A few small chunks

Kelly Alexandra Roe

2013

Contents

1 The first 'small chunk' 1

1.1 Problems in the philosophy of psychiatry 1
1.2 Motivating the problems 2
1.3 Normativist concerns 3
1.3.1 Psychiatry as a form of social control. 3
1.3.2 Eliminativism 4
1.3.3 Aristotelian teleology 5
1.4 The two-stage view: The science of dysfunction 8
1.4.1 The DSM view 9
1.4.2 Wakefield's 'harmful dysfunction' analysis 12
1.4.3 Evolutionary dysfunction 19
1.5 Biological wrongness and the division of labour 22
1.5.1 Natural norms 22
1.5.2 The proposed division of labour 34

2 The second 'small chunk' 38

2.1 Some assumptions of evolutionary psychology 38
2.1.1 Modularity 38
2.1.2 Behavioural, cognitive, neuro-biological, genetic mapping 39
2.1.3 Dysfunctioning module 41
2.2 A shifting focus 45
2.2.1 Adaptive plasticity 45
2.2.2 The accumulation of social capital 45
2.2.3 Minimal cognitive requirements 45
2.3 Health in physiological systems 45
2.4 From Simple Co-Ordination to Co-Operation 46
2.4.1 Cheating vs Defection / Incapacity 48
2.5 Responsibility / capacity / culpability 50
2.6 Rethinking Evolutionary Accounts 52

3 The third 'small chunk' 53
3.1 Interactive kinds and the impact of classification 53
3.1.1 From essentialism to nominalism (with looping in between) 53
3.2 Inductive Generalization 54
3.3 Homeostatic Property Clusters 56
3.4 Grounding Mechanisms 57
3.5 Socially Constructed Kinds 58
3.6 Looping Kinds 59
3.7 Hacking's Imitation and Internalization (I&I) Model of Apathetic Children 62
3.8 Trauma Model 63
3.9 Implications of implicit looping for taxonomy 69
3.10 Implications for problem cases (e.g., addiction, sociopathy) . . 71

4 References 74
4.1 References 74

Chapter 1

The first 'small chunk'

1.1 Problems in the philosophy of psychiatry

A number of problems or questions have been thought to motivate our interest in the philosophy of psychiatry. I will begin by introducing some of these. I will then turn to two main approaches to answering the questions of the nature of mental disorder and the science of psychiatry. On the one hand we have the critique that there is no such thing as mental disorder, or alternatively that there is little more to mental disorder than a certain kind of social and / or moral norm violation. The idea is then that psychiatry is a form of social control like law and it is not appropriately regarded to be a specialist branch of medicine. On the other hand we have the two- stage view whereby suffering, distress, and / or normative violation is by itself insufficient for mental disorder. In addition to a normative aspect there needs to be something biologically wrong with the person. The sciences of mental disorder and psychiatry as a clinical field are thus conceived of being concerned with discovering and treating that biological wrongness. After considering Wakefield's argument for identifying the biological wrongness with an objective, scientific notion of evolutionary dysfunction we will then be in the position to consider whether it (or an account like it) can assist us with the questions that initially motivated our interest in the account.

I will argue that scientific accounts of biological wrongness seem problematic with respect to the development of a science that is capable of answering the questions that initially motivated our interest in the topic. I will urge that we face up to this and hold some of the normative assumptions up for examination and critique. This is not to say that sciences have nothing to contribute. The sciences are, of course, coming to discover important information that comes to bear on these issues. The next two parts will be

concerned to develop a better understanding of the contribution of sciences to the study of mental disorder.

1.2 Motivating the problems

We are repeatedly told that we are living in the midst of a mental health epidemic. As many as 1 in 4 are thought to be afflicted by mental disorder at some point in their life. The social cost of mental disorder is thought to be staggering and a number of theorists speak out vocally about the need for it to be increasingly prioritized in government health budgets and individual health insurance plans.

The main issues in the philosophy of psychiatry have been those of defining and classifying mental disorder. Defining disorder involves delineating the boundaries of madness, or differentiating between the sick and the well, the mad and the bad, the just odd, the disgusting, and the morally reprehensible. It involves distinguishing between the non-pathologically grieving - and the depressed, the attention deficient and the naughty, the deformed, and those in search of the more ideal. Some individuals might think they are ill, but really they turn out to be mistaken and they are not. Surely not everybody who believes themselves to be ill gets to be ill. We need to distinguish between the genuinely ill and the fakers and malingerers. We also need to distinguish between the genuinely ill and those who are in pursuit of perfection. This issue seems especially pressing with respect to prioritization of funding for health care and whether or not individuals should receive health insurance reimbursement or government assistance for certain sorts of intervention.

We also need to distinguish between the conditions that are medical disorders and those that are not[1]. Are such conditions as attention deficit, pathological gambling, alcohol addiction, binge eating, paedophilia, and narcissism issues of mental disorder or is something different going on? Is dissociative identity disorder merely a different morphological variant on borderline personality, is it importantly different and / or is there really no such thing? Are apathetic children suffering post- traumatic stress or is something different going on? These issues seem important as they impact on whether or not individuals are due special consideration around failure to meet social

[1]Throughout I shall use the term 'condition' as a value neutral way of referring to different states of people. What I mean is probably made clearest by way of example: Having small breasts, being pregnant, having cancer, being homosexual are all examples of 'conditions' in my neutral sense. We would like to know which (if any) of these are bio-medical disorders.

obligations such as those to employers or dependents. It seems important in determining whether or not they are justified in adopting the sick role. These issues seem to relate to whether the appropriate attitude towards failure to meet social obligation is one of empathy, pity, and assistance, or whether it is appropriately one of blame, punishment, and condemnation. These issues seem to impact upon whether governments and health insurance plans are under some sort of obligation to provide help or assistance and perhaps parity of treatment funding for mental disorders. Is the appropriate avenue for assistance that of psychiatry, clinical psychology, social worker, religious leaders, self help consumer groups, or some combination?

These issues seem to matter insofar as they impact upon whether an individual does or does not receive treatment and what kind of treatment it is that they should receive. It impacts upon the conception we have of the person who is regarded to be mentally ill. It impacts upon the self-conception of those who are regarded (or not regarded) to be mentally ill. Some people report finding solace in being given a diagnosis and that their diagnosis assists them in their view of themselves and in making connections with understanding others. It also impacts upon from whom they should receive treatment. There are many forms of suffering only some of which are thought to be medical, however. A number of theorists have spoken out against the medicalization of pregnancy and the idea that medical doctors should be authoritative when it comes to managing the condition.

1.3 Normativist concerns

1.3.1 Psychiatry as a form of social control.

The 1960's saw much critique of psychiatry as people started to wonder whether there was a scientific basis to mental disorder or whether psychiatry was more like law or ethics than medicine in the sense of being an institution concerned primarily with the social control of normative violation. Typical examples of cases that brought these issues to the fore include Soviet psychiatrists who liberally diagnosed individuals with schizophrenia solely on the basis of political dissent. The presence of homosexuality in early editions of the Diagnostic and Statistical Manual of Mental Disorders, and the success of lobby groups in its ultimate removal. The suggestion that slaves who desired to escape their owners were suffering with the disorder 'Draeptomania'. While this never came to be adopted in an official system of classification critics started to question whether there was any principled reason to distinguish between such bogus categories as these and other alleged mental

disorders such as schizophrenia and bi-polar. The majority of theorists accepted that those with mental disorders were violating social norms. The controversy was over whether this exhausted the nature of mental disorder and psychiatry or whether there was a non-normative, scientific aspect that served to ground psychiatry firmly in the biological sciences.

Some theorists maintained that psychiatry (and psychiatric or mental disorders) were importantly different from non-mental disorders. The idea here was that mental disorders dealt with normative violations in a way that bio-medical disorders did not and that psychiatry was more like law than medicine in the sense of being an institution and practice concerned with policing normative violation. Szasz maintained that this was because there was something special about the mind. He adopted a dualist position about the mind and dualistic arguments to arrive at the conclusion that the mind could not dysfunction the way that bodily or physical systems could since the mind was different in kind. We might well be surprised at Szasz insistence on dualism in the face of the relative success of materialism or naturalization. I will not enumerate or critique Szasz's position here. It is an interesting claim that psychiatry and the science of mental disorder is importantly different because it deals with the mind. Many have thought that there was something specially different about psychiatry precisely because the mental is importantly different. Usually the idea is that there is something special about mental states such as intending or reasoning. Insofar as psychiatry is concerned with issues of this perhaps it is more problematic in principle than the rest of medicine.

Fulford (2000, p. 83), for example, in writing about the 'naturalization cascade' considers that it is common ground even among naturalizers that some areas of science are more overtly value laden than others. '... psychiatry is more overtly value-laden than more high-tech areas of medicine, medicine more than biology, and biology more than the natural sciences, such as physics'. What naturalizers have in common, however is 'the belief that somewhere deep down in the naturalization cascade there is a value-free foundation, an heuristic holy grail, on which biology, and in turn the theoretical cores of medicine and psychiatry can, in principle, be built up as mature scientific disciplines' (Fulford, 2000, p. 83).

1.3.2 Eliminativism

In response to the normativist critique some theorists have been led to be eliminativists about mental disorder. Eliminativism has a long history in science. Certain notions such as phlogiston and witches were once considered scientifically respectable but, the thought is, we have come to learn that there

is no such thing. Some theorists have been led to conclude that similarly, there is no such thing as mental disorder fairly generally or that there isn't any such thing as particular kinds of mental disorder, or a particular kind of mental disorder (e.g., some theorists have maintained that there isn't any such thing as schizophrenia, rather we have a group of things so are better to speak of 'the schizophrenias').

The idea here is that we take a category like witches or phlogiston. The thought is that witches were thought to have a bunch of properties that enabled us to more or less identify them and thus to learn about them. The thought was that while there were indeed women and some of them had cats and beards and lived a solitary lifestyle. And indeed some of them floated when we attempted to drown them. And yet these individuals turned out not to have special powers. And thus we conclude that since having special powers (and using them for malevolence) is incredibly central to the concept of witchhood that the appropriate thing to conclude here is that there aren't any witches. And thus since there aren't any witches we should probably stop drowning innocent (insofar as they aren't using special powers for malevolence) women. Whether or not they have black cats. And live by themselves. And float.

Or consider phlogiston. The wonderful magical heat fluid that flows from one substance to another that explains how when a hot object is placed near a cool object the cool object heats up. We could have concluded that phlogiston was a wonderful magical heat fluid that didn't weigh anything (since cooler items don't weigh more). We were led to conclude that there wasn't any such thing as a heat- substance that was transferred in the transfer of heat. And of course, on some accounts, at least, there aren't any such things as beliefs and desires according to neuroscience. And there aren't any tables or chairs (or substances, generally) according to subatomic physics.

In the psychiatry case theorists have typically used eliminativism about disorder to justify their notion that psychiatry isn't legitimately considered a specialist branch of medicine and to support their idea that the institution be eliminated or abolished. Sometimes this goes along with the idea of the abolition of involuntary treatment and the abolition of the insanity defence.

1.3.3 Aristotelian teleology

Christopher Megone (Megone, 1998, 2000) thinks that the relevant notion of function and dysfunction for medicine and psychiatry is a blend of fact and value and that two-stage theorists (as we will go on to consider) are wrong in thinking it is solely a matter of fact. Megone defends a broadly Aristotelian view of health and function. There has been a fairly recent upsurgance

in theorists adopting a broadly Aristotelian view of health. The trouble with this kind of approach is that it seems more regressive than progressive. Aristotle's metaphysics, in particular his essentialism, isn't something that the majority of theorists take seriously these days. It seems hard to see what sense we can make of something like the Aristotelian view within the context of more modern metaphysics. Lets have a go, however.

Aristotle had a view of essence whereby each thing was itself and not another thing in virtue of essential properties. Essential properties are had by things of one kind and not by things that are not of that kind. On this account essential properties are thus necessary properties for the thing to exist as or *qua* that type of thing. For example, to say that water is essentially H_2O is to say that a molecule of H_2O that came to lose or gain a hydrogen atom would no longer be a molecule of water. Or alternatively, a substance that appeared to be water that turned out to be comprised of molecules that weren't predominantly H_2O, would turn out not to be water after all.

Aristotle thought that when we asked about the essence of a person we thus needed to try and figure out what properties persons had that other critters lacked. Aristotle was led to conclude that only persons were rational animals and thus persons were essentially rational animals. This might seem a particularly promising line insofar as it seems intuitive to many that mental disorders have something to do with failures of rationality. It is perhaps more surprising that this line is thought to provide a take on non-mental disorders as well.

Aristotle thinks that the function of a person is what is good for a person as or qua persons. He thinks that rationality is what is good for persons, and understands rationality as being very broad indeed to include culture and language and indeed, political science. He thinks that this good for persons is health. That illness is disruption to rationality. Either direct disruption in the case of mental illness or indirect disruption in the case of physical illness. For Aristotle essential potentialities don't need to be actualized and thus it doesn't count against the view that there are people who aren't (perfectly) rational. Their personhood is called into question, however. We do wonder about whether children or mentally retarded are persons. Not whether they are homo sapiens - but whether they are persons. Part of what seems so very problematic about being ill is that ones personhood is diminished.

On the Aristotelian view the person is primary and the parts derive their function from the contribution they should make to the function of the person. He:

> ... argues first that there is a quite general connection between

> the function of a kind and the good of that kind... examples of musicians and sculptors whose good, qua musicians and sculptors, is tied to the function of each skill. Given this general claim, if humans qua humans have a function, that will determine their good too. Aristotle then cites two considerations in favour of the view that humans as such have a function. First he suggests that if carpenters and tanners have a function, then humans as such should do also; and second, he notes that since eyes and feet have functions, the human being as a whole must do so (Megone 2000 p 50).

One interesting feature of the view is that it presents health as an ideal. An ideal that is equated with rationality. What is meant by rationality turns out to be much broader than we are perhaps typically used to thinking of it. Health, too. Political science. Instead of thinking of disease as a thing (an entity) it is lack of approximation to an ideal.

Megone (2000, p. 49) says that according to Aristotle 'Human beings have a function in the sense that there are goals or purposes that good human beings will realize (actualize). The Aristotelian claim is that the ultimate goal for a good human being as a whole is to live the life of a fully rational animal. The function of the bodily and mental parts of a human is to operate in ways that contribute instrumentally or constitutively to the realization of this goal. Thus a bodily ailment such as a lung condition is, at root, an incapacitating failure of the lung to function in the way whereby it contributes optimally to the life of the fully rational animal'.

Megone thinks that empathy with others, the sense of justice, complex social emotions like honour and pride and envy, capacity for language etc. are what Aristototle has in mind as part of what a suitably rich conception of the fully rational life will include (paraphrased p.53) "In particular it makes clear that this claim can be understood only if it is recognized that he has in mind a very rich conception of rationality exhibited by a language speaking, emotionally complex, virtuous, social animal (and an animal also exhibiting the traits of theoretical rationality)." p.53.

> The potentialities that are realized (actualized) when a member of a natural kind fulfils its function also constitute the essential potentialities of members of that kind. Thus that cycle of changes that constitutes the function of the human being as a whole (and the basis for determining the functions of parts) also determines what a human being is. As a result, making the supposedly evaluative judgement as to what constitutes an ill or healthy human

> being is the very same thing as making the supposedly factual judgement as to what constitutes a human being. The judgements are the same. There is no separation of fact and value (Megone, 2000, p.54).

If humans qua humans had a function or goal that would be a suitable end in itself. role of political science to investigate this goal.

> Contrary to this metaphysical scheme stands the Aristotelian framework in which the scientific investigation of the natural world is an investigation of a worlds in which facts and values are fused. Correspondingly, the Aristotelian account rejects the fact/value distinction, rejects a Cartesian substance dualism about the mind and the body, and offers a much richer scheme of genuinely explanatory modes of (scientific) explanation that accords with the nature of the natural substances that constitute the world that science investigates. It is not surprising that an Aristotelian account of illness, which is a facet of a natural kind (in this case, human illness), should therefore reflect this fusion of facts and values and the important role of irreducible teleological explanation within this metaphysical picture (Megone, 2000, p.64).

I will now turn to the two-stage view that attempts to defend a value-free biological foundation for psychiatry with respect to appealing to a non-evaluative notion of biological dysfunction. In the last part I will return to the normative objection and a discussion on whether it forces us to eliminativism.

1.4 The two-stage view: The science of dysfunction

Imagine a tribe of hunter-gatherers where each individual contributes towards the hunting or gathering of food on a daily basis. Consider the following four cases:

- (1) During a hunt one of the members of a tribe is accidentally stabbed in the leg by a spear. When the spear is pulled out the skin is open and there is blood. When the person attempts to walk they scream and cease in their attempts.

- (2) One of the members wakes up in the morning and when the person attempts to walk they scream and cease in their attempts. The leg doesn't look any different to what it looked like yesterday when the person participated successfully in the collection of food.
- (3) As above except in this case the persons prior participation in the collection of food was unsuccessful and the person has previously expressed reluctance to participate.
- (4) As above except in this case the person refuses to attempt to walk.

I think most will find it plausible that the first case is the clearest case of the presence of bio-medical disorder whereas the fourth case is the clearest case of the absence of bio-medical disorder. The following considerations seem to be relevant:

- (1) The presence of physical abnormality / dysfunction / defect. What this is trying to capture is that in the cases where the skin is open and there is visible blood we tend to have the intuition that there is something physically bio-medically wrong with this person.
- (2) The presence of suffering, pain, distress. What this is trying to capture is that the person seems perturbed psychologically and that this is limiting their normal activities.
- (3) We have some kind of duty or obligation to assist since they would be better off if what was wrong were to be put right.

And thus we arrive at a fairly generic version of the two-stage view. According to the two-stage view there is firstly an objective aspect to disorder (provided by the first condition) and secondly a normative aspect to disorder (the second and third condition). There are a number of different particular versions of the two-stage view as different theorists attempt to cash out the features of each stage in slightly different ways.

1.4.1 The DSM view

According to the two-stage view there are two individually necessary and jointly sufficient conditions for mental disorder. Firstly, there is malfunction, and secondly, the malfunction has harmful consequences for the individual and / or society. The clinician's handbook The Diagnostic and Statistical Manual of Mental Disorders endorses the necessity of the first condition when it states that 'whatever it's original cause it must currently be considered

the manifestation of a behavioural, psychological, or biological dysfunction within the individual'. While we will go on to see that Wakefield differs from the DSM by maintaining that inner rather than purely behavioural malfunction is required it is clear that Wakefield and the DSM are similar in regarding malfunction to be necessary for mental disorder. The two-stage view has been extremely influential, partly because it promises, by way of the objectivity of the first condition, to ground psychiatry firmly on a scientific footing. The notion is that scientists can investigate malfunction independently from our normative assessment of harm. It is partly because malfunction is regarded as objective that the two-stage view has been embraced by the majority of psychiatrists.

One advantage of the two-stage view is that the first stage is seen as setting the scientific foundations of the study of disorder such that the scientists can learn about functions and malfunctions and identify them independently of our normative assessment of harm. While harm might be dependent on the values, norms, and activities of particular cultures malfunction is thought to be universal. The sciences can thus get on with discovering the objective facts about function and malfunction independently of our assessment of the normative consequences of the malfunction. There are facts about the individual malfunctioning that are necessary for mental disorder and as such mental disorder is not solely a matter of the individual violating norms. We thus have a picture of the division of labour between scientists on the one hand (engaged in the process of discovering dysfunction) and normative theorists on the other (who consider the notion of harm).

Before offering a definition of mental disorder the American Psychiatric Association (2000, pp. xxx-xxxi) begins with some caveats.

> ...although this manual provides a classification of mental disorders, it must be admitted that no definition adequately specifies precise boundaries for the concept of "mental disorder". The concept of mental disorder, like many other concepts in medicine and science, lacks a consistent operational definition that covers all situations. All medical conditions are defined on various levels of abstraction - for example, structural pathology (e.g., ulcerative colitis), symptom presentation (e.g., migraine), deviance from a physiological norm (e.g., hypertension), and aetiology (e.g., pneumococcal pneumonia). Mental disorders have also been defined by a variety of concepts (e.g., disease, dysfunction, dyscontrol, disadvantage, disability, inflexibility, irrationality, syndromal pattern, aetiology, and statistical deviation). Each is a useful indicator for a mental disorder, but none is equivalent to the concept,

and different situations call for different definitions.

There are several issues that are raised by this section of the DSM. Firstly, the APA is explicit about attempting to offer an operational definition that enables clinicians to identify which individuals are mentally disordered. The APA is also explicit about attempting to offer an operational definition that justifies which conditions are included in the DSM as mental disorders. We may well wonder whether the above definition can do the work that is required of it. On the other hand we may have sympathy that if we needed to offer a satisfactory definition of 'life' before biological science could get up off the ground then we might well still be waiting. The American Psychiatric Association continues on:

> Despite these caveats, the definition of mental disorder that was included in DSM-III and DSM-III-R is presented here because it is as useful as any other available definition and has helped to guide decisions regarding which conditions on the boundary between normality and pathology should be included in DSM-IV. In DSM-IV, each of the mental disorders is conceptualized as a clinically significant behavioural or psychological syndrome or pattern that occurs in an individual and that is associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one. Whatever its original cause, it must currently be considered a manifestation of a behavioural, psychological, or biological dysfunction in the individual. Neither deviant behaviour (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.

Rachel Cooper (2005, 2007) maintains that the definition was instead provided in the attempt to justify why certain conditions were included. Cooper has noted that the attempt to define mental disorder occurs at the time in which the APA was under considerable pressure from gay rights activists and anti-psychiatry lobby groups for the APA to justify how they decided that certain individuals / conditions were mentally disordered. In particular, the attempt was to ground psychiatry (or to justify psychiatry's

status) as a speciality within medicine, and to directly counter the concern that psychiatry was in the business of confining and treating people who were merely in violation of social and moral norms as some anti-psychiatrists maintained. The DSM definition addresses this latter issue quite specifically in the last sentence.

> It is to be understood that inclusion here, for clinical and research purposes, of a diagnostic category such as Pathological Gambling or Paedophilia does not imply that the condition meets legal or other non-medical criteria for what constitutes mental disease, mental disorder, or mental disability (xxxvii).

This carving off of the scientific concerns from the extra-scientific concerns is similar to Murphy's take when he focuses his book on the scientific notion of mental disorder rather than the extra-scientific notion that comes up in issues to do with moral and legal responsibility. Murphy also places treatment as an extra-scientific concern. I don't think that we can separate out these two stages of theory that occurs prior to application. This begs the question with respect to assuming an answer to the questions that interested us rather than discovering an answer to those questions.

The DSM definition includes many of the notions that different theorists have attempted to use to account for disease. It attempts to incorporate the literature. It joins the criterion that have been offered with disjunction. The problem with this definition is that it is hopelessly unclear how to apply it to particular cases. For example, consider homosexuality. If we want to know whether homosexuality is a disorder or not we can apply the criterion to it and see if it fits. The problem is that the definition is fairly unclear. On the one hand a good conceptual analysis is as vague and approximate and rough and ambiguous as the term was initially. On the other hand it is hard to see how it it particularly useful to us.

1.4.2 Wakefield's 'harmful dysfunction' analysis

Wakefield maintains that it follows from our pre-theoretic concept of bio-medical disorder that there are two individually necessary and jointly sufficient conditions for disorder as follows:

- (1) There is evolutionary dysfunction to a mechanism
- (2) This dysfunction results in harm to the individual and / or to society

He thinks that it similarly follows from our pre-theoretic concept of mental disorder that:

- (1) There is evolutionary dysfunction to a mental mechanism
- (2) This dysfunction results in harm to the individual and / or society

We can see from the above account that Wakefield is a two-stage theorist both about bio-medical disorder in general and about mental disorder in particular. The first of the two conditions is intended to be non-normative or non-evaluative, while the second of the two conditions is intended to be normative or evaluative. Wakefield's focus has been on providing an account of the first condition and defending it against critique. He doesn't argue for the relevant notion of harm being normative or evaluative the way that he argues for the relevant notion of dysfunction being objective. This is because the burden of proof has been thought to be on theorists who maintain that there is an objective notion of dysfunction that can show psychiatry to be grounded in medicine to be grounded more generally in the biological sciences.

Wakefield is of particular interest to us not only because of his account of evolutionary dysfunction but also because he has systematically articulated and defended his view over a period of 20+ years (See, for example 1992a, 1992b, 1993, 2000, 2003, 2004). (add 2011 reference and see if there is one earlier than '92)

Wakefield maintains that his position is "black box essentialist" and he explicitly contrasts it with a descriptivist (or "Roschian") cluster view. The following is a reconstruction of Wakefield's argument for evolutionary dysfunction being necessary for bio-medical disorder:

- (1) [P1] It follows from our concept of mental disorder that there is a dysfunction to a mental mechanism (in some pre-theoretic sense of dysfunction) that results in harm to the individual and / or to society.
- (2) [P2] It follows from our pre-theoretic notion of dysfunction that there is an historical process that fixes biological functions and dysfunctions. The nature of that process is to be discovered by science.
- (3) [P3] Scientists have discovered that the relevant historical process for fixing biological functions and dysfunctions is evolution by natural selection.
- (4) [C] Bio-medical disorders are evolutionary dysfunctions of a mental mechanism that result in harm to the individual and / or to society.

Wakefield starts out by appealing to our common-sense concept of biomedical disorder that we have in virtue of being competent speakers of our language. He thinks that reflection upon this concept will lead us to see the truth of the two-stage view[2].

The first premise in the reconstruction of the argument consists in something that is meant to follow conceptually or analytically from our concept of mental disorder. The notion is that in order to grasp the notion of mental disorder one must grasp that there is something wrong or dysfunctional about a person who has one. 'Disorder exists only when an internal mechanism is dysfunctional, specifically in the sense that it is incapable of performing one of its natural functions (Wakefield, 1999, p. 375)'.

He also maintains that 'at this stage of the analysis, natural function is used in an intuitive sense that has existed for millennia, not in a technical evolutionary sense '.

Wakefield draws an explicit analogy between his approach to 'mental disorder' and the causal-historical approach to natural kind terms such as 'gold' and 'water' that was defended by theorists such as Kripke and Putnam in the 60's[3]. A popular view in semantics is that there are two aspects to meaning; what we may (roughly) call a 'primary intension' or an 'A intension' or a 'description' or a 'meaning' on the one hand, and what we may (roughly) call a 'secondary intension' or a 'B intension' or a 'real nature' or a 'referent' on the other.

The primary intension is thought to consist in something along the lines of a description or a list of features that are cognitively significant and that form part of the meaning of the term / the content of a concept. In the case of 'water' / WATER the A intension consists in something along the lines of the colourless, odourless, potable, drinkable stuff that falls from the skies and fills the lakes etc[4]. In the case of 'gold' / GOLD the A intension consists in something along the lines of the yellowy, shiny, malleable, valuable metal. Now, while it is thought to be contingent that the terms 'water' or 'gold'

[2]One issue with putting things this way is that we may now wonder what to make of theorists who deny the two-stage view. Do they have a different concept of disorder since they weren't led to the two-stage view that Wakefield thinks follows from our concept? Perhaps these theorists have difficulty seeing what follows from their concept or perhaps they have a different concept or perhaps Wakefield is wrong in thinking that the two-stage view follows from the concept of disorder. How do we decide? I will return to this issue.

[3]Wakefield doesn't discuss some of the more modern controversies within the two-dimensional semantics framework such as the nature of the a-priori, issues of concept individuation etc.

[4]I follow convention of CONCEPT, 'word' and world. So we have the concept WATER the word 'water' and the water (or watery stuff) in the world.

have the A intension that they have, it is thought that in order to grasp the concept of WATER or GOLD one does need to grasp the A intension. That is what it is to understand the meaning of the terms or to have grasped the relevant concept. As such, it it thought to be a-priori, or a conceptual or analytic truth that the A intension of WATER or GOLD is the description that is listed in the A intension.

Kripke and Putnam went on to argue that while this is one aspect to meaning, there is also another aspect to meaning - reference - that served to link the term or the concept on to something that mind-independently exists in the actual world. The B intension is thought to be discovered by science by way of their discovering what the realizers or the A intension are in our world. In the case of 'water' scientists discovered that the colourless, odourless, potable, drinkable stuff that falls from the skies and fills the lakes etc around here was H_2O In the case of 'gold' scientists discovered that the yellowy, shiny, malleable, valuable metal around here has atomic number 79. The notion then is that certain kinds of terms - natural kind terms - function to track the reference or the B intension. So in Putnam's famous twin-earth scenario if there is a world (not the actual world) in which the watery stuff (the A intension) turned out to be XYZ, then the watery stuff on that world would not be 'water', WATER, or water. Conversely, if it turned out that if there is a world (not the actual world) in which H_2O is black and tarry then the correct way to describe the world is that their 'water', WATER, or water is black and tarry. This is because 'water', WATER and water is necessarily or essentially H_2O given that 'water' functions as a natural kind term (which is to say given that 'water' tracks the B intension) and that H_2O is the B intension / nature on this world.

The second premiss of Wakefield's argument is also meant to follow conceptually or analytically from our concept of mental disorder. Wakefield maintains that it simply follows from our concept that the nature of the bio-medical dysfunction that he arrived at in premise one is something that is for science to discover. This is, in effect, to treat the relevant sense of 'bio-medical dysfunction' to be a natural kind term whose essential nature (reference or B intension) is to be discovered by science.

In the second stage, Wakefield maintains that 'the seemingly anthropomorphic notion of the function of a biological mechanisms is analysed in straightforward scientific causal terms... The language of function is used to indicate that certain effects of biological mechanisms are so complex, beneficial, and intricately structured that they cannot be accidental side-effects of random causal processes but, like the intentionally designed functions of artefacts, must somehow be part of the explanation of why the underlying mechanisms exist and are structured as they are... Assertions that cer-

tain effects of a mechanism are useful do not offer any explanation of the mechanism; the usefulness could be due to chance. In contrast, function attributions implicitly make an explanatory claim, namely, that the mechanism is the way it is partly because of its usefulness. Disorders, then, are failures of mechanisms to perform their natural functions, where natural function is understood in the aforementioned explanatory sense.'

He then maintains that 'strictly speaking, these two steps complete the conceptual analysis of disorder. However, this analysis does not explain how an effect (e.g., pumping, seeing) could explain its own cause (the heart, the eyes), nor does the analysis provide a criterion by which one can scientifically distinguish natural functions from other effects in a manner more precise than that afforded by common-sense intuitions. The analysis inevitably leads to the question, what kind of underlying process could possibly be responsible for such seeming design in natural systems without any designer? To answer this question, there needs to be a scientific theory of how such explanatory effects can come about. The attempt to answer this question leads to a third step in the argument: Evolutionary theory provides the only plausible scientific account that presently exists of how the natural functions of a mechanism can explain the existence and structure of the mechanism... The third, theoretical argument leads to the conclusion that disorders are failures of mechanisms to perform functions for which they were naturally selected'.

The third premise consists in an empirical claim that is meant to be revisable in the face of future empirical evidence. The notion is that as the best current chemical theory holds that water is H_2O and that gold has atomic number 79 that the process for fixing biological functions and dysfunctions is evolution by natural selection. The conclusion thus follows analytically from the premises: Given that our notion of BIO-MEDICAL DISORDER entails BIOLOGICAL DYSFUNCTION (as asserted in premise one); and given that **biological dysfunction** is a natural kind term (which is to say that it tracks the B intension as asserted in premise two); then, given that science tells us that (biological dysfunctions) are fixed by evolution by natural selection (as asserted in premise three); it follows analytically or conceptually from those premises that BIO-MEDICAL DISORDERS are (at least) failures of evolutionary function.

The "black box essentialist" account I (Wakefield, 1997, 1999b, 2000) present is one flavour of such essentialist views; the name and a few nuances are mine but the basic ideas are derived from the noted philosophers. The proposed concepts are essentialist because category membership is ultimately determined not by observable properties (e.g., for water, clear, thirst-quenching liquid) but by the hypothesized theoretical property or "inner nature" that explains the observed features (for water, H_2O). The proposed

concepts are black box because, rather than defining concepts by specific theoretical properties (e.g., "water isH20", such concepts postulate a theoretical explanatory structure and remain agnostic on its identity, which may be unknown (e.g., "Water is anything that has the same substance-essence as the clear thirst-quenching liquid in the lakes and rivers"). This essentialist definition uses the prototype properties not as universal criteria for the construct but only to indirectly refer to its essence. Thus, the definition allows things very different from the prototype set, such as ice, steam, or H_2O atoms floating in space, to be water. The description based on the prototype sample allows us to fix the reference of the construct term by a "closed" concept, and all other propositions remain "open" but are not part of the concept (Wakefield, 2004, p. 79).

Wakefield is thus led to identify mental disorder with a failure of an internal mechanism to perform its evolutionary function. Our concept of water is such that it is transparent, potable etc. Our concept of water is also such that water is a substance. Best scientific theory then tells us the underlying property of the substance that is responsible for the properties that featured in our concept. The essential property of water is thus the property that the scientists have discovered. Wakefield similarly thinks that our concept of mental disorder is such that it is a harmful dysfunction. Our concept of harmful dysfunction is also that there is a causal process that fixes the functions and dysfunctions. Best scientific theory then tells us that the underlying causal process is evolution by natural selection. The essential property of mental disorder is thus the property that scientists have discovered.

Black box essentialism can also be used to identify kinds of mental disorder as follows:

> Roughly, "schizophrenia" might be defined as follows: Take as the prototypical set those who Bleuler originally picked out as clear cases of schizophrenia, when he defined the concept; an individual then falls under the concept of schizophrenia if he or she possesses the underlying psychopathological structure that was shared by most of that prototypical set and explains the symptoms that led to their being placed in the set (Wakefield, 2004, p. 81).

On scientific discovery:

> But isn't it possible, however remotely, that temperature could turn out not to be mean kinetic molecular motion after all, or that water could yet turn out not to be H_2O? We could wake

> up tomorrow, for example, and find that chemists had discovered that their instruments had been systematically mis calibrated and their instruments' readings misinterpreted, and that the liquid in the familiar lakes and rivers has as its molecular structure not H_2O but, say, XYZ. If that happened, we would surely not conclude that there is no water in the lakes and rivers... Rather, we would conclude that water is not H_2O after all, but XYZ showing that even theoretical reductions get their legitimacy... from whether they in fact match out pre-theoretical concept (Wakefield, 2004, p. 81).

Two-stage theorists attempt to distinguish the normative aspect of disorder and separate it out from the non-normative aspect. The basic idea behind such as approach is to isolate out a scientific theoretic core that is common to psychiatry, medicine, and the biological sciences more generally. This aspect is thought to exist prior to, independently of, and upstream from issues of normativity or other application. This aspect can be investigated and learned about independently from the normative aspect. While the normative issues are important we can put them to one side and get on with developing the science of psychiatry. The status of psychiatry as a specialist branch within medicine is assured, however, since the subject matter – that of mental disorder – forms a subset of bio-medical disorder more generally. There is more to mental disorder than normative violation. Much ink has been spilt as theorists have attempted to provide a more precise specification of the nature of the non-normative aspect. Theorists have turned to work that has been done in the philosophy of biology on naturalizing norms with respect to naturalistically respectable accounts of talk of 'function' and 'dysfunction' in evolutionary biology.

More recently theorists have turned to physiological accounts of function and the idea of diseases or disorders as (at least partly) due to disruptions or failures of homeostasis. The idea here is that insofar as we grant that function and dysfunction talk is scientifically respectable in evolutionary biology and physiology medicine and psychiatry inherit that scientific respectability. The notions of disease and dysfunction in medicine and psychiatry are no more problematic than they are in medicine, physiology, and evolutionary biology. We can thus proceed with the science of disorder in the face of the anti-psychiatry critique.

1.4.3 Evolutionary dysfunction

Philosophers have been interested in naturalistic accounts of mental functions and dysfunctions in the hope that it would help us naturalize intentionality or mental content / representation. The idea is that if there is a scientifically respectable notion of function and dysfunction perhaps this can be utilized to account for mental representation / misrepresentation. The following is a bit of a journey of the history of accounts that have been offered. A criterion of adequacy is that an account of function must also be able to account for dysfunction and (importantly) distinguish both of those from dif-functions (different functions).

One way of naturalizing normality vs abnormality is to employ the statistical notions. If normal just means statistical average then there are objective facts about normality or abnormality that are non-normative or statistical. There is controversy over whether mental retardation is merely a statistical notion or whether there are mechanisms that result in mental retardation. While intelligence is described on a bell curve there are clusters of people who are found at the low end of the range and it could be the case that certain mechanisms are responsible for the clustering. Hypertension is also a controversial example. It is unclear whether we are best to think of hypertension as a disorder that is defined in terms of heart rate at the high end of the statistically normal range or whether we focus our attention on these people because they are prone to disorders or dysfunctions. It seems clear that only some statistical abnormalities are relevant and it seems possible that in the cases that are statistically deviant their status as a disorder is dependent on something other than the fact that the conditions are statistically deviant. It would seem possible that the entire population could suffer from parasites or broken limbs, for example, yet the fact that such conditions were statistically normal would not seem to change our intuitions about the conditions pathological status.

The problem of offering a naturalistic or scientifically respectable account of biological function and dysfunction has long been a concern for philosophers and for philosophically inclined biologists. In the 1960's a number of philosophers attempted to naturalize talk of function and dysfunction in biology by appealing to evolution by natural selection as the naturalistic process that fixes functions and dysfunctions. The thought is that if talk of 'function' and 'dysfunction' in biology can be successfully translated into talk of evolutionary functions and dysfunctions then biologists use of the terms are unproblematic from a scientific point of view.

The basic thought here is that evolution by natural selection requires only three things. The first is that there be variation in some trait. The second

is that there be competition for resources such that some variations of the trait result in greater relative fitness than others. The third is that that variation in the trait be heritable such that the offspring are more likely to have the variant of their parents than the other variants in the population. The thought is then that if that obtains evolution by natural selection could result in the population coming to be fixated on certain variants of the trait. Having a heart that pumps blood, for example, could be such that individuals that lacked that variant would be at such a disadvantage that it would surely make sense to say that they had a dysfunctioning heart in the evolutionary sense.

The thought is that there are types of organisms in an environment that have variation in the tokens they have of some kind of trait. If some variants of the trait result in greater relative fitness than other variants and if that trait is more likely to be inherited by future generations of that organism then that trait can come to proliferate over the other variants. In this way certain features can come to be fixated in organisms (e.g., the majority of people are born with hearts and brains and limbs etc).

Three things are often thought to be required for evolution by natural selection to occur. Firstly, there needs to be a trait that has different forms or variants. Secondly, some forms or variants need to be 'better adapted' to their environment such that they result in that variant surviving better than others. Thirdly, there needs to be a mechanism of heredity such that the 'better adapted' traits survive better than others in the sense that the proportion of variants alters due to the offspring of one variant being more likely to resemble their parents than the parents of others. If these three things obtain then evolution by natural selection will occur - which is just to say that the relative frequency of the traits will?

The notion of biological function has inspired a great deal of controversy within philosophy. Some theorists maintain that appealing to the process of evolution by natural selection is not enough to fix biological functions and malfunctions. One would need to add a clause to the effect that eyes enable us to see under 'normal conditions', for example, but it is unclear how one is supposed to go about specifying 'normal conditions'. One can't simply appeal to conditions that were statistically frequent in our evolutionary pasts, for example, as it would have been dark around half the time and yet we wouldn't say that a person had a malfunctioning eye if they couldn't see in the dark.

Evolutionary theories appeal to certain kinds of special causal properties that are meant to fix functions and dysfunctions. The idea here is that evolutionary processes fix 'natural' functions. Or that there is something special about these particular causal properties and processes. The idea is that whatever it was about past tokens that resulted in their outperforming

alternative variants such that they proliferated is the function.

> M has the function of causing behaviour B* iff 1) M has been naturally selected in virtue of causing B*

Critics have rightly pointed out that M could have been selected for B in our evolutionary past, but be maintained in current populations in virtue of causing C. One example of this would be that the mechanism that subserve language were selected for one function in our evolutionary past, and yet they seem to have an acquired function of subserving language now so that if language was impaired due to their failure this would be a genuine instance of malfunction. Wakefield responds to this objection by clarifying the role of evolutionary history by natural selection:

> an effect is a function only if it plays a continuing role in explaining the maintenance into the present generation (i.e., continued existence) of the mechanism in the species. A former function that ceased exerting selective pressure long ago is not currently a function because it has no role in explaining current species-typical structure (Wakefield, 2003 p 979).

Thus Wakefield's revised view thus seems to be that:

- M has the function of causing behaviour B* iff
- M is maintained in the population (by natural selection) in virtue of causing B*

The biological notion of function is thus thought to be fixed by objective facts about the mechanisms and facts about evolution by natural selection.

Thus, according to Wakefield a clinician is justified in maintaining that X is mentally disordered iff

- (1) The clinician judges that according to the best theory of B*, B* is caused by a malfunctioning mechanism

'The HD analysis is an analysis of the concept of disorder, not a theory of the mechanisms or dysfunctions underlying disorders' (Wakefield, 2003, p.978). Systemic capacity theories think that there is a notion of function and dysfunction in physiology (for instance) that is scientifically respectable but not evolutionary. The idea here is that there is a sense in which Harvey understood the function of the heart centuries before Darwin. In other words, the discovery that there was a closed network of arteries, capillaries, and veins and that the heart served to pump the blood around this closed network told us something important about the function of the heart (to pump) without appealing at all to evolution by natural selection.

1.5 Biological wrongness and the division of labour

1.5.1 Natural norms

One of the features of the natural world that biologists have been charged with explaining is the adaptedness of a number of its features. For instance, birds have wings which are adapted to flight and fish have gills which are adapted to underwater. Eyes seem adapted to seeing. Or consider Darwin's Finches where different beaks are adapted to support different foraging patterns. Adaptedness was traditionally explained by appealing to the intentions of an intelligent designer God. Creationists commonly cited this feature of the natural world as something that could only be explained by appeal to the intentions of an intelligent designer creator God. One of the successes of modern biological science has been the provision of an alternative, naturalistic account of how evolution by natural selection could result in the emergence and fixation of adaptive traits.

Since evolutionary processes have been useful to explain adaptive morphological features it seems only natural to suppose that they might also be useful to explain psychological features. Evolutionary accounts of the evolution of mind proliferated. Several assumptions of Evolutionary Psychology (Cosmides and Tooby) were that the mind is massively modular. While there is no one characterization of modularity that all theorists would agree are central to modules most regard informational encapsulation to be central even though encapsulation may be a property that comes in degrees regarding how accessible information is to more or less channels.

Once upon a time in our deep evolutionary pasts there weren't any norms. Today there are. One might then have a view whereby a successful naturalization project involves telling the story of how these norms emerged. On the other hand theorists have noted that you 'can't derive an 'ought' from an 'is" or, alternatively, that it doesn't follow from any description of the way things are any prescription of the way that things should be. We might thus have serious concerns about evolutionary accounts of norms insofar as they attempt to derive prescriptive conclusions from descriptive premises.

Naturalistic theorists are fond of drawing our attention to naturally emergent properties such as that of liquidity. It might not seem obvious that liquidity would inevitably derive from certain arrangements of atoms – and yet this is precisely what liquidity is. Similarly, it might not seem obvious that norms arise from certain kinds of naturalistic processes and yet it might turn out that that is precisely how certain norms came to be.

Here theorists are fond of attempting to distinguish between hypothetical and categorical imperatives. Hypothetical imperatives are of the sort 'if you desire to maximize y then do x'. They say that these can successfully be naturalized. Categorical imperatives of the sort 'do x' can't be naturalized, however. Let us suppose (for the sake of argument) that categorical imperatives cannot be naturalized. One might well say 'so much the worse for categorical imperatives' (e.g., there aren't any such things insofar as they can't be naturalized). One might say that it turns out that the best candidates for categorical imperatives aren't really categorical after all.

What about bio-medical and mental disorders? What is good for a person? Health? Flourishing? Ideal views. Consider plants. Nutrient deficiencies. Death. Causes of death. Suicide. Coma etc (anorexia).

It has long been noted in philosophy that you 'can't get an 'ought' from an 'is" or, alternatively, that it doesn't follow from any description of the way the world is what (if anything) should be done. Or, again, it doesn't follow from any description of the world what prescription there is for the way forwards. This idea doesn't sit very well with naturalization, however. The whole idea of naturalization is to show how norms arose.

Fulford talks of 'event horizons' or places where theorists have introduced normativity in their account of disorder. While different theorists use their terms slightly differently (and indeed different theorists often focus on slightly different terms) most have some idea of which notions they consider to be normative and which notions they consider to be non-normative. And where in their little nested hierarchy they place the event hierarchy for the introduction of norms or values. For instance, we might think that mental and bio-medical disorder are normative notions but that they can be grounded in the non-normative dysfunction – in which case the event horizon occurs between those terms. An even horizon must occur somewhere, however.

Kantians think that norms or values can't be naturalized for something like the above considerations. Other theorists think that they can be naturalized, however. What we have is some kind of emergence. For example, liquidity is an emergent property that can be explained by the nature and interaction between smaller bits that comprise the liquid that are not themselves liquid. In this way values can be naturalized by appeal to something logically or perhaps causally prior that is not itself normative.

We need to consider again the work that the dysfunction criterion is meant to do: We start with a candidate phenomenon. We then want to know whether the candidate phenomenon is really a disorder or not. If there is a biological dysfunction within the individual then yes, we have a case of disorder. If there is not a biological dysfunction within the individual then no, we do not have a case of disorder. There are different ways in which

we can query the above account. The first is to maintain that it is *over inclusive*. This is to say that some biological dysfunctions that cause some normative violations are not disorders. The other is to say that it is *under inclusive*. That is to say that some disorders are caused by other things. It isn't necessary for disorder at all. Some disorders don't involve dysfunction. Dysfunctions (perhaps) come far too cheap.

Sometimes the attempt to naturalize function and dysfunction is characterized as an attempt to naturalize norms. The idea here seems to be that function and dysfunction are normative notions but that a successful naturalization of them will be an explanation that appeals to purely natural (non-normative) properties and processes. Other times the attempt to naturalize function and dysfunction is characterized as an attempt to show these notions to not be normative insofar as a successful naturalization of them will be an explanation that appeals to purely natural (non-normative) properties and processes.

Sometimes evolution by natural selection is claimed to be the naturalistic process that naturalizes the notions of function and dysfunction. The thought seems to be that insofar as evolution by natural selection can fix functions and dysfunctions of biological phenomena (e.g., eyes, hearts etc) these norms have been successfully naturalized. Other times the success of the above is claimed to have shown us that evolutionary functions and dysfunctions aren't normative after all. The idea here is that they are purely naturalistic, respectable phenomena. Either way, we can see how the notion of evolutionary function and dysfunction is supposed to legitimize our regarding people to be functioning or dysfunctioning in a way that isn't problematic. There are mind-independent scientific (e.g., biological) facts about function and dysfunction and it isn't a matter of our expressing disapproval or moral indignation etc.

The notion here is that we begin with some feature of biological systems that we would like to explain. We might want to explain vision, for example, or the circulatory system. What we then do is discover how there are mechanistic components that contribute to the explanandum. In the case of vision we discover that there are parts to the eye (e.g., the cornea and the lens) and that they each seem to contribute differently to the explanandum - vision. The systemic notion maintains that the functions are fixed by the contribution that the component part makes with respect to the relevant output of the system that was our initial explanandum. This is thought to account for functions and dysfunctions in physiology in particular where physiologists often make no reference to evolution by natural selection.

The obvious move to make is to maintain that functions attach to trait types rather than trait tokens. A token of the type can thus be a functional

token or a dysfunctional token. Davies has argued that in order to do this we need some independent way of characterizing types.

Cummins offered his systemic account of function as an analysis of what was going on in at least some areas of physiology. While evolutionary biologists may at times make use of an evolutionary notion of function it seemed clear to Cummins that there was a notion of function in play that didn't explicitly make reference to evolutionary considerations and he attempted to analyze this. Many theorists have found Cummins notion of systemic function to offer a plausible analysis of function talk in physiology in particular. One might thus think that this notion of function might be more relevant to medicine and to psychiatry.

There has been much controversy over whether Cummins has offered a genuine rival to the evolutionary account of function. One might consider something like an ecosystem, for instance, and then take the systemic approach by attributing functions to components of the ecosystem such as clouds and predators etc. Theorists have argued that there needs to be some non-arbitrary way of fixing the relevant systems. Systems can't be arbitrary mereological fusions, for instance. Thus one way of restricting the range of systems that the systemic notion employs is to use evolution by natural selection. Similarly, one might argue that the evolution by natural selection is more fundamental than systemic analysis because the systemic analysis only works in virtue of evolution by natural selection operating over the systems.

This alternative view of function arose as a development of the work of Cummins. The basic idea is that there seems to be a sense of biological function that is not essentially historical. In order to see this we just need to consider the obvious truth of the claim 'Harvey understood the function of the heart centuries before Darwin'. The idea here is that when Harvey came to understand that the heart functioned as a pump within the circulatory system we learned something about the function of the heart even though we didn't learn anything at all about the general or specific historical processes that have resulted in hearts. While for the evolutionary theorist the questions 'how did x come about' and 'what is the function of x' are to be given the same answer for the systemic theorist these questions come apart while the questions 'what is the function of x' and 'what role does x play in some greater system' are equivalent.

On the systemic capacity account functions are assigned to components in virtue of the role that they play in the production of an output in some greater system. If one wants to give a systemic account of some trait or variation on a trait then firstly one appeals to some system that produces the phenomena that one wants to explain. Once one has the relevant system then one proceeds to analyze the system into components and assign func-

tions to the components in virtue of the role they play with respect to the production of the phenomena that one wants to explain. Davies maintains that it is important to note that assignment of function to components is relative in two respects. Firstly, which components are relevant is going to partly depend on what phenomena the researcher is interested in offering an account of. Secondly, which features of the components are functions is going to partly depend on what phenomena the researcher is interested in offering an account of. Despite these two aspects of the systemic capacity view being partly determined by the interests of the researcher Davies maintains that there are also several features of systems that are not dependent on the interests of the researcher.

We have already seen that systemic capacity analysis involves appealing to two distinct levels. There is the level of the phenomena and the system that produces it and there is the lower level with the components and their functions. Davies maintains that systems must consist in two distinct levels and once we hit a level at which the outputs are basic where the 'system' cannot be analyzed into further components then we have reached the end of the systemic capacity chain of explanation. Aside from this bedrock we can often reiterate the systemic capacity framework down - explaining the workings of the circulatory system, the heart, certain kinds of tissue, certain kinds of cell, and so on.

Davies enumerates the systemic view as follows:

- (1) I is capable of doing F,
- (2) A appropriately and adequately accounts for S's capacity to C,
- (3) A accounts for S's capacity to C, in part, by appealing to the capacity of I to do F.
- (4) A specifies the physical mechanisms in S that implement the systemic capacities itemized in A.

While the majority of theorists attempt to show systemic capacity functions to be grounded in evolutionary functions or to show that the different accounts are involved with different explanatory projects Davies argues that evolutionary functions turn out to be a certain kind of systemic capacity functions. He maintains that by viewing a population as a system and viewing members of a population as constituents of the system we can offer a systemic capacity analysis of the phenomena that we want to explain in a way that captures all the verdicts of the evolutionary view. He also maintains that it is an advantage of the systemic capacity view that it can help us

understand what is going on with evolutionary explanation or modelling of other phenomena that the evolutionary function view can't explain such as drift. Davies seems right to be putting pressure on the evolutionary view to move from the adaptation assumption to other phenomena that is of evolutionary 'interest' even if it isn't straightforwardly explained by evolution by natural selection. While it might be thought to be a feature of Davies view that it can be applied to phenomena that the evolutionary view (at least in its simple version) can't explain it might be thought to be a vice of the view that it is over-inclusive. While Davies has no trouble applying the view to artefacts that produce things such as assembly lines intuitions are divided as to whether we want a unified account of artefacts alongside biological phenomena. While Davies briefly considers Godfrey-Smith's concern that the systemic capacity and evolutionary views are both important because they pick out importantly different causal chains at different levels of analysis he moves on from the objection and doesn't consider it further. I don't see the problem in reserving the term 'proper function' for solely functions arising from evolution by natural selection or for solely functions arising from the historical analysis of biological phenomena. Hard to know where to draw the line on mental phenomena but hard to distinguish psychiatry from neurology at any rate.

One problem with attempting to ground medicine and psychiatry in the systemic rather than the evolutionary notion of function is that the systemic view (as enumerated by Cummins, anyway) doesn't allow us to differentiate dysfunction from the absence of function. Cummins enumeration is that the function of some part mechanism x is fixed by the causal contribution makes towards the output of the system. So the function of a heart valve might be (roughly) to regulate blood flow as the casual contribution the heart valve makes to the hearts pumping of blood is to regulate blood flow. The trouble is that if the valve fails to regulate blood flow then this view doesn't provide us the resources to say that the valve is malfunctioning. This is because if the valve doesn't play that causal role then it simply fails to have that function rather than it dysfunctioning.

Other theorists have attempted to develop the systemic notion of function in such a way that it can account for dysfunctions. One way of going about this would be to make use of a type and token distinction. On this view the function of the heart valve is to regulate blood flow because this is what mechanisms of the valve type do with respect to contributing towards the hearts pumping of blood. Because the functions are type-functions rather than token-functions a token valve that failed to regulate blood flow could be described as malfunctioning because it is not playing its type function.

The problem with this view is that we need some independent way of

stating how tokens get to be members of a type. If we are attempting to explain how the type has the function that it does then we can't say that a token is a member of a type in virtue of exhibiting the type function because part of what we are trying to explain is how the types have their function. We don't seem to have grounds for saying that a token is a dysfunctioning member of a type rather than saying that insofar as the token doesn't play its usual contributory role it fails to be a member of a type. And thus it lacks a function rather than dysfunctioning.

Davies develops a systemic view of function and he simply acknowledges that it doesn't have the resources to handle dysfunction talk - but then he maintains that the evolutionary notion can't adequately account for dysfunction either so that is no reason to adopt the evolutionary notion over the systemic notion. While it might not be an adequacy constraint on function talk in general that it can account for dysfunction it does seem that insofar as medicine and psychiatry attempt to ground their subject matter in dysfunction an adequate account of medical functions must be able to account for dysfunctions, however. If Davies is right that neither the systemic or the evolutionary notion can allow for dysfunction then this will have very skeptical implications for medicine indeed. While Davies does talk about medicine a little he doesn't seem to realize the role that dysfunction talk is supposed to play with respect to grounding medicine and psychiatry in particular in the natural sciences. He thus doesn't realize how significant his finding that neither can account for dysfunction would be with respect to medicine and psychiatry.

The obvious way to provide an account of dysfunction is to see functions as properties of types. On this account a type (e.g., hearts) have a functional property (e.g., functioning as a pump). Particular tokens or instances of hearts can thus be functional or dysfunctional hearts depending on whether they function as a pump or not. What is needed for this style of account is for there to be properties that are sufficient to make a particular instance a member of the kind but where the functional property itself is not needed in order for the instance to be a member of the kind. If the functional property was needed for kind membership then we wouldn't have dysfunctioning hearts because an alleged heart that didn't have that property would not be a heart after all.

In defending the systemic account of function as being primary (where evolutionary functions are thought to be a subset of systemic functions) Davies offers an argument that seems to create a problem for naturalistic accounts of function more generally. While we considered briefly above that a number of theorists think that attempts to naturalize function are doomed to fail because of problems with dysfunction being normative and biology

being non-normative Davies maintains that while it is commonly thought to be a virtue of evolutionary accounts of function that they can offer a naturalistic account of dysfunction he maintains that evolutionary accounts fail to do so. It is thus no objection to the systemic view that it cannot either. Davies doesn't seem to explicitly consider the role that dysfunction has played in attempts to naturalize bio-medical disorder. As such it is hard to know whether he would be happy or unhappy with this implication of his view. It is worth considering whether one can get dysfunctions out of the evolutionary notion of dysfunction (or an alternative naturalistic account of systemic). This will better help us understand the role and limits of sciences contribution to fixing what conditions are bio-medical disorders.

Davies has fairly recently raised a couple of objections to the evolutionary view that are worth considering. If Davies objections are well founded then there would seem to be significant problems with appealing to the evolutionary view that haven't been properly unpacked in the literature. Davies notes that one of the great appeals of the evolutionary function view is that theorists have done much in order to show that it can provide an account of malfunction or dysfunction. The main objection to the systemic function view is that the systemic view does not have the resources to account for dysfunction. Davies maintains that despite this common wisdom the evolutionary view is not able to provide an account of dysfunction. He maintains that as such it is no objection to the systemic view that it cannot. If Davies is right that neither the systemic or the evolutionary view can offer accounts of dysfunction then this will create a significant problem for the two-stage view insofar as the appeal to evolutionary and / or systemic functions is supposed to provide an account of biological dysfunctions which is supposed to be what science discovers about psychiatric and mental disorder. Davies also maintains that the evolutionary view can be shown to be a particular kind (or variant on) systemic capacity functions. The main objection to this line has been that one can account for dysfunction but the other cannot. This part is less relevant for here. If we can't get dysfunctions then that seems very problematic. Davies does not seem to have considered the implication of this or how theorists have attempted to use the terms in medicine and psychiatry.

Firstly, the notion of a 'set point' or 'set point range' is introduced. The notion is attributed to Cannon. The thought was that in studying cells he noticed that the internal temperature of the cell was fairly invariant to change despite the alterations in external temperature. He noticed that the internal temperature tended to not move much around a fixed point. The average???? Was the set point. The degree of variation is the set point range. What happens when the internal temperature varies outside the set point

range? Sickness and death. Whatever staves off death. Whatever preserves the characteristics of life. Important to note that thermoregulation isn't a characteristic of life, but perhaps there are subsidiary functions that are required for those characteristics to be present. Death seems to be an objective measure (while there is trouble characterizing death 'around the edges' we have a fairly intuitive understanding of the notion in the majority of cases). 'Health' or 'well functioning' is harder. There seems to be another level of functions (of which thermoregulation is one). Other things thermoregulate of course and in explaining this a little more we turn to considering the philosophers favourite example of a thermostat.

We are now in the position to see how anatomy, physiology, levels of analysis, characteristics of life, the notion of a set point are inter-related from the perspective of anatomy and physiology. While anatomy and physiology come apart they seem to relate to and constrain each other in important ways. While philosophers often think that H2O could have a different chemical constitution and play the same qualitative role it is interesting to note that the question takes on a different problematic aspect when considered from the perspective of the properties of the atoms and how they confer properties on the molecules which in turn confer properties on the observable interactions. They are more tightly bound from the perspective of different levels in science than philosophers have often supposed with variations to lower levels not conferring much in the way of change to variations at higher levels. Philosophers who have taken the sciences seriously seem less inclined to multiple realizability intuitions. Might be that they are missing something of philosophical importance here or might be that philosophers are missing something about what scientists have to show us about the way the levels are related in science (much less multiple realizability for scientific kinds than philosophers have supposed).

Despite Wakefield's taking the evolutionary approach to be the only scientific game in town, the evolutionary view is not without its critics and we are far from a consensus on the correct analysis of function and dysfunction talk in biology, general medicine, or psychiatry. The systemic capacity view provides another way of understanding function talk in biology. The systemic capacity account is different from the evolutionary view in that it makes no essential reference to historical processes and instead attempts to ground functions in component capacities of systems. While the systemic capacity account originally offered by Cummins did not have the resources to account for dysfunction many theorists have thought that the approach could be adapted so as to do so. While evolutionary theorists often consider the main virtue of the evolutionary approach to be that it provides a naturalistic account of dysfunction this has recently come under fire by Davies who

defends a modified version of the systemic capacity view. He argues that the evolutionary view is not really an independent theory and that evolutionary functions turn out to be a particular kind of systemic capacity function. He also argues that neither the evolutionary nor the systemic capacity view have the resources to offer a naturalistic account of dysfunction. If this is correct then it seems that we are left with a significant problem. In particular, if Davies is correct that neither view has the resources to offer an account of biological dysfunction then this would seem to undermine the two-stage views assumption that the role of science in bio-medicine and psychiatry is to discover facts about biological dysfunction. In what follows we will consider both the evolutionary and systemic capacity views of function and then turn to problems that each view has in providing a convincing account of function. We will then consider how each fares with respect to providing an account of dysfunction and end with some thoughts on the role of science in discovering facts about bio-medical and psychiatric disorder.

Thus far most of the discussion has focused on the problem of fixing functions and very little has been said about dysfunction or malfunction. Something clearly needs to be said about dysfunction as there is a third option that any theory must be able to rule out: the problem of distinguishing the functional from the non-functional from the malfunctional. While some theorists might not consider it a criterion of adequacy on a theory of function that the theory have the resources to account for dysfunction (as opposed to dif-function or non-function) given the role that dysfunction is supposed to play in medicine and in psychiatry being able to account for dysfunction must be a condition of adequacy on any account of function that purports to be relevant for general medicine or psychiatry.

There seem to be two ways that we can approach the problem on the evolutionary view. The first is to consider traits where the idea is that traits are binary (all or none) and mutually exclusive. On this view where we have a case of stabilizing directional selection for one trait then we might consider we have the best case of selection against the other traits and thus the other traits are dysfunctional. A similar alternative would be to consider there to be different values within a variant. Similarly, where we have a case of stabilizing directional selection for one of the variant (or values of the variant) then we seem to have the strongest case of selection against the alternative variants or values. Both of these seem to amount to a similar thing. The main issue that arises here is how we individuate or type traits or variants on traits. We have already considered above the considerable problems that arise when we try and assign function to traits or variants of traits. Similar issues seem to arise when we try and assign dysfunction to traits or variants of traits. But perhaps this whole approach to the issue is misguided. Maybe

what we really need is a type and token distinction.

The usual way that evolutionary theorists talk about dysfunction is to distinguish between a type that has a function (or a variant that has a function) and particular tokens of that type or variant that lack the function. On this account the function of the type heart is to pump blood because pumping blood is what resulted in past hearts proliferating such that there are token hearts now. A heart can malfunction by not pumping.

Davies objects to the above characterization maintaining that the evolutionary view does not have the resources to account for malfunctioning instances of types. Davies argument for this is that evolutionary theorists individuate types according to their functions. Since the functions are thought to be necessary and sufficient for membership in the type it is thus impossible for an instance to both be a member of the type (possess the necessary and sufficient condition or function) and yet lack the function and hence dysfunction. Davies thus maintains that instead of a heart malfunctioning all the evolutionary view gives us the resources to say is that the instant that does not pump is not a heart after all and thus it doesn't have the function to pump and thus is isn't malfunctioning or dysfunctioning so much as lacking the function that we wanted to assign.

Davies argument relies on the evolutionary theorist individuating types according to the function that the theorist assigns to the type. Insofar as types possess their function as a matter of necessity he seems correct that a instance of a type cannot malfunction. Despite his maintaining that this is the way that every evolutionary theorist has individuated types it seems that there is another way that seems more licensed by the evolutionary view. He also admits that evolutionary theorists individuate types according to their aetiology. This seems naturally at home with the evolutionary view and it is important to note that the best theory we have of species membership is aetiological rather than morphological (where morphological might be thought to be more in line with the systemic capacity view). Davies argument that the evolutionary view does not have the resources to account for dysfunction relies on traits being types according to their function. The problem is basically that *if* traits are typed according to their function *then* an instant that fails to exhibit the function fails to be a member of the type and hence we do not have the resources to say that the instant is a malfunctioning or dysfunctioning member of its type. Davies claim seems correct in the sense that if having some function F is both necessary and sufficient for F's being classified as a member of the functional kind K then if F were to lack the necessary and sufficient condition for being a member of kind K then it would simply stop being a member rather than being a dysfunctioning member. By analogy if we consider an instant of gold and we then apply a proton gun

and remove one of the protons then the instant isn't a malfunctioning or dysfunctioning or abnormal instance of gold in virtue of having one less proton. Rather, the thing to say would be that the instant that was a member of the kind gold is no longer a member of the kind gold - rather it is a member of kind (whatever has one less proton than gold).

In response to Davies objection one needs simply note that it will not do to individuate kinds functionally rather some other criteria must be used for kind individuation. While Davies writes that all evolutionary accounts appeal to functional kinds there is an ambiguity with respect to what is meant by 'functional kind' here. In particular, by functional kind one could simply mean 'kind with a function' where the conditions for kind membership come apart from the function that is attributed to members of the kind.

Earlier we considered four broadly different approaches to accounting for function and dysfunction including: teleological, bio-statistical, evolutionary, and systemic. We can also consider different notions within each of those as there are different ways of specifying statistical normality / abnormality, for instance. I am not wanting to get caught up in the issue of whether these are simply different notions or whether these are rival accounts of the same notion. What seems important is which (if any) version of any of these notions can be made to work to do the work that is required of it for psychiatry.

In order to assess which (if any) can do the work that is required we need to return to the issue of what work is required. The idea seems to be that firstly we observe that behaviour is deviant or problematic (in some yet to be specified way). This isn't sufficient for disorder, however. It is widely thought that if that was all there was to it then psychiatry would in fact be illegitimate. What makes psychiatry legitimate, however, (the practices of treating and incarcerating including involuntarily) is that there is some relevant dysfunction that is the cause of the deviant behaviour.

While theorists can agree that we can define up any notion of function and dysfunction we want they can simply say that this doesn't impact upon what notion is employed in medicine / psychiatry or (perhaps more importantly) the issue of which notion should be employed in medicine / psychiatry. Theorists are fond of maintaining that there is a 'natural' notion and it is this notion that is important. For instance, sometimes we read of theories of 'natural norms'. We can ask what this theory of 'natural norms' is supposed to be.

1.5.2 The proposed division of labour

Theoretic science. Proceed independently from the clinical science and independently from normative concerns. Clinical laboratory science. Progress in genetics, neuroscience etc.

Clinical science. Proceed independently from the clinical science and the normative concerns insofar as it focuses on hypothetical imperatives if you want to stop x behaviour then provide y drug.

Ethics. Proceed independently in asking whether we are justified in intervening against will and the rights duties etc.

Each of these three aspects considered to be fairly independent with not much of a meeting. Instead of this view I think we need to see them as being integrated. Or: If we want the findings of one to be relevant to the findings of others (as I think we do) then we need to see them as being integrated from their foundations. It won't do to have theorists utilizing radically different notions.

Rethinking the division of labour: The problems as explanandum.

Murphy maintains that the malfunction assumption does for psychiatry what the adaptationist assumption does for evolutionary biology. He goes on 'which is to say that sometimes the malfunction assumption is false, sometimes we don't know whether it is true or false but that does not impugn diagnosis'. One thing that concerns me about the malfunction assumption, however, is that it is supposed to be what grounds psychiatry as a non-evaluative science and that it seems to recommend a methodology for modelling mental disorders. The methodology seems to be that we model 'normal' or 'functional' biological or psychological processes and then we explain disorders by appealing to breakdowns in the model. Much work in the cognitive neuro-sciences and the bio-medical sciences has been done utilising this approach. We have explanations that characterise delusions as being the result of some kind of breakdown in belief formation and / or retention mechanisms; we have explanations of autism as a theory of mind deficit and so forth. The malfunction assumption can't make much sense of other projects that have been done, however. Instead of working with the malfunction assumption some theorists have worked with a function or adaptationist assumption where certain traits (such as histrionic or psychopathic) may be modelled as evolutionary adaptive strategies. Some theorists have attempted to characterise disorders such as depression, schizophrenia, and anxiety as evolutionary adaptive strategies that result in harm in present environments because environmental circumstances are far removed from those

in savannah life.

While I'm not going to look at the plausibility of particular theories that have been offered my main point here is that the malfunction assumption does not seem to be required in order for us to study mental disorders scientifically. Instead of attempting to model mental disorders as deviations from some standard one could simply describe the causal processes that seem relevant for some behavioural output while remaining neutral on whether that behavioural output is adaptive or maladaptive. Science can thus model the causes of certain kinds of behavioural symptoms even in the absence of the malfunction assumption. What seems harder to do in the absence of the malfunction assumption, however, is to say what it is about certain conditions or people that means that they are disordered.

Attachment and psychodynamic (neurobiology of shared emotion, attachment). Telescope sciences instead of microscope sciences. Where are we headed as a society? What do we value? Rising populations support greater division of labour. Greater division of labour allows for super-specialists who maybe aren't so much generalist. Idea of neuroplasticity and plasticity (adaptiveness) as being something very specific. Maybe in terms of individual development... Acquiring specific skills. But maybe individuals can then be relatively fixed. See what happens when segments of the economy collapses. E.g., a meat-works closes. How well do those workers fare at alternative employment? Specialist vs generalist.

The issue of normative violation / harm seems important insofar as this seems to be epistemically the first step. We have this intuition that there is something wrong with the individual and we want to know what is wrong. Is it that they are mentally disordered or they are bad or what is going on. Sceptics maintain that we start with this intuition and then we 'cast about for something to medicalize' and it does seem that something like this is what is going on. We then find something – a candidate for dysfunction. The problem with accounts of dysfunction is that they seem very liberal indeed – fairly much everyone would be dysfunctioning in some way. There is a story to be told. Such stories seem to come cheap. Fairly persuasive. Of course one might maintain that this is so but that things really are very much tighter than that. We are mistaken...

Many theorists have the intuition that there are objective facts about who is and who is not mentally disordered that are for the natural sciences to discover. Geneticists and neuroscientists, for example, are working to discover the relevant facts so we are better able to diagnose or identify individuals with disorders. So we are better able to prevent the development of disorder and to treat it effectively. Newspaper headlines proclaiming 'science has discovered the biological basis of schizophrenia) are surely premature but the

idea is that there is such a biological basis to be found and it is only a matter of time. If we aren't sure whether we are sick or not we can go to a doctor who can do some diagnostic tests and perhaps send away for some and tell from the results data whether or not there is anything medically wrong with us.

Fulford (2000, p. 79) states that:

> Medicine has been successful precisely through its identification with science. No one wants to be a loser, therefore. Everyone wants to join the winning team. Psychiatrists (such as Kendell 1975) want to naturalize mental illness in terms of disease so that they can join the medical team of medical science; medics (such as Campbell, et al. 1979) want to naturalize disease in terms of dysfunction so that they can join the winning team of biological science; biologists (Allen, et al. 1998) want to naturalize dysfunction in terms of function so that they can join the winning team of natural science; natural scientists, on this model, are the winning team.

In considering evolutionary models to be dynamic or temporal such that they capture population dynamics over time the possibility arises that psychiatric disorders may have been adaptive in the evolutionary sense at some point in the past but that due to recent alterations in our environments they are no longer so. This notion brings out an interesting point that variants are only adaptive in relation to other variants and also in relation to the environment that the variants are in.

Homo Sapiens have radically reconstructed their environments. We build buildings, work in high rise office blocks in front of computers, live in inner city apartments, and negotiate public transport systems or navigate our own motor vehicles. We have complex social structures of banks and educational schools and government departments. What is required in order to be considered 'not significantly impaired in ones social, occupational, or educational functioning' are obviously different in at least some respects from the conditions our ancestors operated under in the plestocine. While a phobia of falling might significantly impair one who is expected to live and work in high rise apartments and fly around the world for business and / or family a phobia of falling might have positively benefited those who lived in environments where heights were usually cliffs where wind was particularly strong. Similarly, while social aloofness and communication with non-apparent things might impair ones negotiation with social services, potential employers and educators such behaviour might result in a very different outcome in a society

where such behaviours result in a person being revered as a holy prophet or healer with special gifts.

Cooper's gets us to think the about the notion of a weed where a weed is (roughly) an unwanted plant. The notion of a disease or a disorder seems to be like this. While there can be a science of plants there can't really be a science of weeds. There isn't anything that weeds have in common that is aside from our interests. Medicine is similar to this. It is interest dependent. Consider cancer from the point of view of the cancer or disease as being centred on the organisms welfare rather than our own. Medicine and psychiatry is indeed importantly dependent on our interests and values. With respect to morphology and with respect to behaviour. Psychiatry is more about behaviour than medicine. It seems more intentional. We want to distinguish those with different sorts of intentions. This is seriously problematic. The notion of mental disorder is primarily a moral notion. Secondarily law and medicine. Different sorts of intervention on incapacity – on failure to do that which we think people should . Of course more must be said about the relevant norms. But the point is that more must be said about the relevant norms not that we are waiting on evolutionary biologists discoveries about which conditions are caused by dysfunctions.

Chapter 2

The second 'small chunk'

A small chunk received by Kim Sterelny 2013

2.1 Some assumptions of evolutionary psychology

2.1.1 Modularity

It is now typically accepted that there are a number of features that have traditionally been associated with modularity and that more or less of those features might be present to more or less of a degree for different cognitive or neurobiological structures / mechanisms. Some of the features of modules include informational encapsulation, localization, etc. We can talk of the contents / functions of the modules / mechanisms. Whether they are functioning correctly / representing or dysfunctioning / misrepresenting.

The interest in evolutionary functions as fixing function and dysfunction objectively (non-normatively) has been motivated by accounts that try and offer naturalistic explanations of mental content / representation. The idea is that mental content / representation can be fixed by mental functions / dysfunctions. If mental functions / dysfunctions are objective then we see how the naturalization works. An alternative to this kind of a line is a disjunctivist line that there aren't representations and misrepresentations rather there are just two different kinds of state. In virtue of what is something dysfunctioning rather than differently functioning?

2.1.2 Behavioural, cognitive, neuro-biological, genetic mapping

There has also been a growing realization that mapping different levels or layers of explanation is far from trivial. It was initially thought that the kinds of folk psychology (e.g., 'beliefs', 'desires' etc) would map fairly straightforwardly to neurophysiological states. Or that behaviours (e.g., 'anger', 'key press') would be fairly straightforwardly mapped to neurophysiological states. Or that beliefs and desires would fairly straightforwardly map to behaviours (e.g., reports of beliefs would predict voting behaviour). This has turned out not to be the case, however. Genes to morphology / behaviour similarly problematic.

Insofar as the kinds from cognitive psychology will correlate to the kinds from neurophysiology we need to be revisionist about the kinds from one or other or both of those fields. Similarly, insofar as we want a classification system of behavioral symptoms (for instance) to be useful to researchers in the genetics of mental disorder, the neurophysiology of mental disorder etc we are going to need to be revisionist about the classification system.

The DSM purposely tried to classify on the basis of observable behavioural symptoms since there was so much disagreement in attempts to classify on the basis of theory- laden notions. While many have considered it to be a virtue that the DSM by way of focusing on observable behaviours and achieving relative consensus for the field... It has come under serious criticism for sticking to the observable behaviours and not at least attempting to classify on the basis of deeper essences. That which was most useful for consensus has been least useful with respect to theoretical development, scientific discovery, and (perhaps most worryingly) treatment.

While the DSM was supposed to unify and it did a relatively good job of it perhaps the time has come to move to the second stage of science – that of theory development. The National Institute of Mental Health has rejected the DSM as a useful classification system for the purposes of research. They have provided their own categories for research and these are intentionally supposed to cross-cut the DSM categories.

While the DSM provides a nosology where clinicians identify mental disorder on the basis of behavioural symptoms it would seem to be a separate issue whether mental disorders are constituted or defined by the behavioural symptoms as Behavioural Kinds, however. If one takes the behavioural symptoms to be definitional or constitutive then there could plausibly be borderline cases where it is indeterminate whether the individual is in fact a member of the kind or not. It would seem, however, that the main reason why it is that certain properties are to be found clustered together in nature is because

they share some underlying causal mechanism that are responsible for the properties homeostasis. It is because the causal mechanism is found in the different instances that we are able to make scientific generalisations and predictions. It could also turn out that the same set of behavioural symptoms could be generated in two quite different ways. If we found this to be the case then it would seem better to conclude that there are two distinct kinds of disorders where different interventions are required.

Thus, while we might typically identify or come to believe that instances are members of a certain category on the basis of superficial, observable properties, taxonomy is often revised as we come to define categories on the basis of the underlying causal mechanisms that are necessary for category membership. This is because causal mechanisms seem to be what leads to the properties homeostasis and the more homeostatic a property cluster the more those properties are able to support generalisations and predictions. While Boyd's view focused on internal generative mechanisms it is unclear whether a principled distinction between internal and external generative mechanisms can be sustained. If one views a species as an individual, for example, then lineage would be an internal property to the species. Boyd's homeostatic property cluster view, or something like it, can thus be thought of as consistent with both the essentialist and relational view of categories.

Wakefield attempts to draw a principled distinction between the 'right' and 'wrong' kind of causes for mental disorder. He maintains that when the harmful behaviours are due to inner malfunction the individual is mentally disordered and when the harmful behaviours are the result of external causal mechanisms the harmful behaviours are not indicative of mental disorder and are instead best thought of as a non-pathological problem in living. It would seem that whether mental disorders are constituted by social causal mechanisms would be an empirical matter rather than one to be settled on intuitive grounds or by stipulation, however.

Wakefield is especially focused on the notion of neurological and / or cognitive malfunction which he characterises along the lines of a hardware / software distinction and while he doesn't mention it I don't think he would be opposed to adding genetic malfunction to the mix (supposing that it makes sense to talk of genetic malfunction or kinds of genetic disorder). This way of thinking about inner malfunction seems very much in line with cognitive neuropsychology and it might be the case that the kinds of psychiatric disorder are derived as malfunctions of the causal mechanisms that is identified, at lest in part, by cognitive neuroscientists. Neurological kinds would seem to be fairly straightforwardly thought of as biological kinds. Some theorists have attempted to analyse Psychological kinds as another variety of biological kinds where mental or cognitive states such as belief and desire are the

kind of state they are in virtue of what the mechanisms that support the state have evolved to do.

Sometimes theorists (like Wakefield) appeal to current functions instead of evolutionary functions where the effects of a current function are responsible for the mechanism being prevalent in current populations. Treating mental kinds as biological kinds is controversial, however. The natural categories or kinds would seem to be those of normal functions. Psychiatric kinds are breakdowns of normal symptoms and the breakdowns may be unified only by being breakdowns of a specific mechanism. There would thus seem to be an open ended class of ways things could go wrong. Attempting to list them all with respect to behavioural symptoms is thus bound to get unwieldy and more progress might be made by looking at different ways that normally functioning systems can break down.

> Judging from our experience in internal medicine it is a fair assumption that similar disease processes will produce identical symptom pictures, identical pathological anatomy and an identical aetiology. If, therefore, we possessed a comprehensive knowledge of any of these three fields - pathological anatomy, symptomatology, or aetiology - we would at once have a uniform and standard classification of mental diseases. A similar comprehensive knowledge of either of the other two fields would give us not just as uniform and standard classifications, but all of these classifications would exactly coincide (Kraeplin, (1907). Quoted in O. Reider (1974) 'The origin of our confusion about schizophrenia', *Psychiatry,* 37: 197-208 from Bentall.).

2.1.3 Dysfunctioning module

Disorder vs problems in living

Wakefield maintains that there is an important distinction to be drawn between mental disorders on the one hand and problems in living on the other. Wakefield maintains that an example of a problem in living is illiteracy where a person is illiterate because nobody ever taught them how to read. The idea here is that such a person isn't disordered because their illiteracy isn't caused by an inner dysfunction. Instead the cause of the illiteracy is their impoverished environment. In other words, behavioural symptom profiles alone are not enough to establish the presence from absence of disorder. Causes matter, and the presence of inner dysfunction (to behaviour that is harmful to the organism and / or society) is required for disorder.

Another interesting case that has been raised is the idea of a person-environment mismatch. Murphy and Woolfolk consider an example of a smoke detector that is positioned too close to the stove. In this case the smoke detector gives off a number of false alarms and yet most people have the intuition that the smoke detector isn't malfunctioning (there isn't anything wrong with it) because the problem is rather that it has been positioned incorrectly. One suggestion of phobias or of certain triggers for anxiety is that they were adaptive to certain periods of our evolutionary history and rather now we are in mismatched environments. For instance, having a startle response to loud noises would be adaptive in environments where loud noises reliably signalled dangers. Living in present day city environments repeated startle responses might result in significant problems due to high levels of stress / cortisol etc.

Most of the work in the philosophy of psychiatry has focused on attempts to offer a non-normative value-free foundation for the scientific aspect or critique that this is not possible. Relatively little work has focused on the normative or evaluative aspect of psychiatry. Are the norms that are relevant for bio-medical and mental disorder the same norms as those that are relevant for our (correct) judgments of social and / or moral deviance or are the norms importantly different? When are we justified in intervening to bring the individual in line with the norms rather than intervening to bring the norms in line with the behaviour of the individual? Most of the work has focused on accounting for a non-normative, scientific aspect that converts normative violation into a genuine case of disorder. Not much work has focused on the normative aspect. Most of it has focused on the non-normative aspect. The idea seems to be that the same kinds of norms are operative and the bio-medical / mental disorder vs other kinds of normative violation are due to an additional factor that is present in the medical / mental case that is not present in the other cases. Namely, the presence of a dysfunction that is the cause of the norm violation.

Wakefield claims that the biological sciences have discovered that the relevant causal process for fixing function and dysfunction is evolution by natural selection. He claims that this is the relevant causal process not only for the biological sciences, but also for medicine and psychiatry. There is far from univocal support for the priority of evolutionary function and dysfunction for the biological sciences, however. A number of theorists have argued that a non-evolutionary notion is employed in at least some aspects of biology (See, for example Davies, 2000a, 2000b, 2001; Godfrey-Smith, 1993). Similarly, there is far from univocal support for the primacy of a biological notion of function and dysfunction for psychiatry. At least some theorists have argued that the two-stage view presupposes an evaluative or otherwise

normative aspect in the relevant notion of dysfunction (Megone, 1998, 2000).

While the *Diagnostic and Statistical Manual of Mental Disorders* presents a similar picture to Wakefield's in being two-stage and requiring dysfunction, the DSM presents a much more liberal view of dysfunction. The DSM regards the relevant dysfunction to be biological or psychological or behavioural. Wakefield argues that the relevant dysfunction cannot itself be behavioural, and instead the relevant dysfunction must be to a mechanism that is the cause of the behaviour[1]. He maintains that to consider behaviour to be functional or dysfunctional aside from how the behaviour is caused is to fail to capture a distinction between "problems in living" and cases of genuine disorder[2]. His line of reasoning is as follows:

Consider a person who is not able to read. Is the person suffering from a disorder? Reflecting on our concept we are led to the conclusion that whether or not the person is disordered depends on the cause of their inability to read. If we learn that the person has a dysfunctioning mental mechanism then we are correct to conclude that the person is suffering from a mental disorder. If we come to learn that their failure to read is due to nobody ever having tried to teach them then while we are correct in seeing that this person has very real problems in living, the person is not appropriately regarded as disordered.

Critics of the two-stage view maintain that all there is to mental disorder is "problems in living". This is just to say that we (normatively, evaluatively) judge the behaviour to be problematic. Wakefield maintains that there is a distinct class of disorders that are caused by dysfunctioning mental mechanisms. Critics maintain that the notion of dysfunction cannot be cashed

[1]The notion of a mechanism is problematic. There has been controversy over whether evolution by natural selection is a mechanism, for example. Some have argued 'no' because mechanisms need to be internal and localized - which seems to be how Wakefield is thinking of them. On the other hand, theorists have pointed out that paradigmatic mechanisms like neural states are spatio-temporally distributed and thus spatio-temporal distribution cannot count against evolution by natural selection. I want to put such issues aside for now and focus on the idea that Wakefield needs to do something to distinguish "disorders" from "problems in living". Indeed, this is the burden that two-stage theorists have chosen to bear.

[2]Wakefield is actually fairly liberal as to what sorts of things can be dysfunctioning mechanisms e.g., Component of them (genetic, cardiac, neurological etc) that is dysfunctioning. Wakefield's notion of a mechanism is fairly relaxed. He maintains that there can be relevant mechanisms on different levels of analysis. In the general medical case he considers mechanisms at the organ level and at the level of the cell. In the psychopathology case he considers mechanisms at the level of neurology and at the level of cognitive psychology. It is just that behavioural dysfunction isn't enough, it must be a component mechanism of a person.

out so as to distinguish a genuinely disordered group (since there aren't any) from those suffering from "problems in living" which while a very real cause of suffering aren't disorders.

Wakefield and Szasz seem to agree that persons or persons behaviours cannot malfunction. Or that alternatively, there is no objective, non-evaluative notion of function / dysfunction of a person or a persons behaviour.

Murphy seems correct to observe that Wakefield does seem to be capturing intuitions that we have about causes mattering. Some kinds of causes of behaviour (e.g., play acting, co-opting the sick role to avoid duties and / or to gain attention etc) seem to be exclusion criterion for our regarding a person to be mentally disordered even if they display relevant behavioural symptoms. I will have more to say about causes later when I come to offer a characterization of the nature of the contribution from theoretic science. For now, let us turn to Wakefield's methodology that he uses in arriving at his conclusion that the relevant kind of dysfunction is evolutionary. Once I've laid out the view we are in the position to consider alternative accounts of dysfunction and we can assess which (if any) of them seem to have prospects for distinguishing between "problems in living" that arise as the result of genuine disorder from those that don't.

Wakefield critiques the DSM view by being too liberal in maintaining that the relevant dysfunction can be biological, psychological, or behavioural. In particular, he maintains that behavioural dysfunction is insufficient for the dysfunction criterion. He provides the example of a person who is unable to read. They thus meet the DSM criteria for a reading disorder. If we find out that the person had received instruction that was comparable to other people and they had learned to read whereas this person had not then we would have the intuition that there was something wrong with this persons biological / cognitive mechanisms and they were dysfunctioning. If we learn that the person had not received adequate instruction, however, then even though this person might have the same behavioural presentation as the other person we would not have the intuition that the person was disordered, however. Wakefield uses this example to attempt to persuade us that behavioural symptoms are insufficient, and that the causes of the behavioural symptoms matter.

2.2 A shifting focus

2.2.1 Adaptive plasticity

An alternative to modularity is plasticity. The idea of distributed coding / processing. The thought is that many of our higher cognitive capacities might turn out to be realized on a central processor that is plastic and it might only be relatively low level cognitive functions that are likely to be hardwired.

Sterelny draws our attention to the massive variation in our environment that makes hardwiring fairly unlikely for humans. He points to how our cognitive development is importantly scaffolded by tools and by learning. Insofar as we have hardwired capacities they are more likely to be prosocial ones with respect to learning since we can't get by without that.

2.2.2 The accumulation of social capital

2.2.3 Minimal cognitive requirements

2.3 Health in physiological systems

Other things could be going on aside from dysfunction, though. Poorness of fit (operating outside designed range). Much of the literature on the evolution of co-operation focuses on explaining how co-operation could be selected for (or at least tolerated well enough to not have been actively selected against) given the problem of free riding and the personal cost of policing that. In *The Evolved Apprentice* (2012) Sterelny shifts the focus from Machiavellian policing to that of mutual positive benefits of pro-sociality and co- operation. For example the mutual benefit of pleasure in joint attention and joint action where that pleasure is hard to fake and thus provides evidence for ones commitment to collaboration.

The accumulation of social capital depends on sophisticated knowledge of how to extract resources from our environment in various co-operative and high tech ways. The vast range of habitats, technologies, and co-operative strategies employed by humans suggest that particular methods are unlikely to be innate. This knowledge is not something that every individual could learn in their lifetime by trial and error learning. It is in virtue of our gregarious hyper-sociality that we have the high fidelity oblique transmission of information and technical competencies to meet the performance specifications of our lifetime and contribute actively towards the accumulation of social capital.

Sterelny writes that we are obligate co-operators and that humans(?) have probably not been able to go it alone for the past 100,000 years or so[3]. Ones ability / willingness to participate in joint attention and the pleasure that one takes in joint action is hard to fake and might well be reliably predictive of whether one will engage in joint activity when it might seem to be in ones personal interests to skulk safely in the rear while the others are engaged in predator defence (for instance). It has often been thought that a certain level of emotionality is hardwired or innate and crucial for the development of pro-sociality and morality. Sterelny's focus seems to be on joint attention and joint activity accompanied with positive affect which is reinforcing rather than particular emotions (e.g., fear, shame), however.

2.4 From Simple Co-Ordination to Co-Operation

Sterelny points out that coordinated even if not co-operative defence is more likely to be effective than solitary defence. It seems fairly straightforward to think that if throwing a stone at a predator works somewhat well to scare it off then simple coordinated action of multiple people throwing stones would work somewhat better. We can imagine how throwing stones at predators is a fairly automatic self- defence fight response. When coordinated action is successful we can see how critters come to find it reinforcing / to take pleasure from it[4]. In this way we are more likely to engage in coordinated action next time.

Machiavellian concerns raise the issue that it would be safer to defect from stone throwing (where the predator might attack in the direction of a successful hit) and skulk round the back while reaping the benefits of others being engaged in coordinated defence. If one had the cognitive capacity to inhibit an automatic response of throwing stones then it would therefore seem wisest to do so. This does requires one to forgo the reward of joint activity, however, and would come at personal cost even if one was not risking punishment from other critters. Pleasure from joint activity could thus serve as a commitment mechanism. We could trust people who took pleasure from joint activity (and who expressed shame for defecting from it) because those

[3]I'm not sure why you say this because I thought there were (some though perhaps somewhat rare) people who do shun society and go bush. Perhaps this is more of a phase state, though. E.g., people might go wander in the desert for 7 years or go off to hunt possums or fish – but the point is that they come back. Is your thought perhaps also that they rely on the accumulation of cognitive capital in order to be self sufficient?)

[4]Pleasure is something that we find reinforcing but of course not all things that reinforce behaviour are pleasurable

emotions are hard to fake.

Sterelny considers that the vervet alarm call system is puzzling insofar as the benefits to the signaller seem unclear. While the form of the vocalization might be arbitrary perhaps it isn't arbitrary that a vocalization is made, however. Let us consider (a perhaps different) sort of alarm system that arises as part of automatic response of noise associated with preparing to fight or flight[5]. If different noises are involuntarily associated with fighting (a sort of automatic battle cry) and fleeing (a terrified scream, perhaps), then these noises would serve as indicators. But indicators of what? It indicates whether the critter making the noise has assessed the threat as being surmountable or insurmountable. It indicates what that critter is going to do in response to the threat. These two things are coupled for the critter. Insofar as one critter fairly automatically responds to another critters noises as it does to ones own it would fairly reliably predict what the critters will do. Perhaps vervet alarm calls are a bit like my swearing in English when I hit my hand with a hammer. The precise form of my vocalization might well be arbitrary (e.g., that I swear in English rather than German) but the fact that I swear might be automatic or inhibited only with difficulty).

If I'm a mature critter and the critter assessing the threat / making the fleeing noise is juvenile then the best course of action might be for me to inhibit fleeing and check out the level of threat for myself. I might easily overcome what is insurmountable to the juvenile. Thus there is a benefit from being able to inhibit an automatic response to a call. To take into account the characteristics / perspective of the signaller. To assess the trustworthiness of the signaller might be one way of putting it – though the issue here isn't one of intentional deception but perhaps rather or personal relevance.

From this scenario we might get the ability for critters to inhibit automatic response and respond to threat differently. If I am aware that the noise I make influences others actions then my assessment of threat can be different. It can be more prescriptive of what others should do / it can take into account the actions of others in assessing the degree of risk. For instance, if a threat is insurmountable by me but surmountable by us all and y'all are likely to come if I sound a battle cry then it could be advantageous to sound the battle cry. Maybe plausible insofar as territory defence is concerned. Or where fleeing is unlikely to grant much asylum e.g., because of pursuit.

[5]There is of course a freeze response, too, though perhaps unlikely to be accompanied by noise.

2.4.1 Cheating vs Defection / Incapacity

Some individuals are incapacitated (of pleasure / activity of joint attention and / or of pleasure / activity of joint action. Of course there are other reinforcement mechanisms than pleasure but people do find pleasure intrinsically rewarding. When it comes to individuals who aren't participating in join activity of work we distinguish between 'won't' – which is punishable and 'can't' which is not. The very young and the very old may be incapacitated. The temporarily sick or injured. More permanent disability. Physically and mentally handicapped. Problem of detecting the physically sick from those mimicking or pretending to be physically sick. The latter – mentally sick / criminal. We distinguish between mad and bad. The incapacity can be tolerated with population sizes that can support greater division of labour. Some socially useful functions.

The preference for solitariness might be beneficial when foraging dispersed resources with low threat of predation or hunting small game requiring focused quiet to track. It would make a more robust community if different individuals had different strategies (e.g., preferring to be solitary or gregarious) or if the same individuals were capable of employing different strategies given environmental shift (e.g., seasonally or yearly or given changing conditions). Or maybe a bit of both for extra resilience (at the population level). If some individuals persist with solitary strategies no matter what the environment is up to and the group buffers them by providing food then the group can see when times have changed and solitary hunting is starting to pay off. Is it tolerance for free-riding when someone's work efforts don't pay off? Unlucky. Societies with social capital may benefit from investing even when there is risk.

The gregarious / solitary distinction here might be thought to be done of degree, however. Solitary individual's take pleasure in joint action, too, it is just that they do it less often? They take less pleasure in it? Find it stimulating and there is a finer line between pleasing stimulation and overwhelming overstimulation. There could also important benefits (to society) of a fairly significant degree of solitariness and a- (even if not anti-) sociality. It would make sense to monitor these people more closely and trust them less.

While *'Quiet: the power of introverts in a world that can't stop talking'* (Cain, 2012) made the New York Times best seller list the notions of 'introversion' and 'extroversion' have been problematic for personality psychology. Indeed, the notion of 'personality' is problematic for personality psychology and a number of theorists have drawn our attention to how environmental features are more predictive of behaviour than internal or intrinsic features

of persons that have been suggested as candidates for personality types. Part of the problem of the notion of personality is the idea that personality is invariant across the lifespan and stable across environmental shift. One of the ideas of the distinction between introversion and extroversion is the idea that introverts get or recharge their energy from solitude whereas extroverts get or recharge their energy from social contact, however. Insofar as introversion / extroversion are credible constructs they might be thought to have at least something to do with pleasure from joint (vs solitary) activity.

If the notions of introversion / extroversion are problematic there is a (perhaps) more credible line of research on physiological arousal in response to stimulation. Some critters are more sensitive than others. It takes a stimulus with less magnitude to prompt a physiological response. It takes them longer to return to physiological baseline. People find people (eye contact in particular) to be physiologically stimulating / arousing. Infants vary in their sensitivity. There has been a lot of work on the role of the mother (in the first instance) regulating the physiological arousal levels of infants. The mother can help stimulate the infants physiological system – looking, smiling, giggling, laughing, tickling etc. The mother can help soothe or calm the infants physiological system – looking away, rocking, deep pressure touching etc. Responsive mothers are able to help the infants regulate their physiological systems within pleasant levels. Non-responsive mothers can be over or under-stimulating.

One line of research that might be more promising than introversion / extroversion is the idea of reactivity of nervous system with respect to physiological arousal. All kinds of critters seem to reliably vary on this. It is fairly invariant through lifespan (I think). Not every individual is highly gregarious, of course. There are many who are incapacitated from participating in social activities of contributing to the accumulation of social capital. We allow that there are some defectors, of course. Those who are unwilling that populate the prisons (and perhaps some of them are unable – anti-social / psychopath / sociopath). Those who are unwilling (but aren't quite so proactive in violating the preferences of others) – the lazy / complacent, the narcissistic etc. Those who are unable / incapacitated. The very old. The very young. Those who are sick. In hospital. Those with mental illness.

Check out the personality disorders stuff that has happened recently. Theorists are starting to think the narcissists and borderlines are morally objectionable loafers. Instead of the problem being conceived of as one of why we would punish cheaters – at cost to ourself, the problem might be more of how we refrain from punishing anybody who gets between us and the object of our interest. Sterelny (I think rightly) points out that it is not always obvious who the bullies and the punishers are. Perhaps not so clear

who the free riders / social loafers are either.

Tolerance. How to get the apes to stop killing the things. Small children are cute (less likely to kill them). But behavioural inhibition can come from fear, too. Adaptive action is harder. Behavioural inhibition is a natural response to bullies (if the bullies are in fact more physically powerful). The chimps show behavioural inhibition, surely? Is it more that they are unable to perform actions that require it (e.g., their inhibition is freeze rather than act). As population size increases this enables the division of labour. Fire keeping has been suggested as one of the first. The free-riders are the ones that have control of the institutions that are set up to punish free-riding. They get to label who the free-riders are. They get to say 'we aren't bullies we are punishing defection'. But when we wonder why they have an awful lot of stuff for that (how much money do they earn?) There can be a lot of papers etc... Busywork. Advisors etc. Gets lost in the number of people employed. Hide it in the organizational structure.

2.5 Responsibility / capacity / culpability

The malfunction assumption does for psychiatry what the dysfunction assumption does for biology... sometimes true sometimes false sometimes don't know. Doesn't impugn dx. What would the science of psychiatry (evo psychiatry) look like without it? In 'The Evolved Apprentice' Sterelny offers a scientifically informed account of the evolution of hominem cognition in a way that does not draw heavily from cognitive psychology or neuroscience. Sterelny focuses on the wide range of environments hominids have occupied and how our environment is radically shifting. If we are persuaded by this then instead of focusing on the unfolding of largely innate or genetically determined mental modules we are primed to think of the mind as being fairly plastic. Sterelny further primes us by introducing the idea of 'performance specification' that is abstract enough to divert our attention to the precise neural implementation. This is a useful manoeuvrer insofar as while the significant majority of us might meet the performance specification requirements there might be very little at the neural level (because of neural plasticity, for instance) that underlies these abilities. Either because of the present state of our knowledge, or more generally as a matter of principle.

What implications are there for evolutionary psychiatry and evolutionary medicine with this sort of approach to the role of evolution and science in explaining function and adaptedness? One of them is the idea that performance specification is central. I really like this idea of performance specification. It is behaviourist enough to get us out of the head – where things are ex-

tremely murky. It is cognitivist enough to be behaviour under a description and introduce an aspect of purposiveness or adaptedness. It seems a very useful notion indeed with respect to capturing some mid-way point between the organism on the one hand, and the environment on the other. The focus on our rapidly shifting environment gets us thinking about adaptedness as being something that is a distinctive feature of us. Plasticity. Learning seems to play a bigger role for us than for other critters. Our lifespans are long. It does take time to learn. To become en-cultured. To adapt. But adapt we do.

Wakefield argues that pre-theoretically our notion of function:

> is a shared concept based on prototypical examples of non-accidentally beneficial effects like sight, and on the idea that some common underlying process must be responsible for such remarkable phenomena (Wakefield, 2000 p. 39).

While Wakefield thinks that a-priori God could have turned out to be the relevant process for fixing functions (as Paley thinks is obvious) it turned out as a matter of empirical fact that the relevant process for fixing functions is evolution by natural selection.

One thing that is interesting to consider here is why he thinks that the god hypothesis is ruled out. Is it a matter of scale? Is it a matter of exclusiveness? In his commentary on the first premiss Wakefield maintains that according to our pre-theoretic notions of 'function' and 'dysfunction' we are uncommitted to what it is that fixes the relevant functions and dysfunctions. He states that it is perfectly consistent with our pre-theoretic notion that the functions and dysfunctions are fixed by the intentions of an intelligent designer or by a creator God. Wakefield then employs a sub-argument to lead to the conclusion that the *historical* process that was responsible for certain traits or features being present in present populations that is the process that fixes the relevant functions, however. It seems to be to be far from clear that the *historical process* was the important feature of Paley's argument from design, however. It seems rather that the *obvious purposiveness* was the important feature. Some evolutionary accounts of function (propensity views) focus on the forward-looking aspects of evolution by natural selection rather than the causal-historical aspects in order to respect the teleological intuition and thus it is not clear how the pre-theoretic notion of function is essentially historical. It is one thing to show the negative conclusion that the language of causes is incomplete. It is quite another to show the positive case that it is values, rather than some other Factor X, that is required to complete it (Fulford, 2000 p. 80-81) [he is contemplating Thornton's approach here].

2.6 Rethinking Evolutionary Accounts

Evolutionary accounts of mental disorder seem to have inherited some of the assumptions of Evolutionary Psychology. In particular, they assume modularity of mind. The focus is on dysfunctioning mental mechanisms. Wakefield is explicit in holding that environmental shift is not enough for mental disorder.

In 'The Evolved Apprentice' Sterelny considers an approach to the evolution of cognition that is more informed by ethology than cognitive psychology. He considers the task demands that our environments posed through evolutionary history and is relatively agnostic about the cognitive or neurophysiological mechanisms underpinning these capacities.

What if anything can we take from this with respect to evolutionary psychiatry?

Sterelny focuses on how hominem environments are plastic. Because our environments are plastic we are required to be adaptable. If we view things in this way then instead of seeing a mech operating in a new environment to be dysfunctioning we can consider that the person is dysfunctioning insofar as they fail to adapt to their environment. Mismatch hypothesis.

This seems in keeping with a number of intuitions that we have around the diathesis-stress model.

Chapter 3

The third 'small chunk'

3.1 Interactive kinds and the impact of classification

3.1.1 From essentialism to nominalism (with looping in between)

Essentialism

It just seems obvious that the natural world falls into a distinct number of different types or kinds - of substances such as water and gold, and species such as lions and onions. These regularities and the true inductive generalizations they support have been thought to be enabling conditions for our coming to acquire empirical knowledge. In accounting for *what it is in virtue of* that different instances or particulars are members of this or that kind the traditionalist view was that each thing was what it was and not some other thing in virtue of an *intrinsic, eternal, necessary* essence. While this account seems consistent with current chemistry, insofar as water = H_2O, and gold = atomic number 79 these kinds of essences don't seem to have been forthcoming for many of the natural sciences.

The psychiatrist Emil Kraeplin had a particular view of mental disorders that is fairly essentialist in focus. He thought that there were a number of distinctly different kinds of disorder (e.g., dementia praecox and manic depression). He thought that each kind of disorder had a distinct aetiology, underlying physiology, observable behavioural profile, and course. He thus thought that a classification made on the basis of observation of one (e.g., behavioural profile) would result in the same classification system as one made on the basis of observation of another (e.g., underlying physiology). On this

view we can also see how classification is thought to be useful for intervention purposes since a disorder shares a common aetiology and course we can see how it might be useful for guiding research into effective preventions and treatments for different kinds of condition.

Traditionalist essentialist views of kinds were focused on chemical and biological kinds. While traditional essentialist views seem to offer a somewhat plausible account of kinds for chemistry and perhaps physics they aren't so well suited to explaining biological kinds. There are a number of dimensions of the essentialist view and an account could be more or less essentialist according to various ways that one can relax one or more of the dimensions to a greater or lesser degree. For example, instead of maintaining that essences need to be intrinsic one could allow history to be the relevant kind of essence so that lineage (for example) provides what is essential to species. Indeed, depending on how one individuates particular members of a species it might be that lineage is an intrinsic rather than external relational property of species at any rate.

While this is one move that one can make a far more popular view for the biological sciences has been to move form essentialism to a homeostatic property cluster view.

3.2 Inductive Generalization

Richard Boyd considers categories of minerals, for instance, and the categories of metal, semi-metal and non-metal which all seem to lack traditionalist essences. He considers that a theory of intrinsic essences isn't able to account for meterologial kinds, and that perhaps some astronomical kinds will turn out the same way. 'The kind of stability which defined a natural kind of storm system, for example, may depend unexpectedly perhaps, on the nature of weather systems distant from the storm itself (Boyd, 1999, p. 84).' In this case the essential properties of the kind turn out to be *extrinsic* rather than *intrinsic* the way that traditional essentialism took them to be. According to current best biological theory species are members of a kind in virtue of their *extrinsic, historical* lineage[1]. These kinds seem to be a special case of extrinsic kinds where the essence in their particular case is *historic*.

While some have been inclined to bite the essentialist bullet concluding

[1]Actually, perhaps this is too simplistic. Zachar (2001). has some (possibly) interesting stuff to say about how biologists might focus on slightly different ways of carving up species for slightly different projects. Perhaps it is worth digging out an example or two of this and using it to motivate the homeostatic property cluster account a bit more. Or perhaps I should just start with the homeostatic property cluster account.

that there aren't meterological, mineral, and biological kinds insofar as they lack intrinsic essences other theorists have taken cases such as these to be evidence for our needing to revise our beliefs about the nature of natural kinds. Boyd states:

> It is worth asking whether it [*biological species*] is a natural kind in evolutionary biology. This is a somewhat odd question: It's like asking whether or not the kind *mineral* is a natural kind in mineralogy. In neither case would a negative answer provide a methodological reason to abandon the category (Boyd, 1999, p. 97).

Whether or not we are inclined to apply the term 'natural kind' to such phenomena we surely don't want to deny their inductive utility and hence usefulness for a variety of scientific projects. We seem to have little reason to attempt to *prescribe* the elimination of these categories from the respective sciences. Insofar as inductive usefulness was what was interesting about natural kinds then we need a more permissive account than traditional essentialism.

Zachar (2000, pp. 175-176) describes an example from Corning involving classification on the basis of similarity of presentation:

> Upon returning to his office after the meeting, he found that his son had reorganized his filing system. All his brown folders were neatly stacked... All his manila folders were stacked... The documents in each file had been removed and placed in their appropriate piles. Legal size yellow notepaper was in one pile, legal size white notepaper was in another pile, notebook-size yellow notepaper was in a third and so on. All white paper with typing on it was in its own pile... The important issue is not what rules we use to develop categories. The important issue is deciding how useful the proposed categories are. Organizing offices in terms of files, so that information on attention deficit disorder is in one file, information on lateralization is in another file, and information on cluster analysis is in another file is more useful to an academic than putting all the yellow paper in one pile. Categories are practical kinds. The categories developed by Corning's son were impractical, not arbitrary.

In a more recent work Boyd maintains:

> '...the epistemic reliability of scientific practices in a disciplinary matrix (when and to the extent that they are reliable) depends on many dimensions of accommodation between (on the one hand) conceptual features of practice in that matrix like its theories, concepts, classificatory practices, inferential standards, standards of experimental design, etc., and (on the other hand) the causal powers of the phenomena under study. The claim, about any scientific discipline, that its methods are epistemically reliable with respect to a given range of questions is always an empirical hypothesis not only about the subject matter of the discipline but about a variety of complex cognitive, social, linguistic, and classificatory practices (Boyd, 2010, p. 237).

This allows that whether a kind is natural or not must always be considered with reference to the scientific backdrop or matrix within which the category is utilized. Other theorists have similarly considered a conceptual backdrop to kinds e.g., (Godman, 2012, p. 175; Millikan, 2000)[2]. What might be a useful category for some scientific projects / some scientific methodologies might not be so useful for others.

While the American Psychiatric Association thinks that the needs of researchers, clinicians, and allied health professions is similar enough for a single classification scheme it is not obvious that this will turn out to be the case. Boyd considers that the categories of mammalian liver, vertebral liver, liver (broad enough to include Turing machines) and my own particular liver are useful or not depending on the inferences they license for different explanatory projects.

3.3 Homeostatic Property Clusters

If we focus on Boyd's homeostatic property cluster account then we have a more liberal view of what can count as an essence. Or instead of focusing on essences (insofar as we might be inclined to reserve the term for intrinsic, immutable, eternal, necessary properties) we have a more liberal view of what factors serve to stabilize the property cluster grouping. In virtue of what does knowledge of some properties support (more or less reliable) inferences about the presence of certain other properties? Naturalists typically maintain that what underwrites the success of our inductive practices (for natural kinds) is something about the causal structure of the world.

[2]At least I think something along these lines is meant to be going on with the 'determinables' and 'determinates' thing though I'm having a hard time grasping it.

The homeostatic property cluster view of natural kinds is much more liberal than traditional essentialism in allowing different particulars to be more or less the same and for inductive generalizations to be more or less powerful, more or less reliable[3]. It allows that there can be a multiplicity of relevant causal processes depending on what features we want explained. It allows that we can come to revise our categories so they support more powerful inductions (accommodation). In having done away with the idea of a single essence for each kind we can have the idea of multiple stabilizing causal influences tracking different aspects or features of the phenomenon for this or that explanatory project.

Psychiatry is an interesting case insofar as we might expect to see different kinds of kinds emerging – from the genetic to the social, the neurobiological to the cultural, the anatomical to the artefactual, to the non-existent and fictitious. Or where Boyd considers the minorly embarrassing (hence revised) we may also consider the majorly embarrassing (hence eliminated). Are mental disorders natural kinds? Or perhaps a better question: What kinds of mental disorder might there be? These questions have increasingly become the focus of discussion in the build up to the launch of the new edition of the *Diagnostic and Statistical Manual of Mental Disorders* in May of 2013.

While we might be well motivated to require our theory to be more permissive than traditional essentialism, we do need to draw the line somewhere, however. It is typically granted that any sane theory of needs to exclude arbitrary classes or mereological fusions such as Aristotle's super-lunary objects. A focus on inductive generalization may well have the resources to do just that insofar as knowing that x is a super-lunary object doesn't license us to inductive generalization about it. It is also typically granted that any sane theory needs to rule out conjunctive (e.g., duck-rabbit), disjunctive (either a fly or a bee-bee), and grue-some categories. While the standard line has been to appeal to 'natural categories' that have a 'causal structure' as the homeostatic mechanism I have some reservations.

3.4 Grounding Mechanisms

Szasz e.g., (1992) has perhaps been the most vocal advocate of the position that there 'isn't any such thing as mental illness'. He thinks that because

[3]These are special terms in science. Apparently kappa value is a measure of inter-rater reliability, for instance. I need to learn more about these measures so I can say something about them. I thought that inter-rater reliability for mental disorders was poor – but apparently it is comparable to the inter-rater reliability of many diagnoses in medicine (I think I need to check this out properly).

there aren't any mental disorders the whole institution of psychiatry is illegitimate. Or perhaps it is because he rejects the legitimacy of the background matrix of psychiatry that he rejects the categories that the field trafficks in. He thinks that psychiatrists aren't justified in categorizing people as mentally ill, and they aren't justified in involuntarily confining them or subjecting them to 'treatments' against their will. He does maintain that individuals should be able to voluntarily seek assistance for their suffering from psychiatrists should they choose – but that this is like engaging in consensual sexual relations (rather than being forcibly raped) and is more like choosing a religion or a particular restaurant – i.e., nothing to do with *health*. He thus rejects psychiatry as being legitimately regarded as a field within medicine and psychiatric disorders as being medical disorders (rather than non-medical problems in living or non-medical forms of suffering more broadly).

While Szasz reasons are complex and have been critiqued for being based on binary dualisms (Bracken & Thomas, 2010), one aspect is his belief that mental processes (as a causal mechanism) are simply the *wrong kind* of causal process for disorder. It perhaps does seem intuitive that causal processes *matter* in the sense that some causal mechanisms might be thought to be *legitimating* of mental disorders whereas others might be thought to be *debunking*. If we discovered some pattern on some or other chromosome that tracked attention deficit disorder with .87 reliability we would think we had discovered (a very significant!) biological grounding for attention deficit disorder which would legitimate our regarding attention deficit disorder to be a natural kind of disease category (if anything would).

On the other hand we can consider what sorts of things we might find out about it that would be debunking. Bentall (2004) maintains that schizophrenia and bi-polar don't form two different categories (clusters of properties) at all (by which he seems to mean that knowledge of a few morphological features isn't a reliable predictor of others either with respect to etiology, current presentation, or course). If this were correct then that would seem to undermine the utility of the categories.

3.5 Socially Constructed Kinds

Artefacts like pens and chairs are paradigmatic examples of Socially Constructed Kinds. Instances of the category pens count as members of the category in virtue of having the historical relational property of being designed by an agent for a certain function. As such agents designing them for a certain function is necessary and sufficient for or constitutive of category membership. Because they are designed by agents for a certain function

pens exhibit a cluster of superficial properties in common. Those properties may enable us to identify instances as instances of the category. If we found something that shared the superficial properties with pens but it grew on a tree or materialised out of a swamp then because it was not designed by an agent with the relevant intention it would not count as a pen, however. While pens are dependent on us for their initial existence once the instances have been brought into being then it is a mind independent fact that the instances are in fact members of the category. Even if we lost our concept of a pen or we no longer used pens to perform their function the instances that still exist would continue to exist as members of the category.

Some other socially constructed kinds aren't dependent on the intentions or mental states of agents so much as their social practices. Something might count as a doorstop, for example, not because it was designed with that intention in mind, but instead because it is currently being used to perform that function. If we accept this reading of what it is to count as a doorstop then it would follow that if we were to stop using the object as a doorstop that it would cease to be a member of that kind. There isn't a science of pens or doorknobs. While we might be able to make generalisations such as that pens usually have ink and that doorstops tend to be sturdy or obstructive it would seem that there are significantly less generalisations and predictions available to us than there is with either chemical or biological kinds.

3.6 Looping Kinds

The notion of a Looping Kind was initially introduced by Hacking and it has subsequently been picked up on by other authors such as Griffiths, Mallon, and Murphy. In order to describe the features of looping kinds I need to draw a further distinction between what I shall call explicit looping kinds and implicit looping kinds. Explicit looping kinds are kinds that are constituted by our social practices. While artefacts like pens are mind independent in the sense that they continue to be pens in the absence of our social practices around them, looping kinds are thought to be causally rather than definitionally or constitutively dependent on our social practices. Our social practices cause them to come into being as instances of the category and if our social practices change then this can cause them to go out of being as instances of the category. It is easiest to see this by way of example. Members of Parliament and Licensed Dog Owners are examples of explicit looping kinds. We have social practices around parliament and the election of members of parliament, for example, and in virtue of those social practices individuals come to be Members of Parliament. Unlike pens explicit looping kinds aren't

independent of our social practices because if we alter our social practices so that there isn't a parliament then the individuals would cease to be members of the category Members of Parliament.

Individuals that are Members of Parliament have properties in common such that they may be identified as Members of Parliament. We are able to make generalisations and predictions about Members of Parliament with respect to the properties they exhibit or are likely to exhibit and ways in which they are likely to behave. When the individuals are no longer members of the category Members of Parliament then they lose the properties that they had in virtue of their category membership, however, and we can no longer make such generalisations and predictions about them. These looping kinds are explicit in the sense that we are aware that the categories are dependent on our social practices. We know that there wouldn't be any Members of Parliament if we altered our social practices in certain ways. This doesn't stop us being able to make generalisations and predictions about Members of Parliament, however. It also doesn't stop the special science of politics from taking them seriously as a category.

Implicit looping kinds are similar to explicit looping kinds except that in this instance we aren't explicitly aware that the instances of the category are instances of the category because of our social practices and instead we regard the category as being a natural (or biological) kind. Hacking maintains that in this case if we were to become aware of their status as a looping kind then it would be inevitable that our social practices would change and this would have the result that the instances would no longer being members of the category. Our awareness and subsequent change in our social practices would also result in an alteration to the properties that the individuals shared as members of the category and thus the generalisations and predictions that were made about individuals in virtue of their category membership would no longer obtain.

Examples of implicit looping kinds include categories such as demonic possession and being possessed by a wild pig. The notion is that when we believed in these concepts then our belief in them and our social practices around them results in opening up new ways of behaving that are stereotypic of the category. If we take a person to be a member of the category or if they take themselves to be a member of the category then this may cause them to behave in ways that are stereotypic of the category. Members of the category are thus able to be identified as members of the category in virtue of sharing certain stereotypical properties in common. What is supposed to be distinctive about these categories, however, is that they cannot survive our realisation that they refer to looping kinds. The notion is that once we become aware that the properties are due to our social practices then we

cease believing in them and we inevitably alter our social practices so that the individuals no longer display those common features.

In 'Multiple Personality and the Sciences of Memory' Hacking (1995) tells a story of the evolution of a certain patient population (a controversial and perhaps question begging way of describing it – maybe there is instead a mere succession of different, unrelated epidemics):

> We tend to behave in ways that are expected of us, especially by authority figures – doctors, for example. Some physicians had multiples among their patients in the 1840's, but their picture of the disorder was very different from the one that is common in the 1990's. The doctor's vision was different because the patient's were different; but the patients were different because the doctors' expectations were different. That is an example of a very general phenomenon: the looping effect of human kinds. People classified in a certain way tend to conform to or grow into the ways that they are described; but they also evolve in their own ways, so that the classifications and descriptions have to be constantly revised. Multiple personality is an almost too perfect illustration of this feedback effect (Hacking, 1995, p. 21).

Hacking thus maintains that in the case of implicit looping kinds there is a tension in that possession of the concept and our social practises around this are the mechanism that both stabilises and destabilises the property cluster. With respect to the stabilising function he considers that individuals symptoms are shaped because when the clinician applies the concept to the patient this results in the clinician having either implicit or explicit expectations of the symptoms they expect to find in the patient. This changes the way that the clinician relates to the patient and is thought to lead to the patient exhibiting the symptoms they are expected to exhibit. Another way this can happen is if the clients apply the concept to themselves and thus come to exhibit symptoms that they believe to be stereotypic features of the category. In this way the concept and our social practices stabilise the symptoms that the patient exhibits as they come to behave in ways that are consistent with the stereotype.

Hacking also considers how our social practices can have a destabilising effect, however. He traces how the stereotypical features of Multiple Personality Disorder have evolved through time. Hacking tells a complex story of destabilisation and he draws on a variety of factors including political and theoretical, which lead to our beliefs about the concept evolving and the symptoms evolving in response to this. Some examples he has of this effect

in the case of MPD include how many alters are thought to be typical (one or several or over one hundred); whether there is one or two way amnesia; how long it takes to switch between alters; and reports of abuse. It thus seems that the change seems mostly to be a function of a change in the theoretical views of clinicians. This led to a subsequent change in how they related to their clients and what kinds of symptoms they expected to see. Hacking seems to regard implicit looping kinds as having some homeostasis but the homeostasis is less stable than other kinds of socially constructed and natural kinds in that awareness of their status as looping kinds will result in the dissolution of the category.

He states that sometimes people alter their behaviour to conform to classification (thus the category becomes homeostatic and supports more induction) but other times people alter their behaviour to not conform to classification (thus the category evolves or alters its morphology over time). He says that some categories can't survive our realization that they are social kinds rather than biological kinds. He seems to feel this about Multiple Personality in thinking that the more florid symptoms are the product of collusion between the patient and the doctor / are an artefact of the treatment situation. I have some stuff I need to pull out of my 6 month review on this. He thinks we should be nominalists about such categories since our awareness of being classified in this or that way alters what is being classified (behaviour).

We might think that the destabilizing factors are undermining of the status as kinds – but alternatively we might view them as being causal mechanisms that predict / explain the trajectory of morphological evolution of behavioural symptoms. If we know some of the causal mechanisms resulting in disorders evolving in this or that way we might be able to improve our identification of afflicted individuals and / or take steps to control or prevent certain future trajectories.

He doesn't have a great deal more to say about this, but I will go on to consider threads from Godman, Griffiths, and Tekin.

3.7 Hacking's Imitation and Internalization (I&I) Model of Apathetic Children

From Hacking (Hacking, 2010) A couple of children develop a rare phenomenon (PRS). Then it becomes known through the media and gossip etc and then other children come to imitate / role play. Then the imitation / role playing becomes internalized genuine. This accords with a range of phenomena such as catching faith by associating with believers (Pascal's wager).

A part or a version of his more general account of looping effects that affect kinds in the human science. Refugee children start out as 'social replicas' of the real kinds.

3.8 Trauma Model

High achieving (older children, responsibilities) + trauma. Implies a deeper connection in the sense that they are the same disorder with the same grounding (one is not a copy of the other). Same relevant aetiology factors, symptoms, recommendation for treatment. Doesn't account for epidemic (but not trying to explain population level frequencies).

There are similar trauma vs social role accounts of multiple personality disorder / dissociative identity disorder (I have an extended honours piece on precisely that that I can pull out – debate between a couple psychologists from the 90's). What is interesting is their different treatment prescription – empathy (trauma) on the one hand, punishment (social role) on the other. Swedish children refugee case all over again.

Griffith's (2008, p. 140) considers a similar sort of phenomenon of wild pig possession.

> The syndrome is treated as a disease by the tribe... The disease either runs its course or is ritually cured. Wild-pig behaviour is largely restricted to males between the ages of twenty-five and thirty-five. At this age men are likely to be under considerable economic pressure following the acquisition of a wife. Wild-pig behaviour seems to occur when a man cannot meet his financial obligations. After a display of wild-pig behaviour the individual receives special consideration with respect to these obligations. Newman convincingly explains wild-pig behaviour as a device by which a man can obtain this consideration without denying the fact that the demands made on him are legitimate. The behaviour is an action, but it is not acknowledged as such either by the individual or by society. It is part of the wild- pig role that wild-pig behaviour is involuntary.

Griffith's (2008, p. 143) also considers two variants of the social role model (of some culturally specific emotions). 'The social role model has two variants, the disclaimed action version, in which the behaviour is driven by a deliberate attempt to conform to a social role, and the reinforcement version. In the reinforcement version the social role is not internalized as a direct cause of behaviour, but behaviour is brought into conformity with it

by patterns of reinforcement in the cultural environment. This is very similar indeed to the two explanations of MPD as social role (intent hence punish) vs unintentional (trauma hence sympathy) explanations.

What difference does it make? One issue is on how we individuate the disorder. Do we have two disorders (Hacking things we do) or is it one disorder that has evolved through time? In a way it probably doesn't matter (like drawing the precise line on speciation events) but in another way it might matter. If culture bound syndromes are different manifestations of the same disorder (e.g., because they share similar aetiology and / or because similar treatments are effective for them and / or because they have similar stabilizing / destabilizing mechanisms) then this could be very useful for us to know. Hacking considers a number of questions that the Swedish authorities were asking about the apathetic children and that his theory provides the resources for him to explain:

- Why did the problem first occur only in the 2000s?
- Why this high number of cases in Sweden after 2002?
- Why are only asylum-seeking children in Sweden involved?
- Why are only children who have left certain geo-political-religious regions affected?

It seems to me that they didn't directly ask the question they were most wanting answered: Should we consider apathetic children as grounds to support an application of asylum seeking - or not? I take it that if the children had cancer or were tested HIV positive then this would be considered special (compassionate) grounds to expedite the asylum seeking process. If the parents were *intentionally inducing* this upon their children in order that their family be granted asylum then this would not, however, and if the children were *intentionally role-playing* in order that their family be granted asylum then this would not, however. If the case were the result of some kind of collusion between the child and / or the parents and / or the physicians who were sympathetic to the plight of the asylum seeking family then perhaps many would have the intuition that the phenomenon is somehow illegitimate. That they should perhaps be punished by denial of asylum, even. If, on the other hand, trauma plays a significant role then we might be extremely reluctant to allow them to be further traumatized (e.g., by being the 'victim' of an extended asylum seeking process).

We can consider a parallel to the evolution of hysteria that Hacking considers. Hysterical paralysis as doctors found that interesting, then when this

seemed less interesting (since collusion was suggested as an aetiology and clinician's didn't like people thinking they might be making their patients worse) then new morphologies seemed more interesting. Insofar as we lack empathy for / blame people for their symptoms then I think it likely that this results in symptoms morphing into something that we have a little more empathy for. Or at least that sounds plausible to me (as plausible as the idea that behaviours that are rewarded increase in frequency and those that are punished decrease). Intent aside.

What does one need to do in order to get a little asylum in Sweden? And how (genuinely) distressing is that? The publicity afforded to the family had many feeling empathetic toward the plight of the family which could in turn have an effect on the policies around application for asylum in Sweden (e.g., whether there was public pressure to expedite the process and so on). Which resulted in alleviation of distress as asylum was granted.

The issue of *intent* intuitively seems to be crucially important to our intuitions about quite a lot of mental disorders. Szasz critiques psychiatry for doing away with the distinction between thinking you are sick and actually being sick since anybody who thinks they are mentally disordered gets to be mentally disordered. Not everyone who thinks they are mentally disorders gets the kind of treatment they would like to receive, however. If we consider hysterical paralysis there was a proliferation of cases when clinicians were interested in treating it. When it was suggested that the phenomenon might be *therapist induced* or the product of *collusion* between doctor and patient or that the patient might be *faking* then rather less of it was observed / reported. Indeed, if one is in fact presenting with symptoms *in order to* take up the sick role then the symptoms will only be expected to appear *insofar as they enable one* to take up the sick role. If clinicians aren't sympathetic to hysterical paralysis as a *genuine disorder* then there is no utility to present with the symptoms of hysterical paralysis. This can be given a fairly straightforward behaviourist analysis with respect to what behaviours are reinforced intent aside. Though we might think that intent is a causal factor that has the potential to debunk the legitimacy of the phenomenon.

What is the difference between social loafing, being a lazy shit, and being depressed? Why would one work when one could lounge about freeriding off the products of others? People often write about such a thing in such a way as to suggest that everyone would do it if they thought they could get away with it. That it would be in our own best interests to socially loaf if possible. All the benefits of the labour without the costs – it is better for us fairly much by definition – right? It is unclear that it is on a fuller understanding, however. Most people get a sense of personal satisfaction / pride when they feel like they have contributed something positive to society. A sense of

accomplishment from a job well done. From the assembly worker to the apple picker to the computer programmer. Not all, to be sure, but generally people like to feel useful and they often tell themselves stories in which they are the main player. How important their contribution is. Even assembly line workers tend not to view their contribution as replaceable parts.

Some evidence that people who are depressed have a more realistic self assessment than those who aren't (who tend to over-value themselves). Maybe there isn't any difference between being depressed and being a lazy shit. But maybe instead of thinking that is grounds for treating people with so-called 'depression' like they are lazy shits (having no empathy for them and thinking a kick in the pants will cure them) maybe instead it is grounds for treating people who are so-called 'lazy shits' like they are depressed (and instead figuring how to behaviourally get them into activity schedules, cognitively get them (over)-valuing their contribution, and socially involved with their communities).

William Baum... 'Behaviourism explained' begins with a chapter 'beyond freedom and dignity' on how the behaviourist view can allow for our social practices of punishment (prisons) being legitimate even though the notions of personal responsibility and blame need to go. Maybe I want to apply something like this to psychiatry – do away with the notion of there being a difference whether someone is intentionally or unintentionally behaving a certain way. Maybe intent doesn't matter (aren't we supposed to be being eliminativists about such functional folk psychological notions in our scientific explanations of behaviour anyway)?

Perhaps it is about flourishing or the pursuit of an ideal. And the limiting factors are simply that our resources are limited so we need to prioritize when it comes to charity.

Is mental illness more prevalent than it once was or have we changed our criteria? Do more people meet the criteria instead of other things. By which I mean to ask: Do people channel their distress into ways that are socio-culturally appropriate and in western societies where a lot is invested in destigmatizing mental health issues and where (comparatively to developing nations) a lot of resources are allocated to treatment of mental health problems individuals are taught to channel their distress in that way? Particularly in the depression / anxiety way? With the decentralization of church (as a place to go to seek comfort / solace) do we perhaps have the rise of alternative institutions?

In a forthcoming chapter[4] 'The missing self in Hacking's looping effects'

[4]Thanks to Kim Sterelny for providing me with forthcoming publications from Hacking, Tekin, and others.

Serife Tekin presents the 'multitudinous self' and makes the case that 'we can use it to make sense of how self-concepts change after the subject receives a diagnosis of mental disorder. Self-concepts and behaviour change due to (i) the subject's knowledge of the illness... as well as (ii) the course of illness; and (iii) the psychiatric treatment the subject receives (Tekin, forthcoming, p. 31). This is an interesting proposal to separate out some distinct factors so as to make more specific hypotheses regarding the stabilizing or destabilizing trajectory of disorders (assuming again that we are tracking the same thing as it evolves over time).

An idea of making oneself sick – so then the individual is to blame for their predicament (they are undeserving of help). Of course in the physical illness case we think that if one intentionally or unintentionally gets sick there are objective facts about whether one is sick or not. Perhaps it does affect our empathy. Most don't feel entirely sympathetic of the obesity epidemic in the US or of everybody paying more for accommodations (e.g., larger seat planes) for those with such 'disability'.

Hacking makes much of the idea that bacteria aren't self-aware. That it makes no difference to the bacteria whether or not people have whatever attitudes towards them. The bacteria are only responsive to what people do in response to them. On the one hand I do understand the temptation to consider self-concept with respect to mental disorder. When people identify with their diagnosis / identify with others with the diagnosis. When they start to conceptualize themselves as a person with that disorder. When they start to explain their behaviour by recourse to it 'I'm jittery because I have an anxiety disorder' or make predictions about their behaviour 'I can't go outside because I'm agoraphobic'. Or when they otherwise adapt their behaviour in response to their beliefs about people with their disorder. This is a stabilizing mechanism. The more they learn about the diagnosis the more they learn about 'typical'. The more they learn about the efficacy of this or that treatment the more they are likely to stabilize towards the typical outcome.

There are socio-culturally accepted and not accepted ways of asking for assistance. What ways are accepted and what ways are not accepted alters through time. Why is this the case? We want to make it harder to ask for help than it would be for one to simply help oneself (if one was in fact capable of helping oneself). When it comes to formal (institutional, governmental) forms of assistance we want to make it such that non- formal (e.g., familial, community) forms of assistance are exhausted first. Partly because the closer the ties are the more likely the person is to see that others pay the cost that the person could not (which deters against loafing / expedites recovery). Wild pig people might be multiples in future generations (though these disorders

are demographically distinct).

Very real need... For what? Possibly for some empathetic attentiveness, a human connection. While slightly different. We might see the same thing in physiotherapy. While respiratory physiotherapy traditionally involved a lot of tilting people to varying degrees of upsidedownness in order to belt them in different regions of their lungs to help them cough up stuff there has increasingly become a market for private physiotherapy clinics in affluent areas. One such clinic is 'breathing works'. The idea is that most people's suffering (the nebulous fidgety anxious not entirely happy feeling that plagues many affluent people who think they should be happy with their successful lives) is the product of non-optimal breathing. For a fee a physiotherapist can teach you how to breathe. I know you didn't know it was a problem. You have now been 'educated' however. Or maybe it isn't breathing training (or isn't 'just' breathing training) you need. Maybe it is a little clinical Pilates as well... Maybe that was the cause of your distress all along. People report it is quite effective for treating asthma, anxiety, allergies and so on. This bleeding edge treatment is thought to be pushing the boundaries of physiotherapy practice. It is available to practitioners particularly in New Zealand because patients can self-refer for physiotherapy - they need not be referred by a physician. Physiotherapists can market their services directly to consumers (e.g., market how faulty breathing is the cause of your nebulous complaints) and negotiate whatever payment clients will pay.

Of course a downside to all of this is that physiotherapy is / will be increasingly alienated from 'traditional' healthcare (publicly funded / health insurance reimbursed). Some physiotherapists are happy to view themselves as providing a personal service. Other physiotherapists prefer to see the profession as allied health. One might ask.... Whether there is much of a difference... The rise of allied health as... User pays pursuit of flourishing. And why not? Still left with charity... And where we draw the line on that as a society. How much charity costs and how much the society can afford. But maybe it simply is an issue of that and there isn't anything particularly 'health' about it or the conception of health is broad enough to capture the aesthetic.

Are new physiotherapists being trained in 'breathing works' techniques? Should they be (if there is a market for it)? What difference does it make? We can ask similar questions of plastic surgeons who work in private practice performing cosmetic surgery. Beauty therapists. Massage therapists. Hand and nail specialists. Podiatrists. Personal trainers. Hairdressers. Health? Talk to a religious leader. Talk to citizens advice. Etc. In psychiatry we have origins in Freud and psychodynamic. It hasn't broken from that yet. Certain diagnoses are notorious. The idea that people are 'attention seeking' (that

perhaps we should refrain from paying attention to them as punishment) and so on. The other idea (more common in aspects of 'health' where there is a private market) that people pay for a service and let the market dictate. If a person chooses to pay $200 on a new haircut and color or $200 on fixing their bad breathing or $20,000 of a course of psychoanalysis for an aesthetically pleasing narrative structure for their life or $5,000 for a new painting for the hallway or 10,000 for a nose job or $7,000 for a personal trainer building up for their high school reunion then where's the harm?

Suicide is self inflicted. We tend to encourage people to seek assistance if they feel like killing themselves unless they feel like killing themselves too often in which case we regard them to be 'overly needy' and 'attention seeking' and think they shouldn't be helped / think we are justified in our failure to help them. Some people end up with very real (physical) injuries from failed attempts. We think that these people really are physically disordered but might have different intuitions about their treatment? Perhaps.

3.9 Implications of implicit looping for taxonomy

In these cases because it is implicit that we are dealing with a looping kind we are unaware of the impact of categorisation, our social practices, our expectations, our ways of interacting with the person, and so forth. If we come to believe that a certain kind of mental disorder is a looping kind then it seems that one of three things could happen:

Firstly, it could turn out to be the case as an empirical matter of fact our change in belief does not result in a change in our social practices. While Hacking thinks the relevant social practices are ones that invariably would change if we became aware that the category was a looping kind surely it could be possible that the social practices that are sustaining the phenomena could be resistant to change possibly because they have other beneficial effects. It is unclear whether Hacking would consider this to be an example of an implicit looping kind because it was implicit even though awareness did not result in its dissolution or whether Hacking would consider this to be an example of an explicit looping kind because it does not dissolve in the face of our awareness even though the so called explicit looping kind was implicit for a time.

Secondly, it could turn out to be the case that as an empirical matter of fact that if we came to believe the category was looping and we changed the relevant social practices the stereotypical behavioural features remain.

In this case we seem to be left having to conclude that the category wasn't a looping kind after all. While it could still be socially constructed in the sense that artefacts similarly rely on us for their initial existence the phenomenon wouldn't seem to be dependent on our social practices and thus it would not be an implicit looping kind on Hacking's account. The third thing that could happen would be that our awareness of the category as an implicit looping kind could cause the stereotypic features to shift. If we found that a particular kind of mental disorder was an implicit looping kind this isn't to say that all instances of the category are suddenly cured of all symptoms of psychopathology, however. It is just to say that they won't display features of psychopathology that were stereotypic of the looping kind. They may well go on to display stereotypic features of another psychiatric kind, for example. Social constructionists about Multiple Personality Disorder often say that there is no such category as Multiple Personality Disorder there is only Borderline Personality Disorder that has been worked up into Multiple Personality Disorder in response to our social practices around the concept. The notion here seems to be that if we refuse to participate in those social practices the patients will display stereotypic features of Borderline Personality Disorder instead.

What is unclear, however, is whether this would be so because the clinician's expect them to come to display the stereotypical features of Borderline Personality Disorder or whether this is in response to some other mechanism. If clinicians came to believe that there was no such category as Borderline Personality Disorder then would the individuals continue to behave in a way consistent with a diagnosis of Borderline Personality Disorder or would their behavioural symptoms shift so that they met criteria for another diagnostic category? While Multiple Personality Disorder is often one of the favourite categories of those who maintain that we need to look at social causal mechanisms it is unclear whether other, more paradigmatically biological psychiatric kinds could turn out to be looping kinds or to have a looping kind feature to their behavioural symptoms. It could turn out to be the case that mental disorder more generally has a significant looping kind component. If this was found to be the case then this would seem to have significant implications for both the project of how we identify mental disorders and the project of how we develop a scientific classification of them.

One implication is that focusing solely on behavioural symptoms might be counter- productive. Each subsequent edition of the DSM is praised for making scientific progress with respect to providing categories that better support generalisations and predictions. If the properties relevant for generalisation and prediction are purely behavioural symptoms and if the behavioural symptoms evolve over time in response to the classification system

and a new round of expectations by clinician's then it would seem that the DSM approach will be limited insofar as the property cluster is unstable. The DSM may not only describe current symptomatology but it also may have a causal role to play with respect to future symptom development. One consequence of this might be that the DSM and ICD aren't necessarily converging on constructs that are more valid than the old constructs; rather each edition might recover some of the construct validity that the old one had by adequately capturing present symptoms that may, at least partly, have been evoked in response to previous systems of classification. Construct validity on the basis of generalisations and predictions on the basis of behavioural symptoms may be of limited value with respect to a scientific nosology.

If we identify kinds of mental disorders according to causal mechanisms rather than behavioural symptomatology, however, then this enables us to say that the behavioural symptomatology of a particular kind of disorder can evolve over time. This latter approach also allows that there could be considerable cross-cultural variation in the behavioural symptoms of individuals who have the same kind of mental disorder. While the DSM saw purely behavioural symptoms as progress from the causal mechanisms offered by the psychodynamic theorists cognitive neuropsychology would seem to have good prospects for grounding the next stage of scientific development from observational properties towards a scientific nosology of the causal mechanisms that produce psychiatric disorders. It seems plausible to me that more valid constructs may require us to incorporate causes from multiple levels of analysis. While there will be more to social causes than the looping effects that Hacking deals with the looping kind effect is interesting with respect to the relationship between social cognitive and behavioural facts. If we consider that the cognitive facts are represented within the brains of individuals it seems that whether the cause is inner or outer may be a function of how far back in the causal chain we look.

3.10 Implications for problem cases (e.g., addiction, sociopathy)

The problem here is that whether these conditions are labelled 'mental illnesses' or not has important implications for whether these people are treated or jailed, whether health insurance companies are required to provide treatment or not, whether we are able to discriminate against these people or whether they are covered by mental health laws. It would seem to me that the relationship between mental disorder and right to treatment, moral re-

sponsibility, and legal responsibility is a separate issue really... It is far from clear that these things are part of the concept or if they are connected so as to feature into the Carnap conditional then this is importantly different (there aren't facts aside from our social practices). What is left to argue about how our social practices should be. For example, it could be possible to proclaim that addiction is a mental disorder and yet addicts should be prosecuted. The interest in these being mental disorders seems to be around social and legal responsibility. We already know these come apart.

The answer to these questions will come from a complex interrelationship of honing our intuitions and empirical investigation. It is nice that people are doing the conceptual analysis thing and it is important to not end up with a brain storm of features where some are redundant or fairly irrelevant but by the same token it is important not to make the issue out to be too black and white and it is also important not to isolate part of the project off from the whole. Implications for sociopathy and addiction.

How many features do these conditions share with paradigmatic mental disorders and paradigmatic non-mental disorders? How much do mental disorders really have in common? Problem with the data in that the models seem to assume rather than discover irrationality etc. concern about stipulated malfunctions. Decisions we made around the criteria have consequences for the under / over inclusiveness of categories. Once we realize that is is problematic whether there is a categorical feature to nature such that we get things right or wrong. Once we appreciate some of the subtlety of the situation then we can be more nuanced. Multiple personality (and the sciences of memory). We can cast the net broadly or narrowly. This has consequences for seriousness. institutionalization. medication. and so on. the answers to these questions is dependent on how to choose to id the individuals to start with. the literature on sociopathy. different ways of defining cast it narrow or broad. cast it broad and study undergraduates. but then problematic relationship to the most serious (which is very rare).

The Problem of Disjunctive Properties

The disjunction problem arose in the context of fixing content for teleosemantics. The issue is how we are justified in saying that the frog is misrepresenting FLIES when it snaps at bee-bee pellets rather than saying that the frog is accurately representing FLIES OR BEE-BEES. The obvious answer is that the past tokenings of flies were involved in causing the frog to come to have the concept FLIES whereas there weren't any bee-bees in the frogs environment. A worry that I have about this retort is that flies or bee-bees were in fact in the frogs environment. In response, flies or bee-bees isn't a natural category, whereas flies is, and the natural categories are responsible for causing the content. If we then want an account of what natural categories there are we say that they cannot be disjunctive. What grounds is there for considering the content to be disjunctive? It is meant to be that there isn't a common causal explanation for flies or bee-bees the way that there isn't a common causal explanation for jadeite or nephrite. Maybe.

The Problem of Gruesome Properties

Richard Boyd (2010, pg. 73)'Anyone who can define "grue" can define persistence conditions for grue-ish individuals which mix and match temporal stages of ordinary persisting individuals, or define otherwise unnatural things as well as unnatural properties. Induction and explanation require that we quantify over natural as opposed to unnatural individuals every bit as much as that we deploy only projectable predicates.'

I thought that the problem of gruesome properties ran deeper than this, however. The reason why it was such a problem is that we don't have grounds for concluding that grue is more gruesome than green.

Something is grue if it is green before time t and blue after it.

Something is bleen if it is blue before time t and green after it.

It appears that grue and bleen are time dependent or temporally indexed in a way that precludes their being projectable predicates. They are time dependent in the way that blue and green are not.

But consider the predicates green and blue from the perspective of the grue speaker:

Something is green if it is grue before time t and bleen after it.

Something is blue if it is bleen before time t and grue after it.

From the perspective of the grue speaker it is the properties green and blue that are disturbingly temporally indexed and hence not projectable.

If our account of natural properties appeals to some properties being natural because they are projectable then we don't have grounds for concluding that grue is less projectable than green.

Chapter 4

References

4.1 References

American Psychiatric Association. (2000). The diagnostic and statistical manual of mental disorders

Bentall, R. P. (2004). Madness explained: Psychosis and human nature. London: Penguin.

Boyd, R. (1999). Kinds, complexity and multiple realization: Comments on Millikan's "historical kinds and the special sciences". Philosophical Studies, 95(1/2), 67 - 98.

Boyd, R. (2010). Realism, natural kinds, and philosophical methods. In H. Beebee & N. Sabbarton-Leary (Eds.), The Semantics and Metaphysics of Natural Kinds: Routledge.

Bracken, P., & Thomas, P. (2010). Is private (contract-based) practice an answer to the problems of psychiatry? Philosophy, Psychiatry, & Psychology, 17(3), 241-245.

Cain, S. (2012). Quiet: the power of introverts in a world that can't stop talking: Crown Publishers.

Cooper, R. (2005). Classifying madness: A philosophical examination of the diagnostic and statistical manual of mental disorders. Netherlands

Cooper, R. (2007). Psychiatry and the philosophy of science

Fulford, K. W. M. (2000). Teleology without tears: Naturalism, neo-naturalism, and evaluationism in the analysis of function statements in biology (and a bet on the twenty-first century). Philosophy, Psychiatry, & Psychology, 7(1), 77-94.

Godman, M. (2012). Psychiatric disorders qua natural kinds: The case of the "apathetic children". Biological Theory, 7(2), 144-152.

Griffiths, P. E. (2008). What emotions really are: The problem of psychological categories. Chicago, IL: University of Chicago Press.

Hacking, I. (1995). Rewriting the soul: Multiple personality and the science of memory. New Jersey: Princeton University Press.

Hacking, I. (2010). Pathological withdrawal of refugee children seeking asylum in Sweden. Studies in History and Philosophy of Biological and Biomedical Sciences, 41(4), 309-317.

Millikan, R. (2000). On clear and confused ideas: An essay about substance concepts

Megone, C. (1998). Aristotle's function argument and the concept of mental illness. Philosophy, Psychiatry, & Psychology, 5(3), 187-201.

Megone, C. (2000). Mental illness, human function, and values. Philosophy, Psychiatry, & Psychology, 7(1), 45-65.

Sterelny, K. (2012). The evolved apprentice: How evolution made humans unique: MIT Press.

Szasz, T. (1992). The myth of mental illness. In R. B. Miller (Ed.), The restoration of dialogue: Readings in the philosophy of clinical psychology (pp. 175-182). Washington DC.

Tekin, S. (forthcoming). The missing self in Hacking's looping effects. In H. Kincaid & J. A. Sullivan (Eds.), Mental kinds and natural kinds: MIT Press.

Wakefield, J. (1992a). The concept of mental disorder: On the boundary between biological facts and social values. American Psychologist, 47(3), 373-388.

Wakefield, J. (1992b). Disorder as harmful dysfunction: A conceptual critique of DSM-III-R's definition of mental disorder. Psychological Review, 99(2), 232- 247.

Wakefield, J. (1993). Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder. Journal of Abnormal Psychology, 102(1), 160-172.

Wakefield, J. (1999). Evolutionary versus prototype analyses of the concept of disorder. Journal of Abnormal Psychology, 3, 374-399.

Wakefield, J. (2000). Aristotle as sociobiologist: The "function of a human being" argument, black box essentialism, and the concept of mental disorder. Philosophy, Psychiatry, & Psychology, 7(1), 17-44.

Wakefield, J. (2003). Dysfunction as a factual component of disorder. Behaviour Research and Therapy, 41(8), 969-990.

Wakefield, J. (2004). The myth of open concepts: Meehl's analysis of construct meaning versus black box essentialism. Applied and Preventive Psychology, 11(1), 77-82.

Zachar, P. (2000). Psychiatric disorders are not natural kinds. Philosophy, Psychiatry, & Psychology, 7(3), 167-182.