Making Wakefield Warranted: The Hierarchy of Healing Harm and Discerning Dysfunction

Adam D. Hunt
Institute of Evolutionary Medicine, University of Zurich

Abstract

The definition of medical 'disorder' is a matter of longstanding debate. A leading account is Jerome Wakefield's 'harmful dysfunction' analysis, a hybrid model which merges normative and naturalist elements to propose disorder exists in cases of concurrent harm (value-defined) and dysfunction (defined evolutionarily). Despite significant impact in academia, this has so far failed to affect mainstream medical treatment or discourse, with a major criticism being that this definition doesn't correctly capture all conditions of medical relevance or actual medical ideals. This paper provides a supplementary structural nuance to better reflect how medical treatment and terminology incorporate naturalistic facts. I argue that Wakefield is right in utilising a hybrid model incorporating naturalism and normativity. However, in understanding how medicine is directed, making the normative and naturalistic equally necessary is problematic, because the imperatives of naturalism and normativity directly impede each other; norms seek to help those who need help, naturalism concentrates on objective facts, but neither can be fulfilled as a coherent goal of medical treatment when combined as equally necessary. Instead, I propose a hierarchical harmful dysfunction model better reflects current medical ideals, where normative values are necessary and sufficient conditions to prescribe treatment, whilst naturalistic knowledge plays a role informing those normative values. This accounts for the edge cases and practise of medicine not fully captured in Wakefield's account of 'disorder'.

1/ Normativity, Naturalism and Wakefield's Harmful Dysfunction

A core debate in the philosophy of medicine concerns defining disease and disorder, and the conflict between normativity and naturalism. Normative approaches regard medical classification as value judgements: broadly, healthy states are desirable, disease states are undesirable (Engelhardt, 1986; Margolis, 1976). Classifications reflect social value placed on particular biological states, not inherent or objective biological facts. Glackin (2010) even claims there are not only no *sufficient* biological criteria to classify a condition as a disease, but neither are biological criteria *necessary*. This is comparable to law – actions are crimes as a matter of value judgement alone (Matthewson & Griffiths, 2017). Normative considerations are, and always have been, the impetus behind medical treatment – as evidenced by the contradiction of a goalless medicine. The fact that medicine is many millennia older than science, and practised by non-scientific communities and individuals, serves as further evidence concepts of health and disease are not inherently scientific (Murphy & Woolfolk, 2000). Perhaps the most active contemporary public manifestation of an explicitly normative approach to medical issues is the 'social model' of disability arising from within the disability advocacy movement (Barnes, 2016). The central claim is that individuals are not inherently disabled but society is disabl*ing*,

for instance by being built inaccessible to certain individuals. Biological differences become disabilities because of values (e.g. building stairs by valuing legs that can climb stairs over legs that can't climb stairs).

The contrasting position to normativism is of 'naturalism'. Most famously associated with Christopher Boorse (1975, 1977, 1997, 2014), naturalism's central claim is that health and disease must be defined in descriptive biological terms:

> "the classification of human states as health or diseased is an objective matter, to be read off the biological facts of nature without need of value judgements. Let us refer to this general position as 'naturalism' – the opposite of normativism, the view that health judgements are or include value judgements" (Boorse, 1997, p4)

This endeavour was at least partly reactive, out of wariness to the social consequences of normativism:

> "the value-free scientific disease concept [is] a bedrock requirement to block the subversion of medicine by political rhetoric or normative eccentricity" (Boorse, 1997, p100)

Boorse proposed 'biostatistical theory' (BST), inspired by physiology, "the paradigm health discipline" (Boorse, 1975, 49). Boorse's BST defines disease as the statistically abnormal functioning of a specific trait in comparison to similar individuals in a reference class. Health to Boorse was the absence of such abnormality. Disease states reduce functional abilities below typical efficiency. Boorse's influence on the philosophy of medicine is undeniable – but not because of the BST's success. In Boorse's own words "the BST's influence is hardly due to its multitude of converts" (Boorse, 1997, 4). One of the most obvious and well-worn criticisms is that by referencing statistical normality, any disease which spreads beyond a certain proportion of the population loses its disease-status. The fact the specific proportion who classify as diseased is arbitrarily line-drawn is another inherent problem (Ereshefsky, 2009) (although evolutionary theory may allow non-arbitrary line-drawing – see Hunt and Jaeggi (2022)). Nevertheless, naturalism has arguably become a mainstream position in philosophy of medicine specifically because of Boorse's influence (Saborido & Moreno, 2015).

More successful 'naturalising' of disease comes from the class of models variously called 'etiological', 'backwards looking' or 'selected effects' (Millikan, 1989; Neander, 1991; Wright, 1976) (henceforth captured under 'etiological'). Closely related to concepts of function and normativity in the philosophy of biology and the life sciences (Varner, 1998), etiological models of disease explicitly reference the theory of evolution by natural selection. Evolutionary theory is uniquely placed to naturalise definitions of health by reference to evolutionary function, because although one might *say* that atoms, molecules and all physical processes are 'functioning', this use of 'function' does not carry the objectively-derived goal-directedness with which 'function' applies in the life sciences, where physical processes exist *because* and *for* the adaptive function of increasing reproductive success. When we say 'eyes are for seeing', we are describing a function of eyes which explains their current form, which also allows us to categorise non-seeing or badly-seeing eyes as dysfunctional.

Various accounts refer to evolutionary history as the crucial process by which normative claims in biology are objective rather than human-imposed. Godfrey Smith (Godfrey-Smith, 1994) and Wright (1976) are influential accounts in philosophy, whilst Wakefield's (Wakefield, 1992, 1997, 2005) 'harmful dysfunction' model of disorder is the most influential in evolutionary medicine and psychiatry (Stearns *et al.* 2010; Nesse 2019; Hunt, Abed and St John-Smith 2022), and the account concentrated on here. Where Boorse sought to disconnect definitions of health and disease from value judgements entirely, Wakefield and others (e.g. Caplan, 1992; Reznek, 1987) take a 'hybrid' view, suggesting that

satisfactory analysis requires both a naturalistic and evaluative component. For Wakefield, this is best achieved by defining disorder as 'harmful dysfunction' whereby a disorder must *both* be dysfunctional in the naturalistic, etiological, evolutionary sense of failing to achieve selected effects, and also harmful in an evaluative, normative sense. Wakefield strongly frames this as a product of conceptual analysis: the 'harmful dysfunction analysis' (HDA) is supposed to best capture the actual sentiment behind everyday and technical use of the term 'disorder'.

Wakefield's account is essentially unparalleled in recognition as a candidate definition for disorder, both outside of philosophy, where it is highly regarded within the field of evolutionary psychiatry, and inside philosophy, where it is suspect to more controversy and criticism. A recently published book "Defining Mental Disorder: Jerome Wakefield and His Critics" (Kincaid et al., 2021) contained 13 chapters written by critics, and 15 lengthy responses by Wakefield, providing an unmatched overview of the HDA as it stands. Wakefield is generally suspect to two overarching criticisms: firstly, regarding the difficulty of utilising the 'dysfunction' component (which I deal with elsewhere (Hunt, 2023)), secondly regarding its relegation of value, the resulting definition straying too far to be warranted as applying to the full range of medically-relevant conditions. This paper concentrates on this second criticism. First, a realistic thought-experiment will be introduced, to help illuminate these problems, before proposing a possible solution for saving hybrid models such as Wakefield's. My aim is not conceptual analysis of the term 'disorder', but to clarify how naturalism and normativity seem to relate to each other in medical practice, which speaks to the dynamics of hybrid models such as Wakefield's. If clarified in the hierarchical way I propose, Wakefield's contribution may be more easily warranted for introduction to practitioners as the correct model for thinking about health and disease.

2/ Mary's Mind

Imagine the case of Mary, who has recently finished home schooling and entered mainstream education as a young teenager. After six months of struggling in class, she has been referred to a child psychiatrist. The child psychiatrist asks Mary various questions regarding her difficulties paying attention, sitting still, making careless mistakes and organising herself. Eventually Mary is diagnosed with Attention-Deficit Hyperactivity Disorder (ADHD). The psychiatrist explains that ADHD is a brain disorder of unclear pathology, resulting from a combination of genetic and environmental factors. Mary is prescribed stimulant medication to hopefully improve her focus in class and the ability to complete her school work.

During the school summer holiday, Mary goes to work on her aunt and uncle's farm. There she assists with all sorts of farm activities: milking cows, mucking stables, fixing fences, driving tractors and so on. She enjoys the activities, is an energetic worker, and is quick to pick up new skills, impressing her relatives. Over dinner towards the end of her stay, Mary's aunt and uncle tell her how well she has done: she has exceeded their expectations, actually proving more useful and competent than others who have worked on the farm before. They tell Mary that they don't think there is anything wrong with her brain – that she is probably just not suited for sitting in classrooms. They hope she will return to work on the farm whenever she can.

Within a period of a few months, in different environments in the same country, Mary has been diagnosed with a mental disorder and then admired for her cognitive capacity. The child psychiatrist considers her unlucky, disabled, in need of restitution towards the average with medical means; the farmers consider her lucky, entirely able, and an example to be followed. These are not merely differing opinions, they are genuine contradictions, resulting in opposite diagnoses of sickness or health.

This example is not merely an implausible thought experiment; it is realistic, reflecting an ongoing conflict between those who claim psychiatry over-medicalises and those who defend current psychiatric diagnosis and treatment approaches as appropriate (an especially common debate regarding ADHD, including by Wakefield himself (Wakefield 2016)). The example of Mary's mind can also be used to consider how naturalistic and normative approaches interact in medical decision making.

3/ Mary's Mind in Normativity and Naturalism

An initial reaction might simply be this: *disorders are categorised normatively.* In moving to the farm, Mary is reclassified from sick to healthy. If the goal is to understand current usage, this example seems to prove disorder (at least in this case of ADHD) is not defined naturalistically.

However, this is too shallow an analysis, as psychiatrist and farmers *believe they are describing naturalistic facts,* not merely expressing socially contingent values. The child psychiatrist believes that Mary has ADHD, a brain disorder of unknown pathology, just as the farmers believe Mary has no such pathology and simply isn't suited to classrooms. Although the psychiatrist has not observed brain pathology directly, nor the farmers observed non-pathological brain differences, both are likely to claim their judgements are related to an objective truth about Mary, and that the other is mistaken in their assumption about the objective health of Mary's brain.

Thus, what seems like evidence of disorder being normatively defined to us as onlookers would be denied as purely normative by the responsible parties, who believe their diagnosis is naturalistically justified. Their attitudes support the naturalist's argument: disorders should be categorised naturalistically. If every agent delineating disorder from health believes that the definition depends on objective facts rather than value judgements, we have a strong case for believing naturalism is the primary conceptual method they use. The apparent normativity occurs because objective causes are inadequately understood, not because value judgements take precedence in principle. There would assumably be no disagreement over whether childhood anencephaly is a disease because agreement would be reached on the objective facts. Still, the fact that Mary's disorder status changes between classroom and school requires some reference to normative influences.

This paradoxical situation clearly favours hybrid accounts of some kind. The normative and naturalistic aspects of our diagnosis of disorder need accounting for. Wakefield's answer is the HDA; only when Mary is in the classroom *and* has a genuine brain dysfunction would we classify Mary's ADHD as a disorder. As a conceptual analysis, this seems approximately right (although it's plausible that a psychiatrist shown proof that Mary's ADHD is a dysfunction would maintain she has a disorder on the farm, albeit one which is temporarily harmless). Yet although Wakefield's definition of 'disorder' moves in the right direction by integrating normative and naturalistic aspects, two key problems arise due to the HDA's stipulation of the harmful and dysfunctional components as equally *necessary*.

Firstly, the necessity of the harm criterion means the HDA fails to satisfyingly answer the problem of cultural relativism originally motivating naturalism. When Mary joins the farm, she no longer experiences harm, so can no longer have a disorder. When she returns to the classroom, she can have a disorder again. If she returns to the farm, the disorder disappears. This is exactly the type of relativism that naturalists want to solve; political rhetoric or normative eccentricity could still eliminate disorder at a whim within the HDA (but not create disorder without evidence of dysfunction, to the HDA's credit). Whether Mary's mind is dysfunctioning in the evolutionary sense is a separate issue to the fluctuating criterion of harm, but both are necessary. Naturalism's aim at objectivity is thus counter-acted by the necessity of accounting for values. The advance provided by the hybrid view beyond simple normativity is also requiring an analysis of evolutionary dysfunction. However, once a

condition is deemed evolutionarily dysfunctional, relativism ensues. This is a bullet that Wakefield can bite – the HDA holds that we will simply update whether Mary has a true 'disorder' depending on her environment. Still, it is notable that this omits any useful role for naturalism once dysfunction is confirmed, which seems to fall short of the naturalists' desire.

Secondly, and more problematically from an ethical and medical point of view, the necessity of the dysfunction criterion in the HDA causes a decoupling of the 'disorder' label from situations worthy of medical intervention. Mary's diagnosis and treatment occurred in response to her disability in the classroom, aiming to improve her life, irrespective of biological facts about the status of her brain as evolutionarily functional or dysfunctional (even if the psychiatrist assumed such biological facts). Harm was perceived, a medical response was deemed justified. What happens if close analysis suggests that Mary's mind is not evolutionarily dysfunctional, simply different, as suggested by the farmers (and certain researchers (Shelley-Tremblay and Rosén 1996; Williams and Taylor 2006))? The harm remains: the education system will not change, and Mary will remain unsuited to it. However, her status as having a disorder *will* change under the HDA. The implication is that medical intervention is less justified – even though the harm, and the potential to alleviate that harm with psychiatric treatment, remains exactly the same.

Here is the primary problem with applying an etiological account to define health and disorder which makes it unwarranted for directing medical practise. Evolutionary classifications as dysfunctional or functional revolve around facts of evolutionary history. This is entirely ignorant of suffering – indeed, much suffering comes from pain, but pain itself has been specifically selected for a functional purpose! And indeed, is often deemed worth treating medically, as researchers in evolutionary medicine and psychiatry note (Nesse and Schulkin 2019). Etiological facts have no necessary relation to the primary goal of medicine: *to help those who seem to need help.* Even if biological function is best understood through the etiological account, medical decisions do not, and arguably *should not* take it as directive; we should not restrict treatment to conditions classifying as evolutionarily dysfunctional. Wakefield actually accepts this:

> "I have argued that if limitations to an individual's opportunity are primarily due to socially negative valuation of parts of normal variation, it is a matter of justice that the individual should be offered treatment, even though the condition is not a disorder." (J. Wakefield, 2021).

The HDA may provide a relatively plausible definition of 'disorder' as generally implied in common parlance or indeed used by medicine, but meets this challenge in cases where the naturalistic identification of function or dysfunction doesn't concord with social values of health or disorder. This is a problem with naturalism generally: Boorse has called his definition of disease 'ultraconservative' and tellingly said "contrary to the usual view, medicine has no essential connection to disease or health" (Christopher Boorse, 2016). Here Boorse seems to be redefining health and disease beyond the limits of acceptability, admitting medicine's goals are disconnected from naturalism. This undoubtedly will raise serious questions in the minds of practitioners and patients. The primary use of 'disorder' terminology is in medical, everyday settings, not in philosophical or scientific research where strict categorisation by objective facts is desirable (see Hunt (2023) for the set of strict evolutionary classifications of 'dysfunction'). In cases of functional systems currently labelled disorders and treated as medical problems, the philosophical account strays too far from everyday use (Hamilton, 2010). This inapplicability is undoubtedly one of the reasons etiological accounts of dysfunction and health have not reached the level of acceptance proponents had hoped for (Godfrey-Smith, 2004).

Thus, the achievements the HDA strives to make by accounting for normativity and naturalism falls short of an infallible hybrid concept reflecting medical norms, both in terms of language and treatment prerogatives. The equal necessity of value and dysfunction are somewhat self-defeating. The fact that Mary moving to the farm removes the disorder means relativism remains – except this version of relativism seems to encourage the ignorance of her suffering if she is struggling in the classroom yet her ADHD is not evolutionarily dysfunctional. Naturalism's aim at objectivity is counter-acted by the necessity of accounting for values, and normativity's aim to help whoever needs help is restricted to only being acceptable when a naturally evolved system is dysfunctioning, ignoring the plenty of times when evolved systems cause suffering (e.g. pain, fever, obesity, anxiety (Nesse, 2019)). The HDA thus falls just short of providing a socially or scientifically satisfying account, because the imperatives of naturalism and normativity directly impede each other; society wants to help those who need help, science wants to act upon facts, but neither can be fulfilled when combined as equally necessary. This is a barrier to it being widely accepted as accurately representing medical sentiment.

4/ Hierarchical Harmful Dysfunction

I here propose a structural supplement to Wakefield's HDA, aligning it more closely with medical practise. The aim is to reframe the constitutive elements in a way which avoids the problems noted above but retains the strengths of the hybrid model. The crux of the alteration here is recognising a *hierarchical* relationship of normative values and naturalistic facts. These interact, but not as criteria which must be simultaneously met as necessary conditions, as in Wakefield's HDA. Instead, the proposal is that normative values are necessary and sufficient conditions to prescribe treatment, whilst naturalistic knowledge plays a role informing those normative values. This is hierarchical in the sense that naturalism is incorporated at the theoretical foundation of values, which then holistically direct action. This may be formulated as 'hierarchical harmful dysfunction' (HHD):

> *It is necessary and sufficient in medicine to act upon values of alleviating suffering and harm, to help those who seem to need help. In forming those values, scientific understanding (naturalism, etiology, medical research, patient experience and so on) is considered in integration with all other values (religious, traditional, political, humanist; social, individual, familial and so on).*

The HHD formulation is supposed to more accurately capture medical goals than the HDA, where each component is necessary. Healing harm is the overriding goal of medicine, but in modern medicine we also deem it important to discern dysfunction – and that scientific understanding has an effect on our overall values regarding treatment in a case-specific manner. Naturalism bears upon deliberations, but values are formed holistically, and eventually decide treatment (and perhaps labels). Discovering evolutionary dysfunction or function may plausibly alter our values regarding a trait and if so, can alter our decision to believe medical intervention (or disorder labels) justified. That we would alter our ascription of disorder upon such findings is the major argument of Wakefield's HDA. The HHD recognises that such findings have no necessary connotations for medical practise – it depends on the surrounding considerations of the case.

Consider how this hierarchical model makes sense of Mary's situation. The initial paradox of apparent normativity enacted by two parties who state their judgements are naturalistic is resolved by recognising that their values are informed to some significant extent by naturalism. Notably no naturalistic evidence was given by farmer or psychiatrist. Assumptions of naturalistic justification were made, and played a pivotal role, to the extent that a naturalistic rebuttal against either party (by proving lack of dysfunction to the psychiatrist or presence of dysfunction to the farmer) would have likely had some effect on their attitude towards medical intervention for Mary. However, the effect

on that attitude change would be entirely dependent on the specific holistic value system of the psychiatrist and farmers. If the farmers are part of a religious sect against medical intervention, or have familial experience with prescription drug addiction, being proven that Mary's ADHD is caused by some pathology would likely not strongly affect their view on her treatment, whether Mary was in the classroom or farm. Their previous naturalistic assertion of Mary's mind not carrying pathology could be mentioned primarily because it justified their other, stronger values, but altering the naturalistic knowledge may not affect their holistic values regarding treatment.

Another advantage of the HHD is that it recognises how naturalism does not in principle completely cede to relativism once dysfunction is proven, unlike Wakefield's HDA. Naturalistic considerations can constantly inform values, even within environments where harms aren't evident. When Mary moves repeatedly in between classroom and farm, the naturalistic evidence remains identical, and dependent on the holistic values of Mary and her peers, the naturalistic evidence may be given more or less weight. Some psychiatrists may hold that evidence of dysfunction becomes irrelevant if the dysfunction causes no harm (as suggested by the HDA), but the hierarchical approach also accounts for cases of psychiatrists who persist in recommending medical treatments and disorder attribution for dysfunctions even in environments where those dysfunctions aren't harmful (perhaps preventatively out of caution; or because they believe that even when harms aren't obvious, dysfunction requires remedy). Recognising naturalism's place informing values explains the relativism of diagnosis and treatment decisions. Although this allows for instances of the relativism which Boorse and the naturalist agenda is against, this simply reflects the reality of how medical judgements and terminology are made; the HHD at least explains this relativism in a slightly more accurate manner than the HDA. This justifies thinkers such as Glackin (2019) on necessary and sufficient conditions of disease not requiring biological justification, but maintains a contributing role for naturalism, alleviating the worry of "a conceptual divorce between human disease and pathology as a biological phenomenon" (Matthewson & Griffiths, 2017). There is no conceptual divorce; there are just overriding factors which come into play.

The second problem with the HDA, of the implication evolutionary dysfunction is a requisite in identifying medically relevant conditions, is obviously avoided in the HHD, which explicitly recognises that medical decisions are partly directed by the scientific picture, but also integrated with the mix of religions, traditions, fashions, ideologies, intuitions, cultural institutions and other individual, familial and social attitudes which affect our values. Understanding of evolutionary function and dysfunction could play a part in this, but are treated as neither necessary nor sufficient by medicine. The HHD model can account for how we currently treat and think about all manner of medical conditions, from HIV to obesity, and medical interventions from contraception to laser eye surgery, regardless of functional or dysfunctional status. It will continue to be applicable to medical enhancement and transhumanism. The fundamental concept is so simple it is almost tautological: we act by values, and those values are shaped by our ideas, which are partly informed by naturalism in scientific societies. Wakefield and Boorse have both recognised the propriety of treating socially disvalued conditions which are not necessarily dysfunctional – and their views here are better explained by the HHD than their own accounts! Even the most ardent naturalists assess medical decisions holistically (if not definitions, where purists may be stalwart without seeming inhumane). This is built into the HHD, meaning it better accords with public and practitioner actions – almost in principle, because it essentially asks the individual 'given this naturalistic fact, do you personally think medical intervention is justified?'. Whether or not decisions are altered given information regarding evolutionary function or dysfunction (or any other scientific evidence) is context-dependent – it depends on how much weight naturalism has in the value system of those individuals, in these cases.

The HHD notably has a different aim to Wakefield's HDA. It does not aim to analyse exactly what we mean by 'disorder', but instead is focussed on how we make decisions as to what we should treat medically. This is supposed to be distinct, but complementary. It illuminates the dynamics behind the critical examples which are medically relevant but don't align with 'disorder' attribution – where the naturalistic, dysfunction component is relegated down the hierarchy, to play a lesser role in foundations of holistic values. It also provides a better account of what a medical response will be to finding out if a given trait of medical concern is evolutionarily functional – we may update our perception of whether the trait is a true 'disorder' (following the HDA) but will only change our treatment decision if the shift in naturalistic explanation outweighs all the other motivations to provide medical help.

Some comments are necessary on how different forms of scientific explanation may play a place in the hierarchical model of medical decision making. The HHD is supposed to be a general model, not unique to evolutionary sciences or Wakefield's model specifically. However, scientific explanations don't necessarily play equal roles in informing values – as Wakefield points out, we likely update our considerations of whether a condition is a disorder or not on the basis of our understanding of whether the condition is evolutionarily dysfunctional. If a trait is given an alternative mechanistic explanation, we don't update in the same way – for example, if an anxiety disorder is supposed to be related to a specific type of activation of the amygdala, but is then found to be related to a different type of activation – or instead, is related to a specific activation of the hypothalamus, we are unlikely to update our consideration of whether the trait is a disorder or not. Similarly, being given different levels of mechanistic explanation, for example between genetic or neuroscientific explanations, we are unlikely to significantly revise our consideration of whether the condition is a disorder. This is the insight of Wakefield's HDA – that the dysfunction component which causes us to update our definition of 'disorder' most profoundly is evolutionarily based. The point of the HHD is to point out that medical decision-making makes use of that information hierarchically – but it is worth considering how such different scientific explanations may differentially affect values in the HHD. Just as we may update a condition's status as 'disorder' upon evolutionary but not neurological science, different scientific explanations likely have different effects on attitudes towards medical treatment.

How evolutionary or other scientific explanations affect value systems relevant to medical treatment is essentially an empirical question of treatment decisions, but we should expect different effects depending on the condition in question, the precise explanation, and interacting holistic values. We might speculate that evolutionary explanations which swap a condition's presumed explanation from dysfunctional to functional (e.g. receiving an evolutionary explanation of Mary's ADHD as an adaptation suitable for foraging (Williams and Taylor 2006)) might be more likely to have a significant effect on treatment decisions than explanations altering the mechanistic explanation of the dysfunction (although such explanations surely have some effects – for example, neuroscientific and genetic ('biogenetic') versus stress explanations have different effects on stigma, where stress explanations reduce hopelessness and biogenetic explanations increase endorsement of medical intervention (Loughman and Haslam 2018)). The longstanding criticism of the 'chemical imbalance' explanation of psychiatric conditions is that it was pushed by pharmaceutical advertisers to encourage psychiatric drug sales – a criticism that seems at least partly true (Kemp, Lickel and Deacon 2014). Evolutionary explanations don't seem to have any such simple direct medical relevance, as they don't suggest a specific ideal intervention. Certain explanations (e.g. of evolutionary mismatch) might have more precise connotations for medical response (e.g. reversing elements of the environment causing the mismatch), but in general, their effects will depend on their effects on value systems, as the HHD recognises – and it won't be merely function/dysfunctional status that matters, but complexities regarding the explanation itself.

5/ Criticisms, Replies and Consequences

A major criticism which might be levied against the HHD would be that, even if it accurately represents current medical treatment, this doesn't make it ideal – naturalism should be at least *necessary* in grounding medical normativity, as in the HDA. Allowing values overriding power makes the approach unacceptably relativistic. The relativism carried by the sufficiency of values has a history of encouraging medical harm. The most obvious example is homosexuality, long pathologized and treated as a mental illness, justifying the application of electrocution, nauseating drugs and stigmatisation of homosexuals. We now recognise this as immoral and harmful. Such environmental relativism is arguably what drives critics such as Boorse (1977) and Woolfolk (Woolfolk, 1999) to defend naturalism.

In reply, advocates of any ultimately value-defined approach must bite the bullet and recognise their approach wouldn't prevent the medical treatment of homosexuality in a society which considers homosexuality a disease. However, it is worth being reminded that any degree to which we veer away from values directing treatment is the precise degree to which we choose to *not* use treatment to best improve lives which seem to need improving. Values identify harm, and we should always try and prevent harm (arguably by definition), so values must be treated as sufficient. The fact that values differ between societies and individuals means we inevitably judge other's application of medicine as somehow unethical – again, to the degree that they diverge from our own. Yet naturalism will not save us here, unless it is guaranteed to lend perfect knowledge of how to remove all conceivable suffering, thus providing the ultimate value – but this possibility is somewhat incorporated into the HHD anyway – in this case, naturalism would (hopefully) easily out-trump all other value systems.

We should also note that the HHD does recognise that values are informed by science, which prevents them being entirely relativistic, and that scientific input can be consequential. There are objective facts related to homosexuality which should be fully investigated and incorporated into medicine – for example, one of the key facts referenced during the 1970s and 1980s when the medical status of homosexuality was under review was that it doesn't cause suffering and can be conducive to a productive and happy life (Taylor 2011). Even etiological accounts could call homophobia into question – there are various hypotheses that homosexuality is an adaptation (Roughgarden, 2017). The fact homosexuality was pathologized for so long was because other systems were informing our values – primarily religious – and the objective evidence surrounding homosexuality's harmlessness was not fully recognised. The HHD model explains the medicalisation of homosexuality, and recognises that the way to undo that medicalisation was not by redefining disease, but by shifting the overall values. The relative loss of religious influence is almost certainly a reason homosexuality was reclassified. Of the various foundations for social norms which direct medicine, naturalism and science have a part to play, but this is competitive – and if naturalists competed in this sphere and proved themselves as holding superior ideas on which to base values, they would effectively promote naturalistic conceptions of disease – as would the weakening of competing value systems which disregard science.

In further defence of values as the suitable necessary and sufficient conditions of medical normativity, although acting upon values allows society to occasionally commit atrocities against disvalued conditions, taking values as sufficient seems more likely to be *protective* against committing atrocities. A purist etiological health and disease advocate is led down a road to curing dysfunction with chilling similarities to Nazi social hygiene and eugenicist approaches. Giving priority to values does allow us to harm people we decide not to value, but also encourages us to *save* people we *decide* to value, disregarding naturalistic analysis that deems them as dysfunctional. On the flip-side, given that much of evolutionary medicine and psychiatry is involved in describing instances when biological goals

(defined evolutionarily) specifically cause health problems (Nesse, 2019; Stearns et al., 2010), falling into the naturalistic fallacy would similarly increase suffering by implying such conditions should *not* be medically treated. The HHD avoids both problems.

Some worry that an overriding reliance on values endangers medicalisation of all aspects of our lives (Schramme, 2007). As Rose (2007) has pointed out, the negative connotations of such medicalisation is unclear. When is it ever bad to apply medicine to improve people's lives? Once again, the alternative is following some theoretically objective principle which allows people to suffer. Painkillers and anaesthetics are not natural – yet we don't say we are medicalising pain and shouldn't prevent it in an operating theatre. Boorse has called certain psychiatric medicalisation 'grotesque' (Christopher Boorse, 2016), but if a child with ADHD has their overall quality of life improved by medication, it seems good to provide it for them, even if ADHD is not a true pathology from a naturalistic perspective. Naturalism could easily lead to *under*-medicalisation, being too restrictive in following biological principles of achieving 'natural' states, failing to help people whose lives we could and should improve.

The danger of medicalisation may exist more in what we might call '*over*-medicalisation', with an obvious recent example being rampant prescriptions of opioids in the USA for minor pain ailments, causing addiction, ruination, overdosing and death to hundreds of thousands. The crucial point is that this actually caused *harm.* If better science had been foundational in informing medicine here – if we had better information about whether these treatments were helping people overall – our values would have changed, and this over-medicalisation would have been prevented. The HHD certainly implies that the ideal value system incorporates the maximum possible scientific knowledge, and the recognition of that knowledge's place in the hierarchy of holistic values also implies that, for most people, the stronger the scientific evidence, the larger an effect on medicine that evidence will have. It also explains why, in the cases where the scientific evidence is poor, or misrepresented, over-medicalisation can occur. The fact we recognise over-medicalisation has occurred in the past can be incorporated into our holistic value system – and for those who were most personally affected, this consideration probably has particular weight.

You can't understand medicine without values, and you can't understand values without naturalism, but you *can* understand medicine without naturalism, because naturalistic considerations may not be incorporated into our deciding values in particular cases. Naturalism forms a *subset* of the competing ideas which inform values. When deciding the correct medical response, we may consider etiology, religion, individual suffering and social acceptance, and eventually decide individual suffering justifies medical treatment, even though the condition is etiologically functional, socially and religiously permissible, and so on. The same condition with the same level of suffering may *not* be treated if it goes against religious doctrine and the individual and their doctor are religious.

Evolutionary definitions of disorder, and indeed the debate over definitions of disorder in the philosophy of medicine generally, have remained largely academic. Normal people, and practitioners, don't consider precise definitions necessary to engage in what is important to medicine: helping people who seem to need help. The long history of a medicine without science, evolutionary or otherwise, has not needed to deeply consider its use of terminology, even once there arose the possibility of objective standards for defining key concepts. Wakefield's HDA provides the best hybrid model for integrating value and objective components, but it could not be warranted as sufficiently explaining medical terminology *and* norms; it may be the best description of 'true disorder', but medicine doesn't only care about true disorder. By better defining the 'dysfunction' component (Hunt, 2023) and then applying the HHD, we can expand on the HDA to better represent the objects and tasks of medicine. Where harm and evolutionary dysfunction exist, we can identify true disorder. Yet medicine takes harm as sufficient to guide it: we will treat whoever we think we need to treat, in

whatever way we can. Those treatment values are decided by a combination of scientific, social and personal values. Evolutionary considerations bear upon whether we think of the conditions as true disorders, but they may or may not affect our treatment decisions.

ACKNOWLEDGEMENTS

Thanks to my various recent supervisors and advisors: in particular Adrian Jaeggi, Hanjo Glock, Randolph Nesse, Riadh Abed and Paul St John Smith. Also thanks to my many philosophy teachers, in particular Samir Okasha. Thanks also to all my peers, with particular thanks to Jordan Martin for enjoyable broad discussions, and to the Human Ecology lab and Glock's colloquium for extensive discussion and support. Special thanks to Jerry Wakefield, whose shoulders are wonderfully sturdy for standing on.

BIBLIOGRAPHY

Barnes, E. (2016). *The Minority Body: A Theory of Disability*. Oxford University Press.

Boorse, C. (1975). On the Distinction between Disease and Illness. *Philosophy & Public Affairs*, *5*(1), 49–68.

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*(4).

Boorse, C. (1997). *A Rebuttal on Health* (pp. 1–134). Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-59259-451-1_1

Boorse, C. (2014). A Second Rebuttal On Health. *Journal of Medicine and Philosophy*, *39*(6), 683–724. https://doi.org/10.1093/jmp/jhu035

Boorse, C. (2016). Goals of Medicine. In E. Giroux (Ed.), *Naturalism in the Philosophy of Health* (pp. 145–177). Springer International Publishing.

Caplan, Arthur. L. (1992). If Gene Therapy Is the Cure, What Is the Disease? In Annas. G. Elias. S (Ed.), *Gene Mapping* (pp. 128–141). Oxford University Press.

Engelhardt, H. T. (1986). *The Foundations of Bioethics*. Oxford University Press.

Ereshefsky, M. (2009). Defining "health" and "disease." *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *40*(3), 221–227. https://doi.org/10.1016/j.shpsc.2009.06.005

Glackin, S. N. (2010). Tolerance and illness: The politics of medical and psychiatric classification. *Journal of Medicine and Philosophy*, *35*(4), 449–465. https://doi.org/10.1093/jmp/jhq035

Glackin, Shane. N. (2019). Grounded Disease: Constructing the Social from the Biological in Medicine. *Philosophical Quarterly*, *69*(275).

Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, *28*(3).

Godfrey-Smith, P. (2004). Mental representation, naturalism, and teleosemantics. In D. Papineau & G. MacDonald (Eds.), *Teleosemantics: New Philosophical Essays*. Oxford University Press.

Hamilton, R. (2010). The concept of health: Beyond normativism and naturalism. *Journal of Evaluation in Clinical Practice*, *16*, 323–329. https://doi.org/10.1111/j.1365-2753.2010.01393.x

Kincaid, H., Zachar, P., Murphy, D., Garson, J., Gerrans, P., Cooper, R., Demazeux, S., De Vreese, L., Lemoine, M., Thornton, T., De Block, A., & Sholl, J. (2021). Defining Mental Disorder: Jerome Wakefield and His Critics. In L. Faucher & D. Forest (Eds.), *Defining Mental Disorder*. The MIT Press. https://doi.org/10.7551/MITPRESS/9949.001.0001

Kingma, E. (2012). *Health and disease: Social constructivism as a combination of naturalism and normativism* (pp. 37–56).

Margolis, J. (1976). The concept of disease. *Journal of Medicine and Philosophy (United Kingdom)*, *1*(3), 238–255. https://doi.org/10.1093/jmp/1.3.238

Matthewson, J., & Griffiths, P. (2017). Biological Criteria of Disease: Four Ways of Going Wrong. *The Journal of Medicine and Philosophy*, *42*. https://doi.org/10.1093/jmp/jhx004

Millikan, R. (1989). In Defense of Proper Functions. *Philosophy of Science*, *56*, 288–302.

Murphy, D., & Woolfolk, R. L. (2000). Conceptual analysis versus scientific understanding: An assessment of Wakefield's folk psychiatry: Reply. In *Philosophy, Psychiatry, & Psychology* (Vol. 7, Issue 4, pp. 271–293). Johns Hopkins University Press.

Neander, K. (1991). Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science*, *58*(2), 168–184. https://doi.org/10.1086/289610

Nesse, R. M. (2019). *Good Reasons for Bad Feelings: Insights from the Frontier of Evolutionary Psychiatry*. Allen Land.

Nordby, H. (2006). "The analytic-synthetic distinction and conceptual analyses of basic health concepts." *Medicine, Health Care, and Philosophy*, *9*, 169–180. https://doi.org/10.1007/s11019-006-0002-7

Reznek, L. (1987). *The Nature of Disease*. Routledge & Kegan Paul.

Rose, N. (2007). Beyond medicalisation. In *Lancet* (Vol. 369, Issue 9562, pp. 700–702). Lancet. https://doi.org/10.1016/S0140-6736(07)60319-5

Roughgarden, J. (2017). Homosexuality and Evolution: A Critical Appraisal. In *On Human Nature: Biology, Psychology, Ethics, Politics, and Religion* (pp. 495–516). Elsevier Inc. https://doi.org/10.1016/B978-0-12-420190-3.00030-2

Saborido, C., & Moreno, A. (2015). Biological pathology from an organizational perspective. *Theoretical Medicine and Bioethics*, *36*, 83–95. https://doi.org/10.1007/s11017-015-9318-8

Schramme, T. (2007). A qualified defence of a naturalist theory of health. *Medicine, Health Care and Philosophy*, *10*(1), 11–17. https://doi.org/10.1007/s11019-006-9020-8

Stearns, S. C., Nesse, R. M., Govindaraju, D. R., & Ellison, P. T. (2010). Evolutionary perspectives on health and medicine. *Proceedings of the National Academy of Sciences*, *107*(suppl_1), 1691–1695. https://doi.org/10.1073/pnas.0914475107

Varner, G. E. (1998). *In Nature's Interests: Interests, Animal Rights, and Environmental Ethics* (Vol. 109, Issue 435). Oxford University Press.

Wakefield, J. (2021). Can Causal Role Functions Yield Objective Judgments of Medical Dysfunction and Replace the Harmful Dysfunction Analysis's Evolutionary Component? Reply to Dominic Murphy. *Defining Mental Disorder*, 267–316. https://doi.org/10.7551/MITPRESS/9949.003.0019

Wakefield, J. C. (1992). *Disorder as Harmful Dysfunction A Conceptual Critique of DSM–III–R's..taxonomy Wakefield 92.pdf*. *99*(2), 232–247.

Wakefield, J. C. (1997). When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorder. *Development and Psychopathology*, *9*(02), 269–290. https://doi.org/10.1017/s0954579497002058

Wakefield, J. C. (2005). Biological function and dysfunction. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 878–902). Wiley.

Woolfolk, R. L. (1999). Malfunction and Mental Illness. *The Monist*, *82*(4), 658–670. https://doi.org/10.5840/monist199982429

Worrall, J., & Worrall, J. (2001). Defining Disease: Much Ado about Nothing? In *Life — Interpretation and the Sense of Illness within the Human Condition* (pp. 33–55). Springer Netherlands. https://doi.org/10.1007/978-94-010-0780-1_3

Wright, L. (1976). *Teleological Explanations. An Etiological Analysis of Goals and Functions*. University of California Press.