift. There is nothing that compels one to adopt the framework truths or the framework itself. The children could have accepted the functional explanation and been satisfied; it is not that they were simply wrong. We may consider, though that anomalies for this theory are likely to arise in the future, and as such it is not a particularly adequate framework for the explanation of the necessity of the phenomena.

When we consider the necessity of the intra-theoretic laws then they are true by convention, and thus are not naturally necessary though they can be more or less adequate for the representation of reality. The necessity of any given phenomenon thus is relative to a theory, so is cross-theoretic. If we apply this theory to that phenomenon then this theory tells us that this phenomenon cannot occur as a matter of natural necessity. It is not that the framework truths are totally arbitrary, as we are attempting to construct adequate frameworks for the representations that we make of the experiences that we encounter in the natural world. The framework shows us that if the representation is adequate then we will never encounter such experiences in the natural world. Such is the nature of natural necessity.

[10]Analytic truths would seem to be true as a matter of logical necessity whereas the cross- theoretic account of natural necessity is distinguished from this in virtue of its being about the experiences that we can and cannot have in the natural world.

References

Baum, William, M., (1994). Understanding Behaviourism: Science, Behaviour, and Culture, HarperCollins.

Carnap, Rudolph, (1960). 'Elementary and Abstract Terms' 1937 in Danto, Arthur; Morgenbesser, Sidney (Eds), Philosophy of Science, The World Publishing Company.

Caws, Peter, (1965). The Philosophy of Science: A Systematic Account, D. Van Nostrand Company Inc.

Frege, Gottlob, (1997). 'On Sense and Meaning', in Baillie, James Contemporary Analytic Philosophy.

Hempel, Carl G., (1966). Philosophy of Natural Science, Prentice-Hall.

Hempel, Carl G., Oppenheim, Paul, (1970). 'Studies in the Logic of Explanation', in Brody, Baruch A., Readings in the Philosophy of science, pp.8-28, Prentice-Hall, Inc.

Holton, Gerald; Roller, Duane, (1958). Foundations of Modern Physical Science, Addison- Wesley Publishing Company Inc.

Hume, David, (1978). A Treatise of Human Nature, Oxford University Press.

Hung, H.-C., (1997). The Nature of Science: Problems and Perspectives, Wadsworth Publishing Company.

Hung, H.-C., (forthcoming). Beyond Kuhn: Scientific Explanation, Theory Structure, Incommensurability and Physical Necessity.

Kockelmans, Joseph J (Ed.), (1968). Philosophy of Science: The Historical Background, The Free Press.

Kripke, Saul, A., (1972). Naming and Necessity, Harvard University Press.

Kuhn, Thomas, S., (1962). The Structure of Scientific Revolutions, University of Chicago Press.

Kuhn, Thomas S., (1977). The Essential Tension, University of Chicago

Press.

Mill, J, Stuart, (1956). A System of Logic, Longmans, Green and Co Ltd.

Newton, Sir Isaac, (1966). Mathematical Principles of Natural Philosophy and his System of the World, Vol.1 The Motion of Bodies, 1686 translated in University of California Press.

Quine, W.V.O, (1953). 'Two Dogma's of Empiricism', in (1997) Baillie, James, Contemporary Analytic Philosophy, Prentice-Hall Inc.

Salmon, Nathan U., (1981). Reference and Essence, Princeton University Press.

Sankey, Howard, (1994). The Incommensurability Thesis, Avebury.

Scheffler, Israel, (1982). The Anatomy of Inquiry, Routeledge & Kegan Paul Ltd., 1964. Scheffler, Israel, Science and Subjectivity, 2nd Ed. Bobbs-Merrill.

Schilpp, Paul Arthur (Ed.), (1963). The Philosophy of Rudolph Carnap, The Library of Living Philosophers Inc.

Weinert, Friedel (Ed.) (1995). Laws of Nature Essays on the Philosophical, Scientific and Historical Dimensions, Walter de Gruyter.

Wittgenstein, Ludwig, (2001). Philosophical Investigations, Blackwell Publishers.