Universität Zürich[UZH]

Evolving Evolutionary Psychiatry and Explaining Neurodiversity

Thesis (cumulative thesis)
presented to the Faculty of Arts and Social Sciences
of the University of Zurich
for the degree of Doctor of Philosophy

by
Adam Daniel Hunt

Supervisory Committee:
Prof. Dr. Hans-Johann Glock,
Prof. Dr. Adrian Jaeggi

Specimen copy for the doctoral committee

Zurich, 12/01/2024

Universität Zürich[UZH]

Philosophische Fakultät
Dekanat

Universität Zürich
Dekanat
Rämistrasse 69
8001 Zürich
www.phil.uzh.ch

Graduiertenschule
graduiertenschule@phil.uzh.ch

Overview of Contributions to the Cumulative Dissertation

Title Contribution 1: Specialised minds: extending adaptive explanations of personality to the evolution of psychopathology
Author(s): Adam D. Hunt; Adrian Jaeggi
Journal: Evolutionary Human Sciences
Status (submitted/accepted/published): Published
Date (submission/accepted/published): 31/05/22
If co-authored:

- Indication of own contribution[1]: My roles were conceptualization, methodology, project administration, investigation, visualization and writing – original draft. Prof. Jaeggi was involved in funding acquisition, supervision, and writing – review and editing.
- Clarification of rights to data obtained? Yes, under creative commons license and agreement with Prof. Jaeggi.

Title Contribution 2: Making Wakefield Workable: the fitness and function framework for taxonomising evolutionary dysfunction
Author(s): Adam D. Hunt
Journal: PsyArxiv
Status (submitted, accepted, published): Published (without peer review)
Date (submission/accepted/published): 27/07/23

Title Contribution 3: The DCIDE Method: evolutionary explanations beyond just-so storytelling, exemplified with autism
Author(s): Adam D. Hunt
Journal: Behavioural and Brain Sciences
Status (submitted, accepted, published): Submitted
Date (submission/accepted/published): 09/11/23

Title Contribution 4: Making Wakefield Warranted: the hierarchy of healing harm and discerning dysfunction
Author(s): Adam D. Hunt
Journal: Philosophy, Psychiatry, & Psychology
Status (submitted, accepted, published): Submitted
Date (submission/accepted/published): 04/12/23

Title Contribution 5: Changing Perspectives on Autism: overlapping contributions of evolutionary psychiatry and the neurodiversity movement
Author(s): Adam D. Hunt, Tanya L. Procyshyn
Journal: Autism Research
Status (submitted, accepted, published): Accepted
Date (submission/accepted/published): 03/12/23
If co-authored:

- Indication of own contribution[1]: My roles were conceptualization, methodology, investigation, visualization and writing – original draft. Dr Procyshyn was involved in project administration and writing – review and editing.
- Clarification of rights to data obtained? Yes, by agreement with Dr Procyshyn.

[1] according to CRediT – Contributor Roles Taxonomy

Evolving Evolutionary Psychiatry
and
Explaining Neurodiversity

A thesis presented by Adam Daniel Hunt to the Faculty of Arts and Social Sciences of the University of Zurich for the degree of Doctor of Philosophy

Acknowledgements

First and foremost, I would like to thank my PhD supervisors, Adrian Jaeggi and Hanjo Glock. Their support of my career, and help in securing me the freedom to work within the exceptional academic system at the University of Zurich has been invaluable. In particular, I am grateful to Adrian for offering an opportunity to an eccentric philosopher with a book draft in hand, who did need to touch base with the scientific and structural aspects of academia. I have gained immeasurably from my time as a member of his Human Ecology Lab – I have gained friends, experiences, and most importantly, knowledge of facts and theories previously obscure to me, and which have made the resulting dissertation a vastly stronger work than the book draft which preceded it. Hanjo also made it possible to complete this PhD by taking the role as my primary supervisor in the philosophy faculty, and I appreciate that he integrated me into his regular colloquia, where the more philosophical papers of this dissertation could be discussed and developed. Interdisciplinarity is a difficult line to walk in the academic system, and it is only because of the personal support of both Adrian and Hanjo that I could be in Zurich and complete this PhD despite it (and me) falling outside the bounds of what academia normally supports.

Whilst I wouldn't be positioned in Zurich without Adrian and Hanjo, I wouldn't be positioned in a budding field if it weren't for the community of researchers and psychiatrists who have been working on evolutionary psychiatry for the past few decades, and who have also shown me great charity in involving me in the community, educating me, and introducing me to the others passionate about developing the field. Particularly worthy of mention are Riadh Abed and Paul St-John Smith, who published my first academic writing on evolutionary psychiatry and then offered me interesting conversation, tuition and opportunity to get involved in the evolutionary psychiatry community and resources. Their influence on formally instigating the current evolutionary psychiatry community is unparalleled. The next name worthy of mention is one inseparable from the fields of evolutionary psychiatry and evolutionary medicine – Randy Nesse – who has not only published foundational texts in the field which this dissertation builds upon, but shown great personal kindness and support of my career. If I feel like I am coming into a field with sparse literature and structure, I can't imagine how Randy felt! He genuinely built from scratch something which is now (I hope and believe) beginning to blossom, and to know him personally and have his support has truly been an honour. In further need of mention from the community are those who have provided both fun and illumination: Alfonso Troisi, Martin Brüne, Dan Stein, Henry O'Connell, Annie Swanepoel, Agnes Ayton, Tom Carpenter, Pierpaolo Merola and Nikhil Chaudhary. Further gratitude to Aaron Lukaszewski, Laith Al-Shawaf, Marco Del Giudice and Lars Penke, who straddle evolutionary psychology and psychopathology.

Whilst I wouldn't be positioned in Zurich without Adrian and Hanjo, or evolutionary psychiatry without Riadh, Paul, Randy, and the other acolytes of evolutionary psychiatry, I wouldn't be positioned in *existence* as I am without my family and friends. Mum and Dad, for the heritable contribution, and for raising me with opportunity and stability enough to find my own way – and not trying *too* hard to dissuade me from following my dreams, despite the stress it gave them. Emma, my sister, for being exceptionally bright on paper yet clearly not *too* much brighter than me (despite the grade difference) thus implying that everything was in reach. Charles, for giving Emma happiness, and proving for a second time to the parents that stepping outside the norm *can* be successful. My grandparents; Maggie and Peter for encouraging my writing more than any

other (apologies that you never saw the book in print), and Gaga and Grandad for fun and support – especially in the form of a far-away cottage. Of my friends, undoubtedly the most important nonshared environmental factors in my life, and entirely essential for coagulating the phenotype which led me here: thanks to Dom for first teaching me the (hilarious) value of breaking norms, Tom for co-piloting the throes of youth, Rupert for companionship during the journey into adulthood, Ben for friendship, early and multiplied, Robbie for close observation, Hayley for wild times, and Pierre for my earliest memory. Also to George (especially), Garwyn, Tom and the Camberwell crew for providing a home away from home. Nazli, for being necessary. And of course, the many more who I spent my formative years with, having far too much fun.

During my time in Zurich the two people who I most enjoyed chatting with at length as friends and colleagues, in that magical synergy of both learning and having fun, have been Jordan Martin and Tanay Katiyar. Jordan is an exceptional peer, and it was a rare thing to have someone close-by who is so knowledgeable in relevant areas of expertise and able to push me from the perspective of a rigorous evolutionary scientist. Our conversations undoubtedly led to many of the advances found in this dissertation missing from my original thesis and book draft, in particular the emphasis on measuring fitness and considering by-products (it's also fun to share so many individual differences with another researcher of individual differences). Tanay's internship in Zurich was intellectually stimulating on a different dimension, due to our shared interest in technology and how evolutionary perspectives may inform our understanding of human mental health. It was an honour to play the role of mentor for a young man who has the same inquisitive drive and dedication that I recognize in myself, at an early stage of his career. I am unbelievably lucky to have been able to spend time in a lab where I could find two such firm friends.

Additional mention, for all sorts of reasons, goes to Annina Brunner, Jessie Martin, Kerttu Majander, Zehra, Megnana, Erik Ringen, Charlotte Debras, Abby Colby, Camila Scaff, Cameron Alexander, Sammy Jo-Body, Beatrice Borgts, Samir Okasha and nearly-last but *certainly* not least, Frank Rühli.

The final acknowledgement – necessarily – goes to Alexandra Elbakyan. This dissertation was made possible only because of my access to scientific knowledge when living outside of the academic system for several years as an independent researcher. The fact that such knowledge is paywalled, almost entirely against the wishes of the people who worked on it and the public who funded it, is one of the saddest facts in contemporary human society. The shoulders of giants should be free for all to climb, and Alexandra, and all who advocate for open science, are playing a critical role in the forward march of scientific progress. Every contribution I make in my career will be partly due to them.

Thank You All,

Adam Daniel Hunt,
Zurich, Switzerland, 12/01/2024

Evolving Evolutionary Psychiatry and Explaining Neurodiversity – Introduction

"I think it's really important for me to find out... what happened? And why it happened? There is a very important part of my life I haven't been able to explain."
David Harewood. BBC 'Psychosis and Me' Documentary

1/ Persisting Problems in Psychiatry

Mental health conditions represents some of the greatest modern mysteries, and challenges, facing industrialised societies. Psychiatric conditions are hugely costly and very poorly understood despite decades of effort (Alawieh *et al.* 2012). Developments in neuroscience since the late 20th century have been greeted with expectations of discovering pathophysiology, but revolutionised neither treatment nor explanation, and are now suffering from the replication crisis (Nour, Liu and Dolan 2022). Most neuroscientific studies involve around twenty-five participants, but it is likely that thousands are required to find significant replicable effects (Marek *et al.* 2022). The hope of discovering gross morphological disease markers is now essentially forgotten. Developments in genetics have mirrored this trajectory, moving beyond candidate gene studies and into genome-wide association studies (GWAS), again without finding predicted simple pathological explanations for psychiatric conditions, despite the heredity of mental disorder being widely recognised since the birth of modern psychiatry (Schulze, Fangerau and Propping 2004). Far from discovering pathology, findings in both neuroscience and genetics have blurred boundaries between disease states and normal variation (Geschwind and Flint 2015) and pushed psychiatry into even more uncertainty (Cowan, Kopnisky and Hyman 2002). Thomas Insel, who led the National Institute of Mental Health between 2002 and 2015 and oversaw around $20 billion of research spending (Rogers 2017) said, 'Whatever we've been doing for five decades, it ain't working ... Maybe we just need to rethink this whole approach ... With no validated biomarkers and too little in the way of novel medical treatments since 1980 ... it is time to rethink mental disorders' (Greenberg 2013).

Whilst reductionistic biomedical approaches have vainly sought simple pathology, a paradigm has been emerging which offers a very different type of explanation. Concerned with understanding mental health and disorder within the unifying theory of biology, of evolution by natural selection, evolutionary psychiatry (Del Giudice 2018a; Nesse 2019; Abed and St John-Smith 2022a; Hunt, Abed and St John-Smith 2022) borrows key principles from the better established field of evolutionary medicine (Gluckman *et al.* 2016). Biological causation can be considered at multiple levels, ranging from immediate biochemical and cellular processes to processes of phylogenetic change over millions of years (Mayr 1961; Tinbergen 1963). Psychiatry and clinical psychology have concentrated almost exclusively on the individual, their social surroundings, psychological processes, and biological make-up, but very rarely considered the 'ultimate' level of explanation considering adaptation and function, which is the foundation of evolutionary biology. Psychiatry has lauded the biopsychosocial model (Hunt, Abed and St John-Smith 2022a), but it should also be 'evobiopsychosocial' (Hunt, Abed and St John-Smith 2022b).

If there is a single critical justification for serious consideration of evolutionary psychiatry, it is seen in a paradox recognised for many decades as the 'schizophrenia paradox' (Huxley *et al.* 1964), now framed more generally as the paradox of common, harmful, heritable mental disorders (Keller and Miller 2006). Given that all mental disorders have a genetic component – recognised for over a century by observing familial inheritance, now justified in GWAS – why has natural selection not removed the contributing alleles from the population? Keller and Miller (2006) expected findings of mutation-selection balance would be sufficient, but this hasn't been supported by recent genetic data (Keller 2018). This remains a paradox that the correct explanation must answer – whether neurobiological, or psychological, or genetic – the persistence of these conditions in the human species is an evolutionary mystery. It seemingly defies the most fundamental dynamics of natural selection, which implies harmful traits should be selected out whilst helpful traits propagate. This is particularly paradoxical in the cases of personality-like traits diagnosed as psychopathology – sometimes called 'neurodiverse' traits – which cannot be explained as inappropriate emotional reactions caused by unusual novel environments, but reflect heritable, long-lasting, early onset individual differences. This is a paradox that this dissertation attempts to make some progress with, by recognising necessary evolutionary theory (Chapter 1) and offering improved philosophical foundations (Chapter 2) and scientific methods (Chapter 3) for formulating and assessing evolutionary hypotheses of mental disorder. It then proceeds to consider how evolutionary explanations of mental disorders may affect medicine (Chapters 4), and more specifically, how evolutionary psychiatry relates and to some extent supports the growing neurodiversity social movement (Chapter 5).

It is important to note that the choice of language used throughout this dissertation reflects past and common usage in contemporary mainstream scientific research and society – specifically, terms such as 'mental disorder' 'mental illness' 'psychopathology' and similarly pathologising and medicalised terminology are used, although there are strong arguments for

a shift towards depathologising terminology (Kapp *et al.* 2020). The contributions of this dissertation indeed support depathologised language both scientifically and philosophically, to the extent that they argue that many mental health conditions are not simple dysfunctions. However, because the main target audience is the contemporary mainstream, which still predominantly uses pathologising terminology, that language is retained.

2/ Mysterious Individual Differences

The paradox of common, harmful, heritable mental disorders is a key motivation for developing evolutionary psychiatry, but is related to a problem extending across normal as well as abnormal psychology: why should common, heritable variation in behaviour should exist at all? Why, indeed, should personality exist? Chapter 1 considers this problem, and how the answer relates to explaining psychopathology.

Under classic evolutionary theory, heritable phenotypic variation is expected to be positively selected until fixation or negatively selected until elimination (p.ix. Buss and Hawley, 2011), leading to species- or sex-wide traits. Because of this, evolutionary psychology has often framed itself as revealing the *universal* evolved cognitive architecture of the mind (Lukaszewski *et al.* 2020). Adaptive individual variation of universal tendencies only arise due to plastic reactions to individual circumstances; it is suggested that people differ in, for example, conscientiousness due to their environments activating conscientiousness differentially, not because there is any innate adaptive heritable difference which causes differences in conscientiousness. Evolutionary psychologists have even proposed that personality is simply noise – neutral or maladaptive variation around an optimum (Tooby and Cosmides 1990), either unimportant to selection or awaiting selection. But this is surprising. Human personality is moderately heritable; stable over time, context and culture; shows continuities with non-human personality differences; provides predictive power in forecasting behaviour; and has consequences on evolutionarily-relevant factors such as survival, mating success, status, fecundity and parenting (p.x. Buss and Hawley, 2011). Given these factors, adaptive explanations must be explored. Indeed, other evolutionary psychologists (Buss 2009) have spent significant effort on this. The range of possible explanations and evidence supporting them are introduced in Chapter 1, which then relates individual differences in personality to psychopathology, and lays out the case for expecting shared explanatory evolutionary dynamics.

A critical theoretical foundation is in evolutionary dynamics which don't merely trend towards fixation or elimination in the population: what forces allow adaptive individual differences to persist? At the genetic level, this requires some form of balancing selection, with a mix of positive and negative selective forces which maintain stable variation in the population over multiple generations (Smith 1976; Keller, Howrigan and Simonson 2011). At the phenotypic level, this could occur for various reasons, particularly when social dynamics are at play (Martin, Jaeggi and Koski 2023). In particular, individuals may find different niches in the population – so called social-niche specialisation (Bergmüller and Taborsky 2010; Montiglio,

Ferrari and Réale 2013). In theoretical models and non-human research from evolutionary biology and ecology, this is often considered in the context of possible foraging benefits from specialisation in particular types of food resources (e.g. Araújo, Bolnick and Layman 2011). Negative frequency-dependency is a key force to consider: many strategies are optimal when rare, but become less adaptive the more common the strategy in the population. These dynamics are the essence of classic game theoretic models, for example hawk/dove games, and explain stable sex ratios (Fisher 1930). Fluctuating environmental effects outside of the social environment may also cause balancing selection (Abdul-Rahman, Tranchina and Gresham 2021). Populations of 'Darwin's Finches' on the Galapagos Islands show precisely these dynamics, as seasonal changes alter specific food availability and select for different shaped beaks (Grant and Grant 1989). Chapter 1 begins by outlining these theoretical models and recognising various examples of individual differences across non-human species, including differences in mating strategy, food acquisition, and behavioural syndromes (essentially, animal personalities). A general term of 'specialised minds' is introduced to capture the general evolutionary models explaining individual differences in behaviour as adaptive strategies. Adaptation for specialised minds can involve trade-offs and functional and dysfunctional manifestations or aspects, often context-dependent, with stabilising costs and benefits (potentially maintained via inclusive fitness).

In evolutionary psychology, 'Big Five' personality traits (McCrae and Costa Jr. 2008) have been suggested to have evolved as different strategies with costs and weaknesses (Nettle 2011). For example, high agreeableness may gain from prosociality but suffer due to likelihood of being cheated, whilst low agreeableness may gain from cheating but suffer costs via punishment and ostracism. No personality type is perfectly suited for every situation, and each one has costs, plausibly with benefits and costs arising in a frequency-dependent manner. Chapter 1 closely examines the relationship between traits of personality and psychopathology. Although often separated disciplinarily, the objective facts surrounding them are remarkably similar. Both exist on dimensional spectrums (Kotov *et al.* 2017; Widiger *et al.* 2017), with traits visible in the general and subclinical population, especially among family members, and are common. Brain differences are diffuse and complex (Latzman *et al.* 2021) and heritability is often moderate to high (Polderman *et al.* 2015). Environmental components are mainly non-shared (Plomin 2011). Traits are observable relatively early and stable throughout life, with effects on fitness (Power *et al.* 2013; Allen and Robson 2018). Neither personality nor psychopathological traits have discernible pathological cause in most cases.

In general, personality and psychopathology are distinguished by the presence of harm – we draw delineating criteria at some socially-ordained harmful point on the dimensional spectrum of traits to identify 'disorder' – but this distinction is socially rather than naturalistically derived (discussed further in Chapters 2 and 4). Because of this, shared evolutionary explanations for the naturalistic phenomena of both personality and psychopathology should be considered. One point of note is that theoretical models for the

evolutionary explanation of personality and psychopathology have been developed with their different targets in mind – so personality psychologists have concerned themselves with the adaptive explanation of individual differences (e.g. Nettle 2005; Lukaszewski and Roney 2011), whilst evolutionary psychiatry has concerned itself with the maladaptive consequences of natural selection – for example, referencing trade-offs, extremes of spectrums, defences, mismatch and where selection acts for reproductive success over health (Del Giudice 2018a; Nesse 2019; Abed and St John-Smith 2022a). Chapter 1 suggests that merging these explanatory paradigms is necessary to understand both the adaptive and maladaptive manifestations of the traits' dimensions – with both the maladaptive aspects of personality and adaptive aspects of psychopathology often overlooked.

One final major point of Chapter 1 is an emphasis on the relevance of evolutionary anthropology. In the context of understanding evolutionary forces leading to specialised minds, it is particularly important to appreciate the likely social structure of human ancestral groups. This includes likely size of groups, group composition in terms of age and sex, group interaction dynamics and possible social niches. Chapter 1 finishes by noting that such information provides a principled way to draw a line beyond which adaptive individual differences are essentially impossible to maintain – below a frequency of one individual per group. This allows a non-arbitrary cut-off distinction on the dimensional traits familiar to psychiatry and psychology, addressing a longstanding problem in the philosophy of medicine (Rogers and Walker 2017). Applying such a line, cases of severe schizophrenia can be excluded from adaptive explanations, but more common traits such as attention deficit hyperactivity disorder (ADHD) are eligible for adaptive explanations. This situating of traits within estimated ancestral human social dynamics also becomes important for assessing the sufficiency of evolutionary hypotheses, as developed at length in Chapter 3.

3/ Defining Dysfunction

With Chapter 1 introducing the broad motivations and models inspiring further research into evolutionary psychiatry, Chapter 2 moves toward making fundamental theoretical contributions to the pursuit of evolutionary psychiatry. Its aim is to strengthen one of the unique advantages evolutionary approaches hold over contemporary biopsychosocial approaches to understanding medical disorder: the ability to objectively define 'dysfunction', and therefore delineate health from disorder. The Diagnostic and Statistical Manual of Mental Disorders (DSM) fifth edition (American Psychiatric Association, 2013a) defines mental disorder as: "a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behaviour that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning". It does not further define 'dysfunction' or 'mental functioning', leaving them open to attribution based on social norms and values alone, which is a longstanding gripe of philosophers of medicine (Boorse 1975, 1977; Margolis 1976). Some more objective, scientific approach to defining health and disorder is sorely desired: evolutionary approaches offer a unique and powerful standard, by

referencing evolutionary history and adaptation as a biological norm, deviation from which can be defined as dysfunction and disorder (Wakefield, 1992, 2008; Klein, 1999; Spitzer, 1999).

Jerome Wakefield's (1992, 1997, 2015) 'Harmful Dysfunction Analysis' (HDA) is the most prominent attempt to define 'disorder' and often utilised by evolutionary psychiatry (Hunt, Abed and St John-Smith 2022a). For Wakefield, a disorder must *both* be dysfunctional in the evolutionary sense – where a trait *fails to achieve its selected effects* – and also harmful in an evaluative sense, where an individual or society believe the dysfunctional trait is harmful. This definition of disorder has been widely recognised and is referenced most prominently by the evolutionary medicine and psychiatry literature. However, it has also been the target of various criticisms, particularly found in the book "Defining Mental Disorder: Jerome Wakefield and His Critics" (Faucher and Forest 2021) contained 13 chapters by critics and 15 lengthy responses by Wakefield. These particularly concern the 'dysfunction' component. Chapter 2 aims to meet these criticisms, cementing the unique advantage of evolutionary approaches in defining health and disorder.

The overarching theme of the criticisms against Wakefield's 'dysfunction' component is whether it is philosophically sound or scientifically usable. Faucher (2021) labels this the 'epistemic objection'. Can we actually reveal which biological traits are dysfunctional – to Wakefield, interrupting selected effects – and which are not? This is called into question by several more specific critiques. Firstly, it is possible that *extreme tails are selected* (Kincaid 2021). Evolution could select for a trait dimension, but tails of that dimension might be incorrectly identified as dysfunctional disorder (e.g. if individuals with autistic traits gain more benefits than costs but individuals with diagnosable autism suffer more costs than benefits) – and there is no way to verify which part of the dimension was functional. Secondly, *evolutionary by-products* are common, but have no selected effects, so cannot develop disorders under Wakefield's definition (Murphy 2021). Thirdly, *evolutionary mismatch* (Garson 2021) means social devaluation can cause traits to fail to achieve their selected effect (e.g. if aggression becomes punished rather than admired), but this means the dysfunction component is not as objective as is desired. Lastly, there is *indeterminacy of function* (Thornton 2021). Specifying a single function such as 'eyes evolved to see' is appealing, but when extrapolated to psychological traits such as aggression or jealousy, it is much harder to determine such specific functions which can then be called dysfunctional when interrupted.

Wakefield generally responds by either re-affirming that in precise counter-examples, an analysis would indeed identify dysfunction correctly, or claims that it is a matter of waiting for scientific development until such analyses could be undertaken. However, in Chapter 2, other complexities are noted which further trouble Wakefield's stipulation of the dysfunction component. If 'interrupting selected effects' implies dysfunction, we could suggest that lust interrupts the selected effect of rational thinking, or excitement interrupts the selected effect of sleeping – adaptations interrupt the selected effects of other adaptations, so identifying selected effect interruption does not suffice to prove evolutionary dysfunction in the way Wakefield aims to capture. It is also unclear what extent of interruption should count as a

dysfunction. For example, eyes seeing – does every mutation or environmental effect that has a tiny negative effect on sight count as a dysfunction? Is all eyesight except perfect eyesight dysfunctional? Taking these problems alongside those of his critics, a strict tractable scientific paradigm investigating whether traits are dysfunctional seems beyond reach to Wakefield's specific definition. Without a scientifically tractable definition of evolutionary function and dysfunction that can be applied to mental disorders, this unique advantage of evolutionary psychiatry – plausibly the one paradigm which can offer a truly objective definition of function and dysfunction, health and disorder, by having an objective historical standard to appeal to – cannot be utilised. This is the problem that Chapter 2 aims to solve.

The core advance Chapter 2 makes is expanding the function/dysfunction dichotomy into the 'fitness and function' framework (FAFF) of definitions. This distinguishes the objects of selection (e.g. adaptations such as eyes, disruptions such as schizophrenia) from the processes causing selection (e.g. functions such as seeing, dysfunctions such as paranoid thinking), defining these objects and processes with the three possible directions of selection – positive, neutral and negative. The particular advantage of this definition derives from its move of emphasising *fitness effects* as well as functional and dysfunctional processes which cause those effects. The empirical programs required to show biological objects' effects on reproductive success (fitness) are relatively straightforward, because they rely on a fundamentally *quantifiable* variable – number of offspring – compared to the complex and fundamentally *qualitative* processes of function and dysfunction that affect fitness. Ascertaining fitness is relatively simple; ascertaining effects of function or dysfunction is not. We can measure and agree upon the negative effects of schizophrenia on reproductive success fairly uncomplicatedly, but precisely describing the dysfunctional effects of schizophrenia is a problem of entirely different magnitude. The FAFF leapfrogs this problem, allowing the attribution of dysfunctional *status* by observing negative fitness effects without needing to specify the exact dysfunctional *processes* responsible. It makes it possible to say "there is dysfunction – we know not what". The simplicity and tractability of measuring or inferring fitness effects becomes the key scientific metric for distinguishing dysfunction, and thus identifying disorder.

The FAFF provides a definitional framework for better investigating dysfunction, but to use it with precision and surmount the criticisms aimed at Wakefield, two additional considerations are required. These are *specifiers*. The first specifier involves clearly identifying the trait to be studied and categorised. Fitness and function must be measured *of* something. This is called *description*. The second specifier clearly identifies a reference environment. Reproductive success only occurs due to fitness within a particular environment, and a trait's success is relative to that environment. This is called *framing*. Both of these specifiers are not just desirable, but genuinely necessary for any sound application of evolution-based definitions – if any theorist proposes an evolutionary hypothesis without specifying exactly what trait they are referencing, or an environment in which the trait is supposed to manifest, we cannot accurately judge that hypothesis. Eyes are only functional in the presence of light. Evolutionary

classification of depression is impossible without specifying what 'depression' means and in which environment it is being assessed. Traits can be functional for hundreds of thousands of years but become dysfunctional when the environment changes; or vice versa. Describing anxiety in its most extreme form may find negative effects on fitness, whilst describing it in a more general form could find positive effects. When these specifiers are ignored or left implicit, the resulting analysis is inevitably imprecise.

The basic process of assessing a trait with the FAFF is to describe it, frame it in a particular environment, and investigate its fitness and the processes by which it achieves them. Most importantly, the FAFF and specifiers provide a framework for attributing dysfunction which is widely applicable (it can be applied to any biological trait or process) and exclusive and strict (no analysis can result in fitting into more than one category, because fitness effects can only be positive, neutral or negative). It also allows the attribution of dysfunction to complex cases which troubled Wakefield's dichotomy, where altered specifiers describe by-products (the criticisms of Kincaid and Murphy) or frame traits in environments which exacerbate dysfunction (the criticism of Garson). This makes it a more solid theoretical foundation upon which to clarify the existence of dysfunction.

The FAFF's emphasis on quantitative definitions relying on fitness can be utilised to make the taxonomising of dysfunction more strict. Existing taxonomies in evolutionary psychiatry outline broad categories of *processes* contributing to dysfunction; they are filling the necessary void in qualitative explanation, just as 'eyes function to see' or 'depression functions to disengage'. However, they also carry complexities of the indeterminacy of function (here, the indeterminacy of dysfunction), and are not mutually exclusive or strictly demarcatable. For example, Nesse's 'six reasons for disease' (Nesse 2005) are mismatch, host-pathogen co-evolution, trade-offs, constraints, defences and traits which increase fitness at the expense of health; but a disorder could simultaneously be a mismatched defence with trade-offs (e.g. social anxiety overactivated by social media). Abed and St John-Smith (2022b) suggest additional explanatory processes of life history factors, sexual selection and its consequences, balancing selection and heterozygote advantage, demographic history, deleterious alleles and extremes of adaptations; but a disorder could simultaneously be an extreme of a sexually-selected adaptation influenced by life history factors (e.g. extreme risk-taking in young men). For many of these explanatory categories, then, there are no strictly demarcated empirical definitions (e.g. of 'defence'; although others such as 'heterozygote advantage' are more specific).

In contrast to this, the FAFF's primary utilisation of quantifiable fitness leads to classes of dysfunction within a *complete* and *strictly demarcated* framework. Once description and framing are specified, there will never be a biological object or process which doesn't fit into *only one* category of affecting fitness positively, neutrally or negatively. Altering description and framing specifiers allow the parsing of complexities by shifting focus of analysis or environment. This makes for a crisper (but less qualitatively informative) taxonomy of disorder, rendered in Chapter 2 into five distinct classes. This is especially suitable for porting

into the HDA and addressing Wakefield's critics: it draws unambiguous lines. Other taxonomies are useful for describing common processes of dysfunction, but inevitably imprecise, which can lead to downstream difficulty in designing experimental paradigms of direct investigation or inference. It is very easy – indeed, automatic – for there to be strict empirical criteria when using a fitness-based framework such as the FAFF. Chapter 2 deliberately avoids returning to the HDA's specific definitional question of what constitutes 'disorder', but does note that the different classes of dysfunction include one which is unambiguously 'pathological' whilst others are much less likely to accord with concepts of true 'disorder' (e.g. where novel environments cause dysfunction).

Once the FAFF is introduced, the major criticisms of Wakefield's rendition of the 'dysfunction' component are surmounted. The central scientific question moves from being qualitative to being quantitative. Where the HDA *needed* to describe the functional processes being interrupted, which requires some complex qualitative story, the FAFF only needs to appeal to evidence of reproductive success – which may be inferred from selection on genes, facts of epidemiology or demographic characteristics, or directly measured in modern reproductive success. Developing a scientific method utilising such evidence to make evolutionary explanations systematic and rigorous is the aim of Chapter 3.

4/ A Method Beyond Just-So Storytelling

Evolutionary hypotheses to explain observed phenotypes are fundamentally historical, often making claims about processes occurring over hundreds of thousands or millions of years, now completely unobservable directly. Since Darwin's proposition of the theory, this has caused problems in scientific investigation, with Popper even labelling it as not testable and thus not a true scientific theory (although he later accepted its validity; Sonleitner 1986). Today, evolution can be directly observed in experimental setups with species with short generation times, in 'natural experiments' that see populations change in response to environmental changes (e.g. the famous 'peppered moth' example; Majerus 2008), in artificial selection in breeding, and in techniques allowing inference of phylogenetic trees by comparing phenotypes and genotypes between species and populations, so the fact that evolution by natural selection occurred and explains adaptive biology is essentially uncontested. However, the evolution of human psychopathological traits is essentially un-interrogable by these methods. This is the scientific side of the epistemic objection levelled against Wakefield and evolutionary psychiatry, claiming it leads to fundamental unsuitability as the next paradigm for psychiatry (Zachar and Kendler 2017).

Given the difficulty of observing human evolution in action, or utilising the methods available to wider evolutionary biology, research in evolutionary psychology usually takes a particular structure, of aiming to directly reveal precisely designed adaptations in action. This begins by hypothesising about a likely adaptation of human cognition, given *a priori* considerations of what evolution would select for. Specific design features of human cognition are suggested, given the proposed adaptive challenge that the cognition supposedly evolved to solve. Then,

experiments are designed and conducted to reveal that human psychological mechanisms do indeed work in this way (Cosmides and Tooby 2013; Lewis *et al.* 2017). This is often bolstered by cross-cultural work showing that the features of human psychology are universal, as predicted of most adaptations.

This scientific strategy has long been criticised as speculative 'just-so storytelling' (Gould and Lewontin 1979; Dupré 2003; Smith 2019). The 'just-so storytelling' critique is a reference to Rudyard Kipling's 'Just So Stories' for children, which explain the characteristics of various animals in imaginative, yet clearly unscientific, ways. The general criticism here is that adaptationist post-hoc hypotheses can be conjured up to fit around observed biological traits with very little scientific merit. More specifically, there are claims that evolutionary hypotheses are unfalsifiable, lack sufficient testable predictions (McCain and Weslake 2013), and are excessively reliant on modern observations of functioning (Henrich, Heine and Norenzayan 2010). Our evolutionary history is still invisible, and this constrains hypotheses (Smith 2019). Evolutionary psychologists typically counter this critique by noting that only selection can produce complex design in nature (Tooby and Cosmides 2015; Lewis *et al.* 2017) building on the classic definition of adaptation by George Williams (1966), and persist in experimental paradigms which try to reveal psychological mechanisms in precisely the way expected of the adaptations they propose. However, the criticism of just-so storytelling remains commonly cited – this response has not been treated as sufficient by detractors, and a more complete or rigorous scientific method has not been developed.

These problems are exaggerated in evolutionary psychiatry. Evolutionary psychology deals with familiar psychological traits with somewhat intuitive explanations. The invisibility of ancestors feeling jealous in romantic relationships and adaptively mate-guarding over hundreds of thousands of years due to that jealousy (Buss and Haselton 2005) might elicit limited criticisms of just-so storytelling, but retains its plausible function in industrialised societies. Mental disorders are, essentially by definition, not viewed as positive or adaptive traits, so evolutionary explanations referencing adaptive processes or functions are immediately less intuitive and directly observable. They justifiably require a higher standard of evidence. Uncertainty of functional role – indeed, doubt whether there is function at all – undermines the strategy of hypothesising of complex design features to be ascertained experimentally. Mismatch with modern environments might mean that adaptive traits are now functioning quite differently – even dysfunctionally – despite historically being adaptive (related to Chapter 2's framing specifiers). Also, additional complexities regarding correct objects of explanation abound in psychiatry (related to Chapter 2's description specifiers). Disorder categories are highly heterogenous and may contain both functional and dysfunctional subtypes – some cases of 'social anxiety' may be adaptive, some maladaptive (Nesse 2022). Relevant subclinical manifestations of a particular disorder may be disregarded from analysis due to clinical focus on disease categories rather than quantitative traits (a problem noted in both Chapters 1 and 2) despite genetic relationships and shared causality between subclinical and clinical forms (Smoller *et al.* 2018). If so, evolutionary explanations

aimed at the disorder itself will be missing the key phenotype with adaptive benefits – akin to trying to explain sickle cell disease without considering the heterozygous carriers benefitting from malaria resistance.

Despite the great fundamental theoretical appeal of evolutionary psychology and psychiatry, and the paradox of common, harmful, heritable mental disorders demanding resolution, explaining psychological and psychiatric traits evolutionarily is thus troubled by several limitations. Chapter 3 aims to overcome them by providing an improved systematic approach to formulating and evaluating evolutionary hypotheses. A vast array of relevant evidence exists (e.g. across genetics, neurobiology, epidemiology and psychology), but attempts to integrate them and make the best inference of evolutionary explanation currently proceeds without a theoretically-justified structure or standardised analytical principles. Chapter 3 offers a resolution by laying out the 'DCIDE' method of systematic review, an acronym standing for its five stages of Description, Categorisation, Integration, Depiction and Evaluation. The problem of uncertainty in evolutionary explanation is a problem born of its historical nature, but better progress can be made by formally structuring evidence to suit evolutionary inference. The DCIDE method provides such a structure, allowing the assessment and comparison of evolutionary hypotheses, avoiding criticisms of just-so storytelling or cherry-picking as much as possible.

In the DCIDE method, a trait is firstly clearly Described (noted in Chapter 2 as a necessary specifier). Categorisation then distinguishes a major kind of heterogeneity in causation, separating cases primarily explained by non-adaptive factors (e.g. *de novo* mutations) from cases potentially requiring some form of adaptive explanation. Non-adaptive cases are considered explained and removed from the ongoing analysis, whilst cases requiring adaptive explanation are taken forward for Integration. In Integration, the Described trait is assessed for its visibility to selection in a standardised manner: its age of effect, duration, sex differences, environmental effects and prevalence are considered as indications of the likelihood of the phenotype in question being visible to evolutionary selection pressures (e.g. does it appear at life stages which will affect reproductive success; if not, does it have correlates which do appear at those stages). Where visibility is low, relevant genetically related traits are necessarily Integrated so that evolutionary analysis is likely to capture the phenotype under selective pressure – in the case of sickle cell disease, this means including heterozygous individuals. Description, Categorisation and Integration thus hone the target of explanation. The stage of Depiction presents evolutionary hypotheses attempting explanation of those traits, which put forward direct evidence of the trait's supposed function and suggest an evolutionary model. Finally, in Evaluation, those hypotheses are assessed in a standardised manner, firstly following normal scientific procedures asking about their reliability, compatibility, and relevance, before the more novel stage of Evaluation of their sufficiency to explain the circumstantial evidence (e.g. do the hypotheses sufficiently explain the precise age of effect and prevalence of the Described and Integrated traits). The result is a full analysis and direct comparison of competing evolutionary explanations for the Described and

Integrated traits, taking into account relevant available evidence in a principled manner, with specific directions for future research and easy integration of new evidence or novel competing hypotheses.

Chapter 3 primarily exemplifies the DCIDE method with autism ('autism spectrum disorder' in current diagnostic manuals). Autism is chosen due to its phenotypic and etiological heterogeneity (Mottron 2021), large body of available evidence, history of evolutionary analysis (Del Giudice 2018b) and social importance, particularly to the neurodiversity (Kapp *et al.* 2020) movement (returned to in Chapter 5). After Describing autism using standard (notably broad) diagnostic criteria, in Categorisation a significant proportion of cases of autism (approximately 5-20%, depending on diagnostic criteria) are recognised as explained by non-adaptive processes (traditionally recognised as pathological), such as *de novo* mutations and foetal damage. These don't require explanations regarding positive selection and extended evolutionary history. The idiopathic cases (often associated with less intellectual disability or co-occurring health conditions) are taken forward for further analysis. Then, in Integration, analysis shows that the visibility of autism itself to selection is high on most variables, implying evolutionary explanations are indeed necessary – with the main exception of its rarity in the population. This means that recognising the full target of evolutionary explanation requires recognising subclinical and familial manifestations, sometimes known as the Broad Autism Phenotype (BAP). Female manifestation is also likely missed, meaning better incorporation of less pronounced autistic phenotypes in females might be important (Hull, Petrides and Mandy 2020). In Depiction, three evolutionary hypotheses are considered (Del Giudice *et al.* 2010; Crespi 2016; Baron-Cohen 2020), which are condensed to two hypotheses in Evaluation, of the systemising social niche specialisation hypothesis and the high intelligence by-product hypothesis. These hypotheses are Evaluated for their relative sufficiency in explaining autism's visibility variables, and a general pattern emerges: for the by-product hypothesis, the circumstantial evidence is theoretically surprising and needs to be excused as a failure of natural selection, but for the social niche hypothesis, the evidence generally appears as expected. Given the DCIDE review, the best evolutionary explanation for autistic traits, including diagnosable autism but excluding non-adaptive cases, is that advantages in systemising and related cognitive enhancements lead to individuals consistently filling valued roles throughout societies over human evolutionary history, with social disadvantages that came as trade-offs.

Although autism is the primary example of the chapter, and the DCIDE method is particularly suitable for formulating evolutionary explanations of psychopathological traits, it is a general inferential method for evolutionary analysis and could thus be widely applied. The most obvious benefits outside evolutionary psychiatry are in evolutionary psychology. Stages of Categorisation and Integration may be less applicable, because objects of analysis in evolutionary psychology are likely to contain less pathological cases or traits with low visibility to selection, but the stages of Depiction and Evaluation can still be of significant benefit. In particular, they offer a stronger structure for comparing the relative merit of competing

hypotheses, and prevent cherry-picking and just-so storytelling by forcing systematic analysis of relative sufficiency in explaining circumstantial evidence. Upon concluding an Evaluation, a given hypothesis must have explained the age of effect, duration, sex differences, environmental effects and prevalence of a trait, all of which have standardised theoretically derived interpretations (e.g. age of effect implies life stage, which narrows down functional hypotheses). Chapter 3 concludes by noting that any explanatory hypothesis must explain this range of evidence: whether explanations are psychological, social, pathological, neuroscientific, and so on. The best hypothesis is the one that explains the evidence most sufficiently. Utilising the DCIDE method of systematic review is a general format for comparing hypotheses to the same, theoretically sound standard, making general evolutionary inference stronger, and somewhat combatting the problem of scientific hypothesising about the historical processes of selection.

5/ The Hierarchy of Harm and Dysfunction

With Chapters 1-3 providing motivation, theoretical foundations and an improved scientific method for evolutionary hypothesising regarding mental health and disorder, Chapters 4 and 5 consider the practical relevance of such hypotheses.

Wakefield's HDA explicitly recognised the relevance of social values in defining 'disorder', making harm a necessary joint criterion alongside the 'dysfunction' component developed in Chapter 2. Chapter 4 provides a renewed theoretical analysis of the dynamics of this value component, moving beyond Wakefield's definitional question and asking where evolutionary (and indeed, all scientific) approaches interact more generally with medical norms and values. The place of science versus social values in defining medical terminology is a core debate in the philosophy of medicine, usually framed as the conflict between normativity and naturalism (Schramme 2007). Wakefield's HDA is a hybrid account, with both normative and naturalistic criteria as necessary features of disorder. Chapter 4 embellishes upon the hybrid account with a more generalisable *hierarchical* hybrid account of how naturalism and normativity interact.

Normative approaches regard medical classification as value judgements reflecting social value placed on particular biological states: broadly, healthy states are desirable, disease states are undesirable (Engelhardt, 1986; Margolis, 1976). Objective biological facts may be irrelevant: 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). One strong point in the normative approaches' favour is that normative considerations are already, and always have been, the impetus behind medical treatment. This is evidenced by the contradiction of a goalless medicine, and the fact that despite almost all of the targets of psychiatry being poorly understood scientifically, their medicalisation has long been justified. Indeed, medicine is many millennia older than science, and practised by

non-scientific communities and individuals – clearly concepts of health and disease are not inherently scientific (Murphy & Woolfolk, 2000).

Where does scientific understanding sit in relation to modern medical judgements, then? In Chapter 4, a realistic thought experiment is proposed, of an individual with ADHD who goes from being diagnosed as disordered to showing above average competence by changing environment from a classroom to a farm. In dissecting the example, it seems that most contemporary parties do presume some objective – or naturalistic – component explaining her cognition, but may disagree about whether it is healthy and functional or disordered and dysfunctional depending on the environment and display of competence (presuming no direct scientific evidence either way). The implication is that medical judgements are not *merely* informed by pure values. Scientific explanation, and reference to dysfunction, seems inherent, and Wakefield is probably right that we, as believers in science, wouldn't attribute 'disorder' without believing in some sort of objective dysfunction.

However, Chapter 4 notes some oddities of Wakefield's approach, if we are to assume that this application of 'disorder' terminology is supposed to be relevant to medical practise. Firstly, the *necessity* of the value component effectively means the dysfunction component can be relegated to total unimportance: if a dysfunction is confirmed beyond doubt to exist, but specific social circumstances mean it is not considered medically harmful, they overrule that scientific knowledge entirely – a cancer patient moving to a society where the cancer is considered a blessing no longer has a disorder, regardless of the dysfunction. This complete relegation of the naturalistic component seems uncomfortably suboptimal. Secondly, the *necessity* of the dysfunction component implies that we could observe clear harm but somehow de-emphasise the imperative to reduce it when caused by an evolutionarily functional system. If we find out ADHD is evolutionarily functional there is *some* implication that we shouldn't be treating it, even if that treatment helps diagnosed individuals. We should help those who need help – this seems a self-evident moral imperative – but pushing for the necessity of the dysfunction component seems contradictory to this. Many functional evolved systems cause incredible amounts of harm (indeed, pain itself is functional) so restricting attribution of 'disorder' to dysfunctional systems may be appealing definitionally, but clearly misaligns with the goals of medicine. Indeed, Wakefield recognises that we should often apply medicine to heal harm irrespective of disorder status (Wakefield 2021). This creates a barrier to accepting the relevance of evolutionary perspectives to medicine – if they are fundamentally disconnected with treatment decisions, why bother considering them?

In Chapter 4 a hierarchical hybrid model is proposed to better fit how naturalistic and normative considerations interact in practise, reframing the relevance of evolutionary explanations to medicine. This 'hierarchical harmful dysfunction' model proposes that normative considerations are necessary and sufficient conditions to direct medical treatment, but naturalism plays a key role *in informing normativity*, interacting with a range of relevant political, religious, personal and social beliefs. Naturalism bears upon deliberations, but values are formed holistically, and eventually decide treatment. Discovering evolutionary dysfunction

or function may plausibly alter our values regarding a trait (e.g. evidence that ADHD is evolutionarily functional may make us think differently about the proper place of stimulant medication) and if so, can alter our decision to believe medical intervention justified, but have no *necessary* connotations for medical practise – it depends on the surrounding considerations of the case, and the competing values involved. This addresses the criticisms which arise due to the equal necessity of naturalism and normativity in Wakefield's hybrid model: naturalism is never completely irrelevant, and medicine continues to help whoever seems to need help, whether or not the trait in question is evolutionarily dysfunctional. The relevance of evolutionary explanations to making medical decisions is the same as the relevance of any science to making medical decisions – they make us better informed and may or may not affect treatment decisions. Like Chapter 2, this is not supposed to be a commentary on the correct attribution of 'disorder' – it is merely supposed to provide a better account of how insights from evolutionary psychiatry affect treatment decisions.

6/ Evolutionary Psychiatry and Neurodiversity

Chapter 5 concludes the dissertation with an elaboration on a specific area where evolutionary psychiatry may have practical effects, in relation to the growing 'neurodiversity' social movement (Kapp *et al.* 2020). The neurodiversity movement seeks reconceptualisation of long-lasting mental disorders of the sort discussed in Chapter 1 – particularly autism and ADHD, but also dyslexia, bipolar, and other conditions – which they claim are harmfully pathologised. The movement has been explicitly socially rather than scientifically justified, but in Chapter 5 areas of alignment between evolutionary psychiatry and the concept of neurodiversity are explored. Their shared aims and connotations are plentiful and remarkable. Both can reframe mental disorders as natural cognitive differences rather than disease; both expand the concept of 'normal' beyond contemporary mainstream psychiatry's standards; both encourage understanding of psychiatric conditions in terms of relative strengths; both recognise that modern environments unfairly disadvantage certain individuals and cause functional impairment; both emphasise socially accommodation and integration of cognitive variation rather than treatment or cure; and both may potentially reduce stigmatisation.

Despite these alignments, there are areas of differential emphasis and clear disagreement. Evolutionary psychiatry emphasises scientific explanation and objectivity, whilst the neurodiversity movement emphasises social values and action, and removing barriers to inclusion. Their direct disagreement is most pronounced in the extent to which cases are reconceptualised away from traditional pathological models. The neurodiversity concept allows for unrestricted identification, so could include non-adaptive cases such as Down Syndrome. Evolutionary psychiatry must specifically differentiate heterogenous cases based on causation, as seen in Chapter 3: not all cases of autism are eligible for the same explanation, and the alignment with neurodiversity perspectives arises for cases which require some adaptive explanation. Evolutionary psychiatry is thus more restrictive.

Chapter 5 considers the consequences of this difference, and what role evolutionary psychiatry may therefore play as a scientific paradigm supporting the neurodiversity movement. One point of note is that evolutionary psychiatry may lend scientific credibility and thus make the aims of neurodiversity advocates more appealing, particularly to parties who value scientific knowledge highly. If reconceptualisation of mental disorders is the aim, evolutionary explanations seem to have an automatic influence which aligns, in many cases, with the neurodiversity movement, and will inform social values regarding these conditions, in the dynamics noted in Chapter 4. Furthermore, despite evolutionary psychiatry not perfectly aligning with the inclusivity of the neurodiversity movement, this difference directly talks to a major point of controversy: the concern that the neurodiversity movement is aimed and relevant for the least disabled cases, but ignorant of individuals with severe disability (Clements 2019; Nelson 2020; Escher 2023). If parsing the various categories of mental disorder by evolutionary principles, as in Chapter 3, the scientific facts somewhat explain this disagreement. Many mental disorder categories are loosely defined (or 'described', as in Chapters 2 and 3) and include clearly differentiated etiological causes, with the adaptive/non-adaptive line generally aligning with levels of disability. If both sides of the debate recognise this fact, it may simmer disagreement and lead to more productive action, encouraging useful accommodations for individuals displaying a range of abilities and disabilities.

Altogether, then, this dissertation presents the theoretical motivation for developing evolutionary accounts of mental traits currently diagnosed as psychopathology (Chapter 1), provides a solid foundational definitional framework which allows for precise empirical identification of dysfunctional biology (Chapter 2), lays out a novel method of systematic review of scientific evidence to best formulate and analyse evolutionary hypotheses in a rigorous manner (Chapter 3), considers how such explanations interact with values to direct medical treatment (Chapter 4), and finishes by relating this to a real world example, of how evolutionary explanations align with the neurodiversity movement (Chapter 5). It thus aims to forward evolutionary psychiatry, make progress in explaining neurodiversity, and consider the impact these explanations may have.

Bibliography

Abdul-Rahman F, Tranchina D, Gresham D. Fluctuating Environments Maintain Genetic Diversity through Neutral Fitness Effects and Balancing Selection. *Mol Biol Evol* 2021;**38**:4362–75.

Abed R, St John-Smith P eds. *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge: Cambridge University Press, 2022a.

Abed R, St John-Smith P. Introduction to Evolutionary Psychiatry. In: St John-Smith P, Abed R (eds.). *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge: Cambridge University Press, 2022b, 1–18.

Alawieh A, Zaraket FA, Li J-L *et al.* Systems biology, bioinformatics, and biomarkers in neuropsychiatry. *Front Neurosci* 2012;**6**:187.

Allen MS, Robson DA. A 10-year prospective study of personality and reproductive success: Testing the mediating role of healthy living. *Psychol Heal* 2018;**33**:1379–95.

Araújo MS, Bolnick DI, Layman CA. The ecological causes of individual specialisation. *Ecol Lett* 2011;**14**:948–58.

Baron-Cohen S. *The Pattern Seekers: How Autism Drives Human Invention*. Basic Books, 2020.

BBC. *Psychosis and Me*., 2019. https://youtu.be/MBduR0BU1vg?si=OL6fS35Ztwa93hnu (Quote at 2:40)

Bergmüller R, Taborsky M. Animal personality due to social niche specialisation. *Trends Ecol Evol* 2010;**25**:504–11.

Boorse C. On the Distinction between Disease and Illness. *Philos Public Aff* 1975;**5**:49–68.

Boorse C. Health as a theoretical concept. *Philos Sci* 1977;**44**.

Buss DM. How Can Evolutionary Psychology Successfully Explain Personality and Individual Differences? *Perspect Psychol Sci* 2009;**4**:359–66.

Buss DM, Haselton M. The evolution of jealousy. *Trends Cogn Sci* 2005;**9**:506–7.

Buss DM, Hawley PH. *The Evolution of Personality and Individual Differences*. Oxford University Press, 2011.

Clements T. What is autism? How the term became too broad to have meaning any more | Tom Clements | Opinion | The Guardian. *Guard* 2019.

Cosmides L, Tooby J. Evolutionary Psychology: New Perspectives on Cognition and Motivation. *https://doi.org/101146/annurev.psych121208131628* 2013;**64**:201–29.

Cowan WM, Kopnisky KL, Hyman SE. The Human Genome Project and Its Impact on Psychiatry. *Annu Rev Neurosci* 2002;**25**:1–50.

Crespi B. Autism as a disorder of high intelligence. *Front Neurosci* 2016;**10**:300.

Dupré J. *Human Nature and the Limits of Science*. Oxford: Clarendon Press, 2003.

Escher J. The Autism Surge: Lies, Conspiracies, and My Own Kids | The Free Press. *Free Press* 2023.

Faucher L, Forest D eds. *Defining Mental Disorder: Jerome Wakefield and His Critics*. The MIT Press, 2021.

Fisher RA. *The Genetical Theory of Natural Selection*. Oxford: The Clarendon Press, 1930.

Garson J. The Developmental Plasticity Challenge to Wakefield's View. *Defin Disord* 2021:335–52.

Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. *Science* 2015;**349**:1489–94.

Del Giudice M. *Evolutionary Psychopathology: A Unified Approach*. New York, NY: Oxford University Press, 2018a.

Del Giudice M. The Autism Spectrum. *Evolutionary Psychopathology*. Vol 4. 2018b, 243–60.

Del Giudice M, Angeleri R, Brizio A *et al.* The evolution of autistic-like and schizotypal traits: A sexual selection hypothesis. *Front Psychol* 2010;**1**:1–18.

Gluckman P, Beedle A, Buklijas T *et al. Principles of Evolutionary Medicine (2nd Edition)*., 2016.

Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc London - Biol Sci* 1979;**205**:581–98.

Grant BR, Grant PR. Natural Selection in a Population of Darwin's Finches. *https://doi.org/101086/284924* 1989;**133**:377–93.

Greenberg G. The Rats of N.I.M.H. | The New Yorker. *New Yorker* 2013.

Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? *Behav Brain Sci* 2010;**33**:61–83; discussion 83-135.

Hull L, Petrides K V., Mandy W. The Female Autism Phenotype and Camouflaging: a Narrative Review. *Rev J Autism Dev Disord* 2020;**7**:306–17.

Hunt AD, Abed R, St John-Smith P. The Biopsychosocial Model Advanced by Evolutionary Theory. In: Abed R, St John-Smith P (eds.). *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge: Cambridge University Press, 2022.

Huxley J, Mayr E, Osmond H *et al.* Schizophrenia as a Genetic Morphism. *Nature* 1964;**204**:220–1.

Kapp SK, Pripas-Kapit S, Dekker M *et al. Autistic Community and the Neurodiversity Movement - Stories from the Frontline*. Kapp SK (ed.). Palgrave MacMillan, 2020.

Keller MC. Evolutionary Perspectives on Genetic and Environmental Risk Factors for Psychiatric Disorders. *Annu Rev Clin Psychol* 2018;**14**:471–93.

Keller MC, Howrigan DP, Simonson MA. Theory and Methods in Evolutionary Behavioral Genetics. *Evol Personal Individ Differ* 2011, DOI: 10.1093/acprof:oso/9780195372090.003.0010.

Keller MC, Miller G. Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behav Brain Sci* 2006;**29**:385–404.

Kincaid H. Doing without "Disorder" in the Study of Psychopathology. *Defin Disord* 2021:123–32.

Klein DF. Harmful dysfunction, disorder, disease, illness, and evolution. *J Abnorm Psychol* 1999;**108**:421–9.

Kotov R, Krueger RF, Watson D *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol* 2017;**126**:454–77.

Latzman RD, Krueger RF, DeYoung CG *et al.* Connecting quantitatively derived personality–psychopathology models and neuroscience. *Personal Neurosci* 2021;**4**:e4.

Lewis DMG, Al-Shawaf L, Conroy-Beam D *et al.* Evolutionary psychology: A how-to guide. *Am Psychol* 2017;**72**:353–73.

Lukaszewski AW, Lewis DMG, Durkee PK *et al.* An Adaptationist Framework for Personality Science. *Eur J Pers* 2020;**34**:1151–74.

Lukaszewski AW, Roney JR. The origins of extraversion: Joint effects of facultative calibration and genetic polymorphism. *Personal Soc Psychol Bull* 2011;**37**:409–21.

Majerus MEN. Industrial Melanism in the Peppered Moth, Biston betularia: An Excellent Teaching Example of Darwinian Evolution in Action. *Evol Educ Outreach* 2008;**2**:63–74.

Marek S, Tervo-Clemmens B, Calabro FJ *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nat 2022 6037902* 2022;**603**:654–60.

Margolis J. The concept of disease. *J Med Philos (United Kingdom)* 1976;**1**:238–55.

Martin JS, Jaeggi A V., Koski SE. The social evolution of individual differences: Future directions for a comparative science of personality in social behavior. *Neurosci Biobehav Rev* 2023;**144**:104980.

Mayr E. Cause and effect in biology. *Science (80-)* 1961;**134**:1501–6.

McCain K, Weslake B. Evolutionary Theory and the Epistemology of Science. *Hist Philos Theory Life Sci* 2013;**1**:101–19.

McCrae RR, Costa Jr. PT. The five-factor theory of personality. *Handb Personal Theory Res 3rd ed* 2008:159–81.

Montiglio P-O, Ferrari C, Réale D. Social niche specialization under constraints: personality, social interactions and environmental heterogeneity. *Philos Trans R Soc Lond B Biol Sci* 2013;**368**:20120343.

Mottron L. A radical change in our autism research strategy is needed: Back to prototypes. *Autism Res* 2021;**14**:2213–20.

Murphy D. Function and Dysfunction. *Defin Disord* 2021:251–66.

Nelson RH. A Critique of the Neurodiversity View. *J Appl Philos* 2020:japp.12470.

Nesse RM. Maladaptation and natural selection. *Q Rev Biol* 2005;**80**:62–70.

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

Nesse RM. Anxiety Disorders in Evolutionary Perspective. In: Abed R, St John-Smith P (eds.). *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge: Cambridge University Press, 2022, 101–16.

Nettle D. An evolutionary approach to the extraversion continuum. *Evol Hum Behav* 2005;**26**:363–73.

Nettle D. *Evolutionary Perspectives on the Five-Factor Model of Personality*., 2011.

Nour MM, Liu Y, Dolan RJ. Functional neuroimaging in psychiatry and the case for failing better. *Neuron* 2022;**110**:2524–44.

Plomin R. Commentary: Why are children in the same family so different? Non-shared environment three decades later. *Int J Epidemiol* 2011;**40**:582–92.

Polderman TJC, Benyamin B, De Leeuw CA *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015;**47**:702–9.

Power RA, Kyaga S, Uher R *et al.* Fecundity of Patients With Schizophrenia, Autism, Bipolar Disorder, Depression, Anorexia Nervosa, or Substance Abuse vs Their Unaffected Siblings. *JAMA Psychiatry* 2013;**70**:22.

Rogers A. Star Neuroscientist Tom Insel Leaves the Google-Spawned Verily for ... a Startup? | WIRED. *Wired* 2017.

Rogers WA, Walker MJ. The Line-drawing Problem in Disease Definition. *J Med Philos* 2017;**42**:405–23.

Schramme T. A qualified defence of a naturalist theory of health. *Med Heal Care Philos* 2007;**10**:11–7.

Schulze TG, Fangerau H, Propping P. From degeneration to genetic susceptibility, from eugenics to genethics, from Bezugsziffer to LOD score: the history of psychiatric genetics. *Int Rev Psychiatry* 2004;**16**:246–59.

Smith JM. *Evolution and the Theory of Games*. Cambridge University Press, 1976.

Smith SE. Is Evolutionary Psychology Possible? *Biol Theory* 2019;**1**:3.

Smoller JW, Andreassen OA, Edenberg HJ *et al.* Psychiatric genetics and the structure of psychopathology. *Mol Psychiatry 2018 243* 2018;**24**:409–20.

Sonleitner FJ. What Did Karl Popper Really Say About Evolution? *Creat Evol J* 1986;**6**:9–14.

Spitzer RL. Harmful dysfunction and the DSM definition of mental disorder. *J Abnorm Psychol* 1999;**108**:430–2.

Thornton T. Naturalism and Dysfunction. *Defin Disord* 2021:469–84.

Tinbergen N. On aims and methods of Ethology. *Z Tierpsychol* 1963;**20**:410–33.

Tooby J, Cosmides L. On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation. *J Pers* 1990;**58**:17–67.

Tooby J, Cosmides L. Conceptual Foundations of Evolutionary Psychology. *The Handbook of Evolutionary Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2015, 5–67.

Wakefield J. Can Causal Role Functions Yield Objective Judgments of Medical Dysfunction and Replace the Harmful Dysfunction Analysis's Evolutionary Component? Reply to Dominic Murphy. *Defin Disord* 2021:267–316.

Wakefield JC. The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values. *Am Psychol* 1992;**47**:373–88.

Wakefield JC. When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorder. *Dev Psychopathol* 1997;**9**:269–90.

Wakefield JC. Biological Function and Dysfunction: Conceptual Foundations of Evolutionary Psychopathology. *The Handbook of Evolutionary Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2015, 1–19.

Widiger T, Gore W, Crego C *et al.* Five-Factor Model and Personality Disorder. 2017, 449–78.

Williams GC. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press, 1966.

Zachar P, Kendler KS. The Philosophy of Nosology. *Annu Rev Clin Psychol* 2017;**13**:49–71.

Evolutionary Human Sciences (2022), **4**, e26, page 1 of 23
doi:10.1017/ehs.2022.23

REVIEW

Specialised minds: extending adaptive explanations of personality to the evolution of psychopathology

Adam D. Hunt* and Adrian V. Jaeggi

Institute of Evolutionary Medicine, University of Zürich, Switzerland
*Corresponding author. E-mail: adam.hunt@iem.uzh.ch

Abstract

Traditional evolutionary theory invoked natural and sexual selection to explain species- and sex-typical traits. However, some heritable inter-individual variability in behaviour and psychology – personality – is probably adaptive. Here we extend this insight to common psychopathological traits. Reviewing key findings from three background areas of importance – theoretical models, non-human personality and evolved human social dynamics – we propose that a combination of social niche specialisation, negative frequency-dependency, balancing selection and adaptive developmental plasticity should explain adaptation for individual differences in psychology – 'specialised minds' – explaining some variance in personality and psychopathology trait dimensions, which share various characteristics. We suggest that anthropological research of behavioural differences should be extended past broad demographic factors (age and sex) to include individual specialisations. As a first step towards grounding psychopathology in ancestral social structure, we propose a minimum plausible prevalence, given likely ancestral group sizes, for negatively frequency-dependent phenotypes to be maintained as specialised tails of adaptive distributions – below the calculated prevalence, specialisation is highly unlikely. For instance, chronic highly debilitating forms of autism or schizophrenia are too rare for such explanations, whereas attention-deficit-hyperactivity disorder and broad autism phenotypes are common enough to have existed in most hunter-gatherer bands, making adaptive explanations more plausible.

Keywords: Evolutionary psychiatry; neurodiversity; evolutionary psychology; hunter-gatherers; personality

Social media summary: Traits of personality and psychopathology could result from a shared evolutionary process of cognitive specialisation.

1. Introduction

Under traditional evolutionary theory, heritable phenotypic variation is expected to be positively selected until fixation or negatively selected until elimination (Buss & Hawley, 2011: ix). Following this, evolutionary psychology has primarily accounted for psychological mechanisms in terms of evolved universal cognitive architecture (Lukaszewski et al., 2020), with adaptive individual variation arising owing to plastic reactions to individual circumstances. This approach is visible in Darwin's *The Expression of the Emotions in Man and Animals*, framing emotional and behavioural responses as shared adaptive strategies. Personality, defined in behavioural ecology as repeatable individual differences in behaviours (Dochtermann et al., 2015) and defined more broadly in psychology as those characteristics of individuals that describe and account for consistent patterns in feeling, thinking and behaving (Gosling, 2001: 46), has thus often been portrayed as noise – neutral or maladaptive variation around an optimum (Tooby & Cosmides, 1990). This perspective is now changing, as growing theoretical and empirical research in humans and non-humans finds evidence for adaptive individual differences.

Human personality is moderately heritable; stable over time, context and culture; shows continuities with non-human personality differences; provides predictive power in forecasting behaviour; and has consequences on evolutionarily relevant factors such as survival, mating success, status, fecundity and parenting (Buss & Hawley, 2011: x). As heritable phenotypes active during reproductive years are visible to selection, an evolutionary explanation is required. Psychologists and psychiatrists also seek evolution-informed explanations for mental disorders (Del Giudice, 2018; Nesse, 2019). Personality and psychopathology overlap substantially – personality dimensions are related to specific disorders (Widiger et al., 2017), and many mental disorders are relatively common, heritable and affect fitness, requiring an evolutionary explanation (Keller & Miller, 2006). Just as personality differences are reflected by gradual neurobiological differences, common mental disorders do not generally have discrete pathological causes, and descriptivist approaches categorising mental disorders into distinct diagnoses ignore the reality that mental disorders are dimensional, sharing symptoms and alleles with the subclinical population (Geschwind & Flint, 2015; Zachar & Kendler, 2017). Even the term 'psychopathology' is misleading in implying identified disease – to follow past literature this paper uses it, but we support more cautious use of medicalising language in psychiatry, as seen in 'neurodiversity' perspectives (Kapp et al., 2020), which better align with the evidence and theory presented here.

This paper is concerned with evolutionary explanations for relatively *stable* individual differences in personality and psychopathology: 'traits' rather than 'states'. The question is why individuals differ in these traits, and why such differences could lead to diagnosable psychopathology. This excludes common emotions or mood disorders of anxiety and depression, although individual differences in traits often manifest in differential propensity towards certain states (e.g. neuroticism is related to commonly experiencing anxiety; psychopathy as experiencing less empathy, guilt or shame). Throughout this paper we shall examine evidence and theory that indicate that traits of personality and psychopathology are disciplinarily separated but biologically related phenomena, in need of related evolutionary explanation. Although wary of the mistake of overfitting multiple phenotypes into a single model, we believe that the idea that human evolution has selected for *specialised minds* (Tooby & Cosmides, 1996; Troisi, 2005) is most useful – like Dall et al. (2012), we believe that the concept of specialisation nicely encompasses the multiple processes leading to adaptive individual differences in behaviour. We shall examine evidence indicating that specialised minds should have adapted around social niche sizes and dynamics, and be sensitive to local ecology and culture (Smaldino et al., 2019). Social niche specialisation, social selection, adaptive developmental plasticity, negative frequency-dependency and temporally fluctuating selection should lead to complex specialising fixed and plastic effects manifesting via innate predispositions and developmental responses (Section 2). Specialisations involve trade-offs and functional and dysfunctional manifestations or aspects, often context-dependent, with stabilising costs and benefits, potentially maintained via inclusive fitness (Sections 4 and 5). Heterogeneity is the rule rather than the exception; recognition of pathological, neutral and adaptive forces will be required in full evolutionary explanations, especially of psychopathology (Section 5). Our main argument shall be that specialisation is the most useful adaptive process for theorists to consider, and that taking into account theoretical models (Section 2), non-human personality (Section 2), human evolved psychology and social dynamics (Section 3), and evolutionary approaches to personality (Section 4) will prove crucial in explaining psychopathological traits (Section 5). We finish by suggesting productive avenues for future research, especially by examining anthropological research with the proposed theoretical models in mind (Sections 6–8).

2. Evolutionary models of individual differences and examples in non-human animals

Adaptive individual differences in phenotype can arise via various mechanisms and selection pressures. Plasticity can adaptively match phenotypes to different environments (West-Eberhard, 2003), whilst maintenance of genetically based alternative strategies requires equal fitness over multiple generations (Smith, 1976). After briefly summarising relevant fixed genetic and plastic effects, we shall consider selective forces for phenotypic specialisations.

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Genetic effects on individual differences

At the genetic level, individual differences have four possible explanations (Keller et al., 2011):

(1) *Neutrality.* Genetic drift is maintaining differences; they have a negligible effect on reproductive success and are under neither positive nor negative selection.
(2) *Mutation-selection balance.* Individual differences are caused by harmful mutations which arise too frequently to be completely removed by negative selection.
(3) *Positive selection on recent mutations.* Responsible variants are under positive selection but have arisen too recently to have reached fixation.
(4) *Balancing selection.* The variant's fitness depends on conditions, with none being optimal for long enough to reach fixation (at the phenotypic level the process may be negatively frequency-dependent, or temporally or spatially dependent, or sexually antagonistic).

Of these, only under balancing selection is individual genetic variation itself selected to produce alternative strategies.

Plasticity and environmental effects on individual differences

Behavioural ecology starts from the assumption that behaviour is optimal in a given situation (Nettle et al., 2013). The presence of individual differences in behaviour which 'hang together' in stable personality types (also called 'behavioural syndromes') is therefore a fundamental problem (Dall et al., 2004; Sih et al., 2004). Why would some individuals be more aggressive than others, and why would an aggressive individual show differences in their interactions with conspecifics *and* foraging behaviour *and* exploratory behaviour, rather than acting facultatively?

One reason evolution can favour behavioural consistency is because social interactions become predictable, allowing social partners to learn and adaptively respond to personality types (Wolf et al., 2011). This especially occurs when individuals coordinate actions or have a mutual interest to avoid certain outcomes. Variation in turn favours responsiveness to consistent personalities, so a coevolutionary process can occur between responsiveness and consistency (McNamara et al., 2008). The ability to learn a skill and occupy a particular niche (see below) also encourages plasticity canalising over time. Potential costs of plasticity exist (Ellis & Del Giudice, 2019); cues received from the environment take time to sample and react to, may be misleading and difficult to reverse, and plasticity itself may be energetically costly. Adaptive plasticity is not always an optimal or possible strategy.

Relevantly to psychopathology, adaptive developmental plasticity and conditional adaptation leading to individual differences in phenotype have been studied and theorised about extensively in relation to stress (Ellis & Del Giudice, 2019). Assumptions that high-stress environments simply dysregulate biology to cause dysfunction are misplaced – adaptive reactions to stressful environments to maximise expected fitness are expected (e.g. Thayer *et al.*, 2018), sometimes involving trade-offs in organism longevity or health, as in certain applications of life history theory (see Section 5).

Key selective forces encouraging individual differences in behaviour

Individual differences in behaviour often occur in social contexts (Dall et al., 2012). 'Social selection' is a relevant process of natural selection involving the fitness effects of social behaviours and social competition (and subsumes sexual selection; Lyon & Montgomerie, 2012; Nesse, 2007; West-Eberhard, 1979). Organisms affect conspecifics, and the potential fitness benefits of relationships and group membership cause social selection for traits optimising access to these resources. This can lead to cooperation with unrelated individuals, apparent altruism and systems of reward and punishment (Barclay, 2013). Importantly, developing valuable specialised skills and attributes and being individually recognised for them is a key way for humans to ingratiate themselves with cooperators (Tooby & Cosmides, 1996; see Section 3).

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Social niche specialisation involves individuals in a population occupying different roles (Bergmüller & Taborsky, 2010; Montiglio et al., 2013) and is a key explanation for individual differences in social species. Non-human experimental evidence indicates that plastic social niche specialisation leads to various lasting behavioural changes: house mouse sibling sex ratio affects behavioural flexibility, great tit food rationing of siblings affects exploration behaviour, and rearing cichlids with or without dominants affects aggressive and submissive behaviours (Bergmüller & Taborsky, 2010). Genetic and environmental effects can interact in social niche specialisation; for example, an initial genetic predisposition and early life circumstances can lead to an individual occupying a certain niche, at which point they maximise fitness by plastically developing to maximise their efficacy in that niche (Smaldino et al., 2019). Multiple factors can affect niche selection, including state dependence (the individual's states predispose towards success in certain niches), frequency-dependence (negative if niches have limited resources or positive if niches' resources are maximised by multiple occupancy) and social awareness and eavesdropping (where conspecifics are aware of an individual's niche and this makes continually occupying the niche more profitable).

Intraspecific trait variation affects the ecological dynamics of communities in multiple relevant ways (Bolnick et al., 2011). Increasing environmental heterogeneity and population density increase personality diversity in small mammals including deer mice, southern red-backed voles and northern short-tailed shrews (Mortelliti & Brehm, 2020). An increased degree of variation makes a population more generalised via individual specialisation, reducing competition between members and increasing species niche width (e.g. in diet; Bolnick et al., 2003) as the community extracts more resources from the environment. Such species are more robust to extinctions. Social structures in humans can integrate personality differences into a 'pool of competence' (Wolf & Krause, 2014), as collectives of differentiated individuals increasing individual and group fitness through cooperation. Some researchers even suggest that the strength of this cooperative effect makes other evolutionary pathways of explaining diversity unnecessary (Nonacs & Kapheim, 2007) and that computational challenges of specialising played a significant role in selection pressures for *Homo* cognitive abilities (Hagen et al., 2021). The division of behaviour then encourages ongoing cooperation (Bergmüller et al., 2010; McNamara & Leimar, 2010). Paradigmatic examples are in eusocial insects, where division of labour and specialisation amongst workers increase colony efficiency, and emerges spontaneously in small groups of morphologically similar workers (Ulrich et al., 2018).

The balancing of selection effects on psychological traits can arise from temporally fluctuating selection (Taylor et al., 2014) when exogenous environmental (e.g. food availability, disease, conflict) and endogenous population factors (e.g. sex ratios; Del Giudice, 2012) change, affecting optimal behaviour, but never bringing a single phenotype or genotype to fixation. Temporally fluctuating selection has historically been questioned for its sufficiency in maintaining variation (but see Del Giudice, 2020: S2), with negative frequency-dependent selection the preferred explanatory force explaining adaptive individual variation via balancing selection (potentially mistakenly ignoring alternative or co-occurring forces; Brisson, 2018). Frequency-dependent selection occurs when the fitness of a genotype or phenotype depends on its frequency in the population. In positive frequency-dependency, higher frequency results in higher fitness. Under negative frequency-dependency, higher frequency results in lower fitness; this can lead to individual differences stabilising in frequency. Negative frequency-dependency is considered a key process maintaining individual variation in populations, exemplified in alternative mating strategies such as 'sneaky' males which imitate females (in isopods, sunfish, garter snakes and shorebirds) or providing exceptional care to young (cichlid fish; Shuster, 2010). Behavioural strategies include attempting to woo females with gifts or alternatively attempting to mate by force, or varying in tendency to seek multiple partners or defend a single partner. These alternative strategies may be plastic or inherited and fixed. Although negative frequency-dependency is often modelled in a simplified manner with few, discrete morphs (e.g. hawk and dove; male and female), it can also maintain multiple morphs on a continuous distribution underpinned by multiple polygenic factors, allowing species to fill niches as they become available (Slatkin, 1979). Note that forces of balancing selection are not mutually exclusive; negative

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

frequency-dependency and temporally fluctuating selection can act simultaneously in maintaining variation. Combining the effects of these forces is realistic – individual differences' adaptiveness depending both on fluctuating environmental factors and group composition where rare traits are advantageous.

Theoretical models of adaptation account for regular maladaptive outcomes, although maladaptation or dysfunction is only 'disorder' if deemed harmful by social consensus (Wakefield, 2015). Evolutionary medicine and psychiatry have identified various ultimate causes of disorders (Nesse, 2005, 2019). These may be broadly classified as genuine dysfunctions; functional mechanisms mismatched to modern environments and currently maladaptive; or generally adaptive traits with maladaptive or undesirable consequences (Del Giudice, 2018, p133). Medically relevant harm can have adaptive causes, and as noted by Troisi (2005), selection for alternative strategies could be a fundamental evolutionary cause of psychopathology. Adaptive dimensional strategies can lead to maladaptive extremes; costs of developmental plasticity could be maladaptive manifestation in a certain proportion of cases; assortative mating between similar individuals may lead to overexpression of an ancestrally adaptive trait; and mismatch may cause adaptations to manifest maladaptively in certain environments, or be defined as disorders by modern psychiatry. Heterogeneity both between and within disorder categories means that they require explanation via multiple models of adaptation and processes of maladaptation or dysfunction (Del Giudice, 2018).

Implications of theoretical and non-human literature in explaining human individual differences

The theoretical and non-human literature unequivocally shows that assumptions that all individual differences in behavioural traits are noise awaiting extinction or fixation are no longer tenable. However, complex dynamics maintain individual differences theoretically and in non-human species (e.g. Bergmüller and Taborsky, 2010; Shuster, 2010; Mouchet et al., 2021). Theoretical models explain different *aspects* of biological phenomena, so specific models may be simultaneously (but probably not equally) explanatory – models are not only potentially compatible but *necessarily* co-occur. Negative frequency-dependency often has effects in social niche specialisation (but see Brisson, 2018). Models of genetic and environmental effects are required simultaneously. Models of adaptive dimensional traits include functional and dysfunctional manifestations. The impossibility of full explanations with single models is a point against overreliance on single frameworks such as life history theory, which has been suggested as a unifying framework for evolutionary psychopathology (Del Giudice, 2018), but is under scrutiny for its somewhat idiosyncratic applications to humans (Frankenhuis & Nettle, 2020). Explaining individual differences in human psychological traits, adaptive and maladaptive, requires integrating a multiplicity of models.

3. Evolved human psychology and social dynamics

Theoretical models of evolved individual differences overwhelmingly depend on social dynamics. Here we consider various lines of evidence pointing towards the likelihood of humans evolving adaptive, relatively stable inter-individual differences in feeling, thinking and behaving. After noting aspects of human psychology which particularly encourage inter-individual specialisation, we consider anthropological evidence on hunter-gatherer skill niches and social structure, and group sizes and interaction dynamics (to later consider the implications of manifestation in ancestral groups; Section 7).

As reviewed above, specialisations are observed in non-human species and predicted by theoretical models under certain circumstances. Those circumstances are not just present, but heightened in human psychology. Human cognition shows high plasticity, allowing specialisation and canalisation which can't be easily reverted or copied; a tendency towards developing individual skills (Tooby & Cosmides, 1996); and behavioural flexibility and intelligence to select oneself into one's optimum available niche and out of inappropriate niches (Hooper et al., 2015). Humans seem to have evolved excellent capacities for noticing and acting upon other people's individual differences (Buss, 1996;

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Buss & Hawley, 2011), allowing intelligent partner selection, cheater detection and knowledge of others' social niches. Forces of social selection and negative frequency-dependent selection may be heightened by human cognitive ability to assess, remember, compare and share information about individuals (Penke, 2010), allowing social status to be ascribed to individuals who optimise group functioning, encouraging frequency-dependent division of labour and cognitive specialisation, widening the human ecological niche. These capacities provide perhaps unparalleled social selection for specialisation in humans. The fact that modern economies are run with division of labour and specialisation as a first principle, and that most modern specialisation is cognitive, with different personality profiles suiting different careers (Denissen et al., 2018; Wilmot & Ones, 2019), may be a formalised continuation of a long evolutionary history of social niche specialisation which contributed to human ecological success.

Human ancestral social organisation: dynamics, group sizes and niches

Accounting for the evolution of human specialised minds needs reference to ancestral social organisation, and thus evidence from anthropology, particularly of hunter-gatherer societies. Whilst the antiquity of this social organisation cannot be known with certainty, at least some elements of it arguably extend back some 2 million years to *Homo ergaster* (Willems & van Schaik, 2017).

Nomadic hunter-gatherer societies are generally egalitarian, sharing food and cooperating on daily tasks; however, social status does differ between individuals and has many fitness-relevant benefits (Von Rueden, 2014; Von Rueden et al., 2008; Von Rueden & Jaeggi, 2016), making status allocation a force of social selection. Reputations for competency and pro-sociality earn social support and attract cooperative partners, particularly in the contexts of collective labour (Macfarlan & Lyle, 2015). Indeed, cooperation earns social status to such an extent that social status allocation has been proposed as a key factor encouraging the evolution of human cooperative behaviour (Von Rueden et al., 2019).

Traits selected as social-niche specialisations and under negative frequency-dependency evolve in the context of niche size, species-typical group size and interaction dynamics with conspecifics. Hunter-gatherer group structure is typically composed of residential units ('bands'), interacting, migrating and exchanging spouses with other bands in metagroups ('tribes'). Individuals can choose to join foraging groups to optimise their relative efficiency (Smith, 1985) and assort into bands to optimise co-ordination and complementarity whilst maximising personal productivity (Hooper et al., 2015). Typical bands have a mean size of 28 individuals (Hill et al., 2011), consisting of about half a dozen adult male and female couples, children of various ages and perhaps a few post-reproductive individuals. Despite large variation in ecology and other aspects of behaviour, hunter-gatherer band sizes are remarkably similar (Kelly, 2013), probably reflecting consistent constraints such as the need to buffer shortfalls through food sharing (Winterhalder, 1986). Tribe size is more variable than band size, and bands may interact in complex networks of multiple languages that defy the notion of small-scale society (Bird et al., 2019). A relevant group size for our purpose is seen in periodic aggregations of bands who exchange marriage partners (to avoid inbreeding) and material goods, with a geometric mean of 165 individuals (95% confidence limit, 152–181; Hamilton et al., 2007; although the range extends from tens to several hundred individuals). These larger pools of social partners affects gene flow and the flow of information – foragers can pass information to and learn from potentially hundreds of individuals in a lifetime (Hill et al., 2014; Salali et al., 2016).

Traditional human societies provided a breadth of social niche possibilities, empowered by individual ability to migrate or settle into optimally suitable groups given their circumstances (Hooper et al., 2015). Ancestral social niches could have been explicit (e.g. shamanism) or implicit (e.g. personality types which 'fit in' with conspecifics). Explicit niches are most recognisable in division of labour and noticeable individual skills and talents (Sugiyama & Sugiyama, 2003) as avenues to gaining social status and cooperating (Jaeggi et al., 2016; Macfarlan & Lyle, 2015). Hunter-gatherer division of labour shows patterns between gender and age (Bird & Codding, 2015; Gurven & Hill, 2009). Most stereotypically, males more regularly hunt large prey, whilst females gather plant foods and engage in childcare.

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

However, this common demographic patterning hides potentially relevant diversity – for example, the Aché note the roles of finder, caller and killer whilst cooperatively hunting (Walker et al., 2002), which may suit different cognitive abilities, despite occurring within a gender-specific activity.

Sugiyama and Sugiyama (2003) extensively list possible niches for hunter-gatherer individuals to occupy. They may be in social roles (e.g. shaman or chief), specific crafts (e.g. pottery, basket weaving, boat building), or other performance or art (storytelling, singing, composition, dance). Social roles go beyond particular skills, for example in cooperators who aid in warfare, punishment or generosity. Knowledgeable individuals in areas of geography, spirituality, medicine and subsistence and technological skills are acknowledged (Lightner et al., 2021b). Individual roles are recognised and often admired: differential quality of manufactured tools is noted (Sugiyama & Sugiyama, 2003), as is efficacy in storytelling or teaching (Smith et al., 2017). Different abilities between specialists are tracked, and the best individuals are acknowledged, earning valuable social support (Schniter et al., 2018; Smith et al., 2017). Individual skills and personality traits are always judged in comparison with the immediate group; exceptionality exists in comparison with others (Tooby & Cosmides, 1996) and the best specialist in a niche can be disproportionately valued. For example, Singh and Henrich (2020) found that across two villages a single shaman performed 32% of ceremonies, the next best performed 14% and the remaining 54% were spread (unevenly) across 37 other individuals.

Different niches (e.g. tool use, shamanism and oratory skill) probably involve different dynamics, and the biological feasibility for optimal specialisation may differ between niches (e.g. shamanistic niche filling may optimise via adaptive developmental plasticity decided by early life experience, whilst tool-making niches may optimise via heritable fixed components). Precise dynamics are undoubtedly complex. For example, social niches in shamanism (Singh & Henrich, 2020) could show simple linear negative frequency-dependency (shamans steadily less successful as more common) or non-linear effects with mild negative effects on fitness for the first few shamans per tribe (as the niche fills up) and strong negative effects beyond that (once the niche is filled), or even positive frequency-dependency for the first few shamans (whose mutual presence reify their special status) and negative frequency-dependency beyond that. Negative frequency-dependency may affect niche occupiers differently, if the costs of increasing frequency are borne by certain individuals more than others (e.g. if the top shaman earns stable benefits despite niche filling, but lesser shamans compete for fixed benefits; Singh & Henrich, 2020). Even a particular trait's dynamics could differ between populations and over time, affecting the adaptive solutions and roles for fixed and plastic effects.

Anthropologists have not generally aimed to describe social niches and individual cognitive specialisation (perhaps excepting shamanism, e.g. Lightner et al., 2021a; Singh, 2017; Winkelman, 1990), concentrating more on age and sex differences. As modern studies consistently find that individual differences in personality and psychopathology are larger than group differences, this could be an area of important future research in traditional-living populations. Considering the non-human examples and theoretical requisites, the available psychological and anthropological evidence implies human evolutionary history would serve as fertile ground for the evolution of specialised minds, despite a lack of specific investigation into this question hitherto.

4. Adaptation and maladaptation in human personality

Evolutionary explanations of individual differences in psychology have largely concentrated on personality (Buss & Hawley, 2011). Mainstream personality psychology currently centres around factor-analytically derived traits representing clusters of co-occurring common personality descriptors. This method is inspired by the lexical hypothesis of personality, that socially important personality characteristics should be labelled and thereby appear in a language's vocabulary. Importantly, items included for factor analysis are chosen by psychologists, so do not include every stable psychological individual difference of evolutionary relevance. For example, variation in sociosexuality is absent, despite potentially being an adaptive individual difference under negative frequency-dependent selection or adaptive developmental plasticity (Bailey et al., 2000).

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

The most prominent model describing such personality factors is the five-factor 'Big Five' model (McCrae & Costa, 1997), identifying dimensions of openness, conscientiousness, extraversion, agreeableness and neuroticism. These traits are relatively stable over the life course (Graham et al., 2020) and affect various outcomes, including suitability to different careers (Denissen et al., 2018; Wilmot & Ones, 2019) and sexual behaviours and fertility (Allen & Robson, 2018; Allen & Walter, 2018). The Big Five has long been touted as a human universal (McCrae & Costa, 1997) and has been replicated in over 50 countries (Schmitt et al., 2007), although a more recent study in a small-scale subsistence society failed to replicate the same factor structure (Gurven et al., 2013) and large-scale surveys in 23 low- and middle-income countries have found the common Big Five Inventory lacks validity (Laajaj et al., 2019), probably because of methodological issues in translation, wording, interpretation and response biases. Sex differences in Big Five personality are generally small, although moderate differences are seen in agreeableness and neuroticism, in which females are higher (Del Giudice, 2015; Kajonius & Johnson, 2018). The sex differences have perhaps more frequently received evolutionary analysis (e.g. Kajonius & Johnson, 2018), even though group variance is smaller than individual variance, implying that adaptive explanations of individual variance are of high importance.

Personality factors seem to be caused by moderate genetic and environmental components, with an estimate of 40–50% heritability (Polderman et al., 2015; Vukasović & Bratko, 2015). Some of this variance is probably neutral or deleterious, with adaptive effects arising via both balancing selection on genes and developmental plasticity (Penke, 2010; Penke & Jokela, 2016). A prominent alternative explanation appeals to 'reactive heritability' (Lukaszewski & Roney, 2011) – other physical traits are heritable, such as body size, height and attractiveness, and personality could be a universal facultative programme reacting to these physical differences. Extraversion has been proposed as such a facultative strategy (Lukaszewski & von Rueden, 2015). In testing this hypothesis, Haysom et al. (2014) found that attractiveness accounted for only a small percentage of variance in extraversion, and other predicted reactive heritability effects were not found. Balancing selection thus probably partially explains personality's heritability (Buss & Hawley, 2011). Analysing the Big Five, Nettle (2006, 2011) identified possible costs and benefits causing balancing selection (Table 1). For example, high extraversion is associated with larger cooperative networks, increased mating success and positions of social status and leadership. However, it also increases the likelihood of experiencing antagonistic conflict, illnesses and injuries (Lukaszewski & von Rueden, 2015), implying variable benefits and costs in hunter-gatherer bands dependent on group composition and specific environmental circumstances.

Important in linking personality to psychopathology is the recognition that Big Five personality dimensions are associated with specific disorders and generally maladaptive outcomes, often at their tails. High neuroticism in particular is associated with anxiety, mood, eating, somatic symptom and substance use disorders, as well as social and specific phobias and various physical maladies (Widiger et al., 2018). A review of over 150 studies investigating how the Big Five relates to personality

Table 1. Plausible benefits and costs of the Big Five, derived from Nettle (2006, 2011)

Factor	Benefits	Costs
Openness	Artistic creativity; intellectual curiosity	Disorganised thought; tendencies to schizotypy and schizophrenia
Conscientiousness	Self-control; perfectionism; care in premeditated tasks; moral principle	Pathological levels of effort; rigidity; lack of spontaneity
Extraversion	Larger cooperative networks; high status; mating success	Increased risk of social conflict; illness and injury
Agreeableness	Empathy; helpfulness; harmonious interpersonal relationships	Vulnerable to cheaters; suboptimising of personal fitness
Neuroticism	Wariness; vigilance; useful anxiety	Debilitating anxiety; depression; physical health costs

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

disorders (PDs) (Widiger et al., 2017) reveals the extent to which personality and psychopathology are related, with both poles of each factor associated with maladaptive outcomes. High extraversion is associated with histrionic PD, high conscientiousness with obsessive–compulsive PD and high agreeableness with dependent PD. Low extraversion is associated with schizoid PD and avoidant PD, low conscientiousness with antisocial PD and laxness, negligence, disinhibition and irresponsibility, and low agreeableness with antisocial PD and the 'dark triad' traits of psychopathy, narcissism and Machiavellianism. It should be noted that such correlations are instrument dependent (Widiger et al., 2017). The fifth edition of the *Diagnostic and statistical manual of mental disorders* (DSM-5) in Section III, for 'emerging measures and models' (American Psychiatric Association, 2013: 728), included a dimensional model of five broad domains of negative affectivity, detachment, psychoticism, antagonism and disinhibition, explicitly recognising 'these five broad domains are maladaptive variants of the five domains of the extensively validated and replicated personality model known as the "Big Five"' (American Psychiatric Association, 2013: 773).

Despite success linking the Big Five with certain maladaptive outcomes and particular disorders, the totality of common psychopathological traits cannot merely be reduced to aspects of the Big Five. Even recent attempts to create entirely dimensional classifications of mental disorders using factor-analytic methods from personality psychology applied to mental disorder questionnaires fail to incorporate prominent conditions such as autism spectrum disorder (Kotov et al., 2017). The 157 diagnoses in the DSM-5 may overlap and correlate with other personality measures and disorders, but certainly require their own analysis.

5. Evolutionary accounts of psychopathology

Personality and psychopathology are described using different methodologies owing to psychiatry and clinical psychology's requirement for categorical diagnoses to inform treatment decisions (see Figure 1). Modern personality psychology is explicitly dimensional, not attempting to separately categorise particular personality types (although this is common outside of mainstream personality psychology, e.g. Myers–Briggs types). Despite the clinical necessity of discrete categorisation, psychopathologies also exist on dimensional spectra (Kotov et al., 2017; Widiger et al., 2017), with symptoms/traits visible in the general and subclinical population, especially among family members. Disciplinary separation obfuscates the many shared biological and epidemiological characteristics between

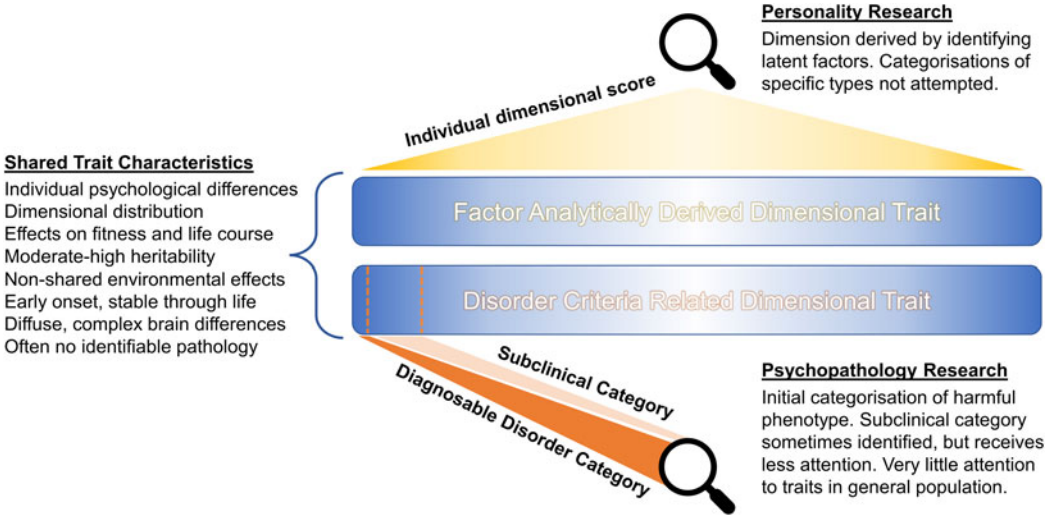

Figure 1. Traits of personality and psychopathology traits share basic characteristics but are studied differently.

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Table 2. Benefits, costs and models in a selection of evolutionary accounts for psychopathological traits

Condition	Proposed benefit	Proposed cost	Evolutionary model
Attention deficit hyperactivity disorder (ADHD) (Williams & Taylor, 2006)	Exploratory behaviour improves collective foraging, risk taking, sexual attractiveness	Physical, mental and social dangers of higher risk taking	Diversity-dependent group selection
Autism spectrum disorder (Crespi, 2016)	Species-wide positive selection on high intelligence, visuo-spatial strengths, attentional focus, deliberative decision making	Dysregulation and imbalance of intelligence causes autistic phenotypes of social, communication and behavioural difficulties	Species-wide positive selection with by-product
Bipolar disorder (Akiskal & Akiskal, 2005)	Subclinical traits beneficial (depressive sensitivity to suffering; creativity and high energy beneficial in leadership, exploration, territoriality and mating)	Extreme manifestations cause severe depression and maladaptive mania	Adaptive spectrum with costly extremes
Psychopathy (Mealey, 1995)	Antisocial strategy of emotionlessness, allowing cheating and optimally selfish behaviour, unconstrained by empathy or guilt	Possibility of detection and punishment	Negatively frequency-dependent and adaptive developmental plasticity
Schizophrenia (Polimeni & Reiss, 2002)	Religious delusions, hallucinations and social withdrawal interpreted as shamanic experiences; shamanic rituals increase group cohesion	Delirium, substance abuse, exacerbated psychotic symptoms in certain cultural contexts	Group selection for specialised roles

personality and psychopathology dimensions (Figure 1). Brain differences are diffuse and complex (Latzman et al., 2021) and heritability is often moderate to high (Polderman et al., 2015). Environmental components seem to be mainly non-shared (Plomin, 2011). Traits are (by definition) observable relatively early and stable throughout life, with effects on fitness (Allen & Robson, 2018; Power et al., 2013), and are common, both in clinically diagnosable and subclinical trait form, which is highly significant for evolutionary explanations (see Section 7). Most individual differences in personality or psychopathology have no discernible pathological cause. Beyond socially ordained descriptive differences and the presence of apparent harm, these traits are naturalistically similar.

Psychopathologies and their evolutionary accounts are too numerous to list here (see Del Giudice, 2018 for a comprehensive review). No single evolutionary explanation is applied between disorders (Table 2), and some accounts rely on less well substantiated evolutionary mechanisms (e.g. group selection) than are focussed on in the (human or non-human) personality literature. Possible costs and benefits of the traits are often emphasised. A point made near universally in the evolution-of-personality but less regularly in the psychopathology literature is of individual differences as *themselves* adaptive. Psychopathology accounts more often hypothesise that differential susceptibility results from simple dysfunction. Key points of five accounts of psychopathological traits are provided in Table 2. These five accounts are selected to show the variability of models used and the range of benefits and costs proposed, not to represent consensus or our preference.

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

For two traits, autism and psychopathy, leading accounts agree on costs and benefits, but differ in details (see below). Accounts of attention deficit hyperactivity disorder (ADHD), bipolar and schizophrenia vary widely between authors. ADHD has been proposed as hunting rather than farming behaviour; or part of a 'fighter' strategy; or a 'response-ready' strategy (Del Giudice, 2018: 266–268); or adaptive at the group level by improving collective foraging (Williams & Taylor, 2006). Hypotheses of bipolar disorder associate manic states with social dominance behaviours; hypomanic states with mastery and success in technical and artistic domains; as adapted to changing climates, with depression suitable for winter and mania for spring and summer (Del Giudice, 2018: 236–238); or manic states as enhancing creativity and energy in leadership, exploration, mating and territoriality, with depressive states enhancing sensitivity to suffering (Akiskal & Akiskal, 2005). Hypotheses of schizophrenia as resulting from specialised adaptation generally concentrate on subclinical schizotypal traits or less severe cases – suggestions are of a group-splitting function to enable smooth fissioning of hunter-gatherer tribes; or as sexually selected for enhanced creativity (Del Giudice, 2018: 219–222); or as causing shamanism which enhances group cohesion (Polimeni & Reiss, 2002). By-product hypotheses of schizophrenia implicate side effects of selection for lipid metabolism which enhanced creativity, religiosity and mentalising; or failure to establish hemispheric dominance for language; or various vulnerabilities in the 'social brain' (Del Giudice, 2018: 217–227).

Accounts of autism spectrum disorder (ASD) show more congruence (but see Del Giudice, 2018: 251–254). ASD is diagnosed by a broad array of differences and difficulties in socialising, communication and behaviour, with significant individual variability in presentation, hence the 'spectrum' (Lord et al., 2020). Some cases (often associated with intellectual disability) are caused by specific genetic or environmental disruptions of normal development, requiring no evolutionary explanation beyond recognising why such vulnerabilities exist (e.g. to mutations or prenatal toxins; Del Giudice, 2018). Other cases are unexplained by such disruptions and less frequently show intellectual disability, and autistic-like traits (which can coalesce in a 'broad autism phenotype'; Sasson et al., 2013) are more frequently seen in family members. These are the main target of evolutionary explanations. Three prominent accounts (Baron-Cohen, 2020; Crespi, 2016; Del Giudice, 2018) emphasise similar positive attributes observed in autistic and autistic-like traits. These include advantages in visual–spatial skills and abstract spatial reasoning, detail-oriented styles of cognition which boost 'systemising', enhanced pattern recognition, sensory acuity and perceptual discrimination and lower susceptibility to illusions (Del Giudice, 2018). Despite agreeing that ASD's evolutionary explanation revolves around these positive attributes, hypotheses differ in specifics. Baron-Cohen (2020) frames autistic traits as systemising, and autistics (without intellectual disability) as hyper-systemisers, occupying a social niche as toolmakers, inventors and experts in areas of their obsession. Sometimes this systemising overexpresses, leading to more disability. Del Giudice (2018) argues that autistic-like traits delay reproduction and increase parental investment as a slow life history strategy, whilst accumulating embodied capital through specialised skills learning, leading to a skilled/provisioning strategy. Maladaptive cases arise from overexpression of potentially adaptive traits, exacerbated by assortative mating, mutations and environmental insults. Baron-Cohen and Del Giudice therefore agree on autistic traits as specialisations, but differ on specifics of function. Crespi (2016) provides an alternative account, of ASD as a dysregulation of intelligence in components associated with strengths (e.g. visual–spatial abilities). He hypothesises that strong recent positive selection for human intelligence has led to cases of ASD as maladaptive by-products in some individuals, without claiming that the individual differences in autistic traits are themselves adaptive (although previous work (Crespi & Badcock, 2008) mentions specialised cognitive strengths and impairments in autistic and psychotic-spectrum individuals).

Psychopathy is characterised by callous and unemotional traits, impulsiveness, manipulativeness and remorselessness. Modern psychopaths can be career-focussed and manipulate their way up career hierarchies (Chiaburu et al., 2013). Aspects of psychopathy seem suitable for politics (Lilienfeld et al., 2012). Psychopathy has received substantial attention as an adaptive strategy, partly because cheating–cooperating strategies are canonical in game theory (e.g. prisoner's dilemma), inspiring early evolutionary accounts (Harpending & Sobus, 1987). Mealey (1995) suggested that psychopathy is

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

maintained by negative frequency-dependency and adaptive developmental plasticity, with a low frequency of psychopaths (to Mealey, 'primary sociopaths' utilising a genetic strategy, 'secondary sociopaths' utilising a reactive plastic strategy) taking a 'cheater niche' in cooperating human groups, accurately assessing the costs and benefits of cheating to cheat as much as is personally profitable. The prevalence of full psychopathy is about 1–2% for males and 0.3–0.7% for females, with a wider psychopathic personality in about 10–12% of the population (Colins et al., 2016). Gervais et al. (2013) suggested that clinical psychopathy is a strategy of unconditional defection, whilst subclinical psychopathic traits promote strategic conditional defection, broadening the adaptive niche of psychopathy in human societies. Other authors have framed psychopathy as a fast life history strategy, noting similar benefits and costs for psychopathic traits (Barr & Quinsey, 2004; Krupp et al., 2013).

These accounts of ASD and psychopathy potentially present examples of theoretically predicted cooperating and cheating evolutionary strategies, respectively. However, in general, accounts of psychopathology are more disconnected from background theory than literature on the evolution of personality (references to group selection specifically imply this oversight), emphasising specific disorder characteristics and hypothetical costs and benefits rather than realistic evolutionary models. This may partly explain the varied accounts of schizophrenia, ADHD and bipolar disorder. We suggest that progress will come by recognising non-human and theoretical models of adaptive individual differences (Section 2), better information regarding evolutionary human social dynamics (Section 3) and clearer acknowledgement of the similarity and relationship between personality and psychopathology trait dimensions (Figure 1), implying a need for shared scientific explanations despite disciplinary separation.

Previous authors have mistakenly inferred that identifiable instances of pathological processes affecting psychopathology (Keller & Miller, 2006) and personality (Verweij et al., 2012) imply that the *entirety* of variance is pathological. The fact that traumatic brain injury can cause aggressive personalities does not imply all aggressive personality is pathological – we cannot infer causation of heterogenous trait categories from causation of single instances. Predictions that mutation-selection balance would explain all heritable variance in disorders have not borne out (Keller, 2018; Keller & Miller, 2006). Pure by-product or pathology explanations of heterogenous phenotype dimensions such as ASD are bound to fail to explain the full clinical and subclinical variance. Similarly, adaptive specialisation clearly will not explain cases associated with *de novo* mutations. Every psychopathology dimension probably contains pathological variance (probably the most debilitating cases) and adapted variance (probably subclinical cases). Better recognition of this heterogeneity and of the plausible evolutionary models of personality adaptation offers promise for guiding evolutionary explanations of psychopathological traits.

6. Minimum adaptive prevalence

To offer some novel contribution beyond the review and commentary above, we now briefly sketch an example of practical synthesis across these areas. This specifically talks to a longstanding problem in the philosophy of medicine, involving the validity of attributing disorder to statistical outliers (Rogers & Walker, 2017). Evolution- and anthropology-informed work could certify low-frequency traits as disorders – or at least, not adaptations – with naturalistic rather than arbitrary justification of line-drawing at a particular prevalence.

We introduce this as a conceptual prompt for future research rather than an immediately applicable model. We encourage formal modelling and recognise that specific empirical details of the trait being analysed will need accounting for, including onset and duration, dimensionality, genetic architecture, environmental effects and probably selection dynamics (as noted in various plausible shamanism niche dynamics in Section 3). Nevertheless, we believe that framing this fundamental concept offers hope of progress in assessing psychopathology evolutionarily, and that further work can even be extrapolated to understanding specialisation more generally, especially in illuminating function, rather than simply providing a lower bound frequency for specialised adaptations.

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Negative frequency-dependency is a key force of balancing selection, regularly included in models of stable individual differences, including niche specialisation. When negatively frequency-dependent phenotypes reduce in frequency, negative selection effects approach 0. Upon reaching 1 individual with a phenotype, negative selection reaches 0 (to apply to dimensional models assume that different positions on the dimension are different phenotypes). Negative frequency-dependence can thus maintain, at an absolute minimum, one individual with a specialised phenotype (or capacity to develop a phenotype) per relevant group of interacting conspecifics. This allows dimensional tails to be deemed almost certain non-adaptations at a non-arbitrary point – when a phenotype is too rare to exist in every interacting group. So, where negative frequency-dependency is maintaining specialisation, we can derive a minimum adaptive prevalence (MAP) for a specialised trait using evolutionarily relevant group sizes. Where group size is represented as G, the simplest estimate of MAP is calculated with the equation:

$$\mathrm{MAP} = 1/G$$

Thus, under simplifying assumptions, in species interacting in groups of 30–40, the MAP of a negatively frequency-dependent phenotype is 1/35, approximately 3%. If 1% of individuals in that species display a phenotype, we can infer that it is not a specialised trait maintained by negative frequency-dependent selection. Note the assumption of equal environmental effects; if the specialised strategy is as an adaptive plastic developmental response (e.g. only becoming a psychopath if raised in an abusive environment) or also under temporally fluctuating selection, then the observed prevalence can be lower than the MAP. Also note the importance of accounting for trait-specific features and possible functions – if the trait is sexually dimorphic (perhaps sexually selected) or shows a delayed age of onset, effective group size is reduced, because negative frequency-dependency occurs via phenotype interaction and competition – the evolutionarily relevant group is then smaller than the whole group size. Beyond 'once per group' specialisations, 'once per males/females in group' or 'once per age group' specialisations are plausible, but need the group size to be adjusted to the relevant population to calculate their MAP (Table 3). Phenotype competition often occurs within demographics – rarely would a human 5 year old's phenotype have negative frequency-dependent effects on a 25 year old's similar phenotype – so this is an important consideration in discerning (and excluding) potential specialisations.

7. Specific prevalence of human specialised minds

Calculating the MAP for human traits requires knowledge of human social group size and dynamics. Band size is around 28 individuals, a relevant figure in assessing specialisation function, but does not provide the MAP, because bands are interconnected. Negative frequency-dependency relies upon interacting phenotypes reducing each other's fitness. The relevant social group size in estimating human MAP must interact frequently enough that individual differences interrupt each other and compete, and be sensitive enough to phenotype frequency change that one individual per group could display an adaptive phenotype, but in two individuals negative selection would reduce the phenotype's frequency back down to one individual per group. This group size will be much lower than lifetime interaction partners (which can be several hundreds; Hill et al., 2014), but could be larger than band size. Anthropologists have not specifically tackled this question, but as a placeholder for purposes of example, a reasonable estimate can be derived from periodic aggregations of bands totalling around 165 individuals (Hamilton et al., 2007). Periodic aggregations interact frequently enough, and phenotypes which do not periodically aggregate seem unlikely to negatively impact each other via their frequency. Although larger periodic aggregations have been recorded, the chance of relevant interaction in such large groups drops, and the relevant group size is that which caused selection over many generations, so the aggregate mean seems more relevant than higher or lower bounds of band aggregations (although this depends on specific selection effects and deserves separate extended inquiry). Taking

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Table 3. Estimated individuals per varying group, and resulting frequency of a particular trait given its prevalence. Zeros indicate prevalence falling below the MAP, indicating that the trait cannot be the result of a negatively frequency-dependent specialisation for that group

		Prevalence; percentage/individuals per group						
		50%	20%	10%	5%	1%	0.5%	0.2%
Group type (size) (Hamilton et al., 2007; Hill et al., 2011)	Aggregated bands (165)	82–83	33	16–17	8	1–2	0–1	0
	Males/females per aggregated band (83)	42	17	8	4	0–1	0	0
	Adult male/females per aggregated band (42)	21	8	4	2	0–1	0	0
	Band (28)	14	5–6	2–3	1	0	0	0
	Males/females per band (8–12)	4–6	2	1	0	0	0	0
	Adult male/females per band (4–6)	2–3	1	0–1	0	0	0	0

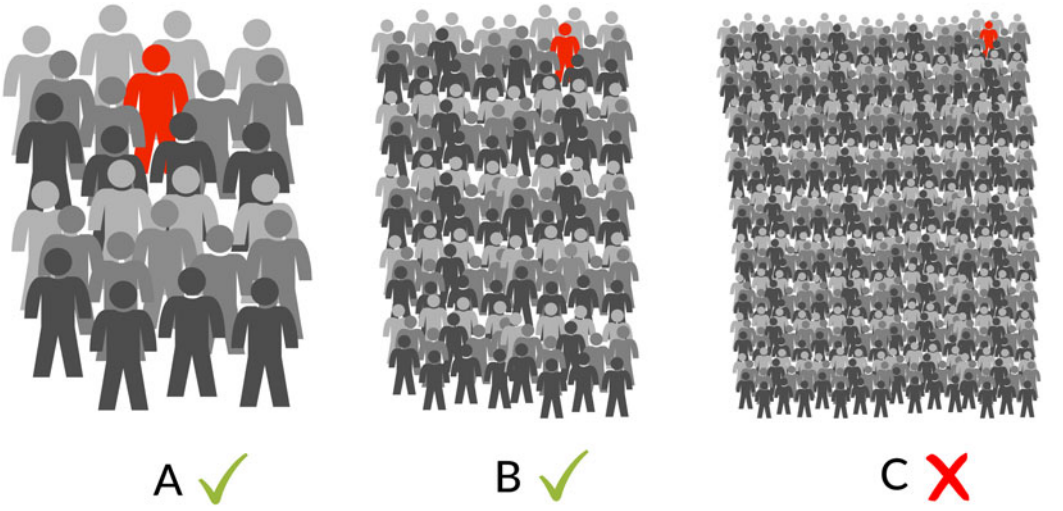

Figure 2. Prevalence per group size in relation to MAP. (A) is approximately band sized, 26 individuals. (B) is approximately band-aggregation sized, 130 individuals. (C) is more than twice band-aggregation size, 416 individuals. Phenotypes appearing once in (A) and (B) could plausibly be adaptive and maintained by negative frequency-dependency; phenotypes appearing once in (C) could not.

shamanism as prototypical role specialisation filling a human social niche, feasibly one shaman per 165 individuals experiences optimum benefits, additional shamans experiencing negatively frequency-dependent declining benefits as the niche fills, and groups much larger than 165 probably cannot have their demand for shamanistic services met by one shaman (Singh, 2017; Singh & Henrich, 2020).

The MAP of a group of 165 is approximately 0.6% so (in this simplified conceptual model) traits rarer than 0.6% prevalence are highly unlikely to be adaptations (Figure 2). *Note this is an absolute minimum*, and traits at 0.6% prevalence are only eligible for a very specific adaptive explanation – a specialisation functional for only one individual per group of aggregating bands – presumably an unusual occurrence. Sex- or age-group specific specialisations (e.g. 'once per adolescent males in aggregating bands') have lower group sizes and thus a higher MAP. Because of regularity of interaction, it is perhaps more plausible that 'one per group' specialisations were maintained by negative frequency-dependency in adults within bands (approximately 8–12 individuals), yet exist in exaggerated maladaptive forms in every aggregation of bands (possible if the inclusive fitness cost/benefit is stable). If so, maladaptive traits related to adaptations would exist above the MAP derived from band agglomeration size – this should encourage caution against using the MAP as a simple cut-off prevalence above which adaptation is expected and below which dysfunction is evident.

This analysis speaks to the paradox of common, harmful, heritable mental disorders (Keller & Miller, 2006), placing the 'common' factor in the appropriate context of ancestral human group sizes. Most common psychopathological traits (including all those mentioned in Section 5) have a lifetime prevalence of around or over 1%. Thus, every human ancestor probably interacted with at least one individual with each condition (assuming fairly constant cross-cultural rates, which is plausible, although difficult to test; Steel et al., 2014). Certain conditions and subclinical manifestations such as ADHD and broad autism phenotype have a prevalence of approximately 7% (Figure 3) – so predictably present once or twice in every band. Crucial, though, in identifying non-adaptive traits is recognising heterogeneity within psychopathology categories and specific trait characteristics. For example, some estimates of ASD base prevalence are 0.97% with approximately 40% having intellectual disability (Fombonne et al., 2021), in which case intellectually disabled cases do not meet the MAP. Lifetime schizophrenia prevalence is approximately 0.7% (although rates differ substantially between populations; McGrath et al., 2008), but schizophrenia onset is in late adolescence. Out of 165 individuals,

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Figure 3. The expected prevalence of five psychopathological traits (Table 2) and their subclinical spectra amongst a band-aggregation sized population of 165 adults and children. Following diagnostic norms and clinical appearance, schizophrenia, bipolar and psychopathy are displayed as adult onset (assumed around 18 years of age); ADHD and autism are active in children. Some accounting for gender differences and comorbidity between ADHD/autism spectra (Antshel et al., 2014) and bipolar/psychosis spectra (Kotov et al., 2020) has been made. Psychopathy and psychopathic personality, Colins et al. (2016); schizophrenia, McGrath et al. (2008) and psychosis spectrum, Guloksuz & Van Os (2018); bipolar disorder and bipolar spectrum disorders, Merikangas et al. (2011); autism spectrum disorder, Lord et al. (2020) and broad autism phenotype, Sasson et al. (2013); and attention-deficit hyperactivity disorder, Polanczyk et al. (2014).

approximately one develops schizophrenia in their life, but out of 165 hunter-gatherers, roughly half are children – one schizophrenic per two band agglomerations – so this falls below the MAP (this presents a further problem for Polimeni and Reiss's (2002) hypothesis of schizophrenia as a group-selected adaption causing shamanism). Heterogeneous genetic etiologies also require recognition. A significant proportion of cases of autism and schizophrenia are caused by rare or *de novo* variants (De La Torre-Ubieta et al., 2016; Legge et al., 2021; Singh et al., 2022), excluding adaptive explanations. Epidemiological estimates of disorder prevalence can contain potentially adaptive and certifiably dysfunctional genetic subtypes – removing certifiably dysfunctional types identifies the prevalence needed to meet the MAP. Combining these winnowing factors, the 1% lifetime prevalence of severe disorders can quickly drop below the MAP. This is compatible with hypotheses of adaptive specialisation leading to personality and psychopathology spectra, with costly extremes diagnosable as disorders.

8. Limitations and future directions

It is, in our view, essentially untenable that specialised minds would be predicted by theory, visible in non-humans, perfectly suited to human social dynamics and evolution, and yet not cause *any* of the variance in personality or psychopathology. Future questioning should not be whether human minds specialised at all, but uncovering the precise details and extent of that specialisation. The clearest areas for development are in fundamental theory – developing formal models accounting for the complexity of specialisation – and better empirical understanding of relevant human psychology and ancestral social dynamics – particularly in anthropology.

Calls for adaptationist explanations of psychological individual differences are plentiful (Buss & Hawley, 2011), but heterogenous and complex causation makes specific hypotheses hard to prove

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

or assess using a rigorous scientific method (although see Hunt, in preparation). We suggest that a first crucial step in untangling heterogeneity, especially in assessing psychopathology, will follow approaches such as Del Giudice's (2018), recognising subtypes eligible for functional and dysfunctional explanations before attempting to assess adaptation. This will probably broadly redraw the line between psychiatric and neurological disorders (Anttila et al., 2018).

Specialisation should be selected for specific social niches and group dynamics – probably dramatically changed since relevant periods of human evolution, and thus less amenable to the usual methods in evolutionary psychology seeking to describe and test complex design (Lukaszewski et al., 2020; Tooby & Cosmides, 1996). Human universals governing sexual attraction, status or anger may be straightforwardly experimentally testable in developed societies, but adaptive functions of individual differences in personality may manifest unusually owing to environmental factors, and certain culturally specific group dynamics (e.g. surrounding shamanism) have probably completely changed. Any fitness effects measured in developed societies are apt to be misleading, as evidenced by dramatic drops in fertility observed in recent decades; only if fertility has dropped equally irrespective of personality and psychopathology since relevant evolutionary time periods would such measures be useful. The gap between ancestral and modern environments also forces close consideration of heritability estimates of these traits, as measured heritability depends on socio-ecological circumstances (Uchiyama et al., 2021) and indirect genetic effects from social partners (Martin & Jaeggi, 2021). Since gene–environment interactions are important in personality and psychopathology phenotypic expression, their current manifestation, population variation and heritability may differ substantially from relevant periods of evolutionary history.

For these reasons, particular importance lies in studying individual psychological differences in hunter-gatherer and other non-industrialised societies, ideally measuring fitness consequences, but at minimum assessing how (and whether) such traits manifest and function in these groups. The fact that anthropologists have concentrated on measuring and analysing overall trends, such as sex or age differences – negligible in comparison with individual differences in modern society – implies that a wealth of relevant information awaits discovery. Inter-individual differences may be pivotal to the division of labour and hunter-gatherer cooperative dynamics. An example of low-hanging fruit might be re-analyses of research on hunter-gatherer behaviour, examining variance. Models of human individual differences in behaviour as noise awaiting negative selection predict trait clustering around the optimum (probably the mean) and negative effects moving away from it. Specialisation predicts a spectrum of behavioural variation related to fixed and plastic optimisation within niches to achieve optimum inclusive fitness – useful differences, not simply noise. The fact that personality and psychopathology play such critical roles in modern life paths may not be an artefact of modernity, but a cross-culturally verifiable result of evolutionary history.

Acknowledgements. The authors thank Louise Barrett for the invitation to submit to this special issue and Ed Hagen and Marco Del Giudice for very helpful comments on the manuscript. They also thank Jordan Martin, Camila Scaff, Erik Ringen, Gabriel Saffa and Giuseppe Pierpaolo Merola for helpful discussion on early drafts of the manuscript. Adam Hunt thanks Randolph Nesse, Paul St-John Smith and Riadh Abed for stimulating relevant discussion.

Author contributions.
AH conceived of the paper and led the writing; AJ gave ongoing feedback and contributed to the writing.

Financial support. AH was supported by University of Zurich Forschungskredit grant no. FK-21-062. AJ and AH were supported by the Pierre Mercier Foundation for Science.

Conflicts of interest. The authors declare no conflicts of interest.

References

Akiskal, K. K., & Akiskal, H. S. (2005). The theoretical underpinnings of affective temperaments: Implications for evolutionary foundations of bipolar disorder and human nature. *Journal of Affective Disorders*, *85*(1–2), 231–239. https://doi.org/10.1016/j.jad.2004.08.002

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Allen, M. S., & Robson, D. A. (2018). A 10-year prospective study of personality and reproductive success: Testing the mediating role of healthy living. *Psychology and Health*, *33*(11), 1379–1395. https://doi.org/10.1080/08870446.2018.1498499

Allen, M. S., & Walter, E. E. (2018). Linking big five personality traits to sexuality and sexual health: A meta-analytic review. *Psychological Bulletin*, *144*(10), 1081–1110. https://doi.org/10.1037/BUL0000157

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.

Antshel, K. M., Zhang-James, Y., & Faraone, S. V. (2014). The comorbidity of ADHD and autism spectrum disorder. *Expert Review of Neurotherapeutics*, *13*(10), 1117–1128. https://doi.org/10.1586/14737175.2013.840417

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., … Neale, B. M. (2018). Analysis of shared heritability in common disorders of the brain. *Science (New York)*, *360*(6395), 8757. https://doi.org/10.1126/SCIENCE.AAP8757

Bailey, J. M., Kirk, K. M., Zhu, G., Dunne, M. P., & Martin, N. G. (2000). Do individual differences in sociosexuality represent genetic or environmentally contingent strategies? Evidence from the Australian twin registry. *Journal of Personality and Social Psychology*, *78*(3), 537–545). https://doi.org/10.1037/0022-3514.78.3.537

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior*, *34*(3), 164–175. https://doi.org/10.1016/J.EVOLHUMBEHAV.2013.02.002

Baron-Cohen, S. (2020). *The pattern seekers: How autism drives human invention*. Basic Books.

Barr, K. N., & Quinsey, V. L. (2004). Is psychopathy pathology or a life strategy? Implications for social policy. In *Evolutionary psychology, public policy and personal decisions* (pp. 293–317). Lawrence Erlbaum Associates.

Bergmüller, R., Schürch, R., & Hamilton, I. M. (2010). Evolutionary causes and consequences of consistent individual variation in cooperative behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2751–2764. https://doi.org/10.1098/rstb.2010.0124

Bergmüller, R., & Taborsky, M. (2010). Animal personality due to social niche specialisation. *Trends in Ecology and Evolution*, *25*(9), 504–511. https://doi.org/10.1016/j.tree.2010.06.012

Bird, D. W., Bird, R. B., Codding, B. F., & Zeanah, D. W. (2019). Variability in the organization and size of hunter–gatherer groups: Foragers do not live in small-scale societies. *Journal of Human Evolution*, *131*, 96–108. https://doi.org/10.1016/j.jhevol.2019.03.005

Bird, R. B., & Codding, B. F. (2015). The sexual division of labor. In *Emerging trends in the social and behavioral sciences* (pp. 1–16). Wiley. https://doi.org/10.1002/9781118900772.etrds0300

Bolnick, D. I., Amarasekare, P., Araújo, M. S., Bürger, R., Levine, J. M., Novak, M., … Vasseur, D. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, *26*(4), 183. https://doi.org/10.1016/J.TREE.2011.01.009

Bolnick, D., Svanbäck, R., Fordyce, J.A., Yang, L.H., Davis, J.M., Hulsey, C.D., & Forister, M.L. (2003). The ecology of individuals: Incidence and implications of individual specialisation. *The American Naturalist*, *161*(1), 1. https://doi.org/10.2307/3078879

Brisson, D. (2018). Negative frequency-dependent selection is frequently confounding. *Frontiers in Ecology and Evolution*, *6*, 10. https://doi.org/10.3389/FEVO.2018.00010/BIBTEX

Buss, D. M. (1996). Social adaptation and five major factors of personality. In J. S. Wiggins (ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 18–207). Guilford. https://labs.la.utexas.edu/buss/files/2015/09/social_adaptation-1996.pdf

Buss, D. M., & Hawley, P. H. (2011). *The evolution of personality and individual differences*. Oxford University Press.

Chiaburu, D. S., Muñoz, G. J., & Gardner, R. G. (2013). How to spot a careerist early on: Psychopathy and exchange ideology as predictors of careerism. *Journal of Business Ethics*, *118*(3), 473–486. https://doi.org/10.1007/s10551-012-1599-5

Colins, O. F., Fanti, K. A., Salekin, R. T., & Andershed, H. (2016). Psychopathic personality in the general population: Differences and similarities across gender. *Journal of Personality Disorders*, *31*(1), 49–74. https://doi.org/10.1521/pedi_2016_30_237

Crespi, B. (2016). Autism as a disorder of high intelligence. *Frontiers in Neuroscience*, *10*, 300. https://doi.org/10.3389/fnins.2016.00300

Crespi, B., & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences*, *31*(3), 241–261; discussion 261–320. https://doi.org/10.1017/S0140525X08004214

Dall, S. R. X., Bell, A. M., Bolnick, D. I., & Ratnieks, F. L. W. (2012). An evolutionary ecology of individual differences. *Ecology Letters*, *15*(10), 1189–1198. https://doi.org/10.1111/j.1461-0248.2012.01846.x

Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. *Ecology Letters*, *7*(8), 734–739. https://doi.org/10.1111/j.1461-0248.2004.00618.x

De La Torre-Ubieta, L., Won, H., Stein, J. L., & Geschwind, D. H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nature Medicine*, *22*(4), 345–361. https://doi.org/10.1038/nm.4071

Del Giudice, M. (2012). Sex ratio dynamics and fluctuating selection on personality. *Journal of Theoretical Biology*, *297*, 48–60. https://doi.org/10.1016/J.JTBI.2011.12.004

Del Giudice, M. (2015). Gender differences in personality and social behavior. In *International encyclopedia of the social & behavioral sciences* (pp. 750–756). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.25100-3

Del Giudice, M. (2018). *Evolutionary psychopathology: A unified approach*. Oxford University Press. https://doi.org/10.1093/med-psych/9780190246846.001.0001

Del Giudice, M. (2020). Rethinking the fast–slow continuum of individual differences. *Evolution and Human Behavior*, *41*(6), 536–549. https://doi.org/10.1016/J.EVOLHUMBEHAV.2020.05.004

Denissen, J. J. A., Bleidorn, W., Hennecke, M., Luhmann, M., Orth, U., Specht, J., & Zimmermann, J. (2018). Uncovering the power of personality to shape income. *Psychological Science*, *29*(1), 3–13. https://doi.org/10.1177/0956797617724435

Dochtermann, N. A., Schwab, T., & Sih, A. (2015). The contribution of additive genetic variation to personality variation: Heritability of personality. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798). https://doi.org/10.1098/RSPB.2014.2201

Ellis, B. J., & Del Giudice, M. (2019). Developmental adaptation to stress: An evolutionary perspective. *Annual Review of Psychology*, *70*, 111–139. https://doi.org/10.1146/ANNUREV-PSYCH-122216-011732

Fombonne, E., Macfarlane, H., & Salem, A. C. (2021). Epidemiological surveys of ASD: Advances and remaining challenges. *Journal of Autism and Developmental Disorders*, *51*. https://doi.org/10.1007/s10803-021-05005-9

Frankenhuis, W. E., & Nettle, D. (2020). Current debates in human life history research. *Evolution and Human Behavior*, *41*(6), 469–473. https://doi.org/10.1016/J.EVOLHUMBEHAV.2020.09.005

Gervais, M. M., Kline, M., Ludmer, M., George, R., & Manson, J. H. (2013). The strategy of psychopathy: Primary psychopathic traits predict defection on low-value relationships. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1757), 20122773–20122773. https://doi.org/10.1098/rspb.2012.2773

Geschwind, D. H., & Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science* (*New York*), *349*(6255), 1489–1494. https://doi.org/10.1126/science.aaa8954

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*, *127*(1), 45–86. https://doi.org/10.1037/0033-2909.127.1.45

Graham, E. K., Weston, S. J., Gerstorf, D., Yoneda, T. B., Booth, T., Beam, C. R., … Mroczek, D. K. (2020). Trajectories of big five personality traits: A coordinated analysis of 16 longitudinal samples. *European Journal of Personality*, *34*(3), 301–321. https://doi.org/10.1002/per.2259

Guloksuz, S., & Van Os, J. (2018). The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. *Psychological Medicine*, *48*(2), 229–244. https://doi.org/10.1017/S0033291717001775

Gurven, M., & Hill, K. R. (2009). Why do men hunt? A reevaluation of 'Man the Hunter' and the sexual division of labor. *Current Anthropology*, *50*(1), 51–74. https://doi.org/10.1086/595620

Gurven, M., Rueden, C. von, Massenkoff, M., Kaplan, H., & Vie, M. L. (2013). How universal is the big five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, *104*(2), 354–370. https://doi.org/10.1037/a0030841

Hagen, E., Blackwell, A., Lightner, A., & Sullivan, R. (2021). Homo medicus: The transition to meat eating, increased pathogen pressure, and the constitutive and inducible use of pharmacological plants in *Homo*. *OSF Preprints*. https://doi.org/10.31219/OSF.IO/76BKA

Hamilton, M. J., Milne, B. T., Walker, R. S., Burger, O., & Brown, J. H. (2007). The complex structure of hunter–gatherer social networks. *Proceedings. Biological Sciences*, *274*(1622), 2195–2202. https://doi.org/10.1098/rspb.2007.0564

Harpending, H. C., & Sobus, J. (1987). Sociopathy as an adaptation. *Ethology and Sociobiology*, *8*, 63–72. https://doi.org/10.1016/0162-3095(87)90019-7

Haysom, H. J., Mitchem, D. G., Lee, A. J., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2014). A test of the facultative calibration/reactive heritability model of extraversion. *Evolution and Human Behavior*, *36*(5), 414–419. https://doi.org/10.1016/j.evolhumbehav.2015.03.002

Hill, K. R., Walker, R. S., Bozicevic, M., Eder, J., Headland, T., Hewlett, B., … Wood, B. (2011). Co-residence patterns in hunter–gatherer societies show unique human social structure. *Science*, *331*(6022), 1286–1289. https://doi.org/10.1126/science.1199071

Hill, K. R., Wood, B. M., Baggio, J., Hurtado, A. M., & Boyd, R. T. (2014). Hunter–gatherer inter-band interaction rates: Implications for cumulative culture. *PLOS ONE*, *9*(7), e102806. https://doi.org/10.1371/JOURNAL.PONE.0102806

Hooper, P. L., Demps, K., Gurven, M., Gerkey, D., & Kaplan, H. S. (2015). Skills, division of labour and economies of scale among Amazonian hunters and South Indian honey collectors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1683). https://doi.org/10.1098/rstb.2015.0008

Jaeggi, A. V., Hooper, P. L., Beheim, B. A., Kaplan, H., & Gurven, M. (2016). Reciprocal exchange patterned by market forces helps explain cooperation in a small-scale society. *Current Biology*, *26*(16), 2180–2187. https://doi.org/10.1016/j.cub.2016.06.019

Kajonius, P. J., & Johnson, J. (2018). Sex differences in 30 facets of the five factor model of personality in the large public (N = 320,128). *Personality and Individual Differences*, *129*, 126–130. https://doi.org/10.1016/J.PAID.2018.03.026

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Kapp, S. K., Pripas-Kapit, S., Dekker, M., Murray, D., Tisoncik, L. A., Baggs, M., … Russell, G. (2020). *Autistic community and the neurodiversity movement – Stories from the frontline* (S. K. Kapp, Ed.). Palgrave Macmillan. https://link.springer.com/content/pdf/10.1007%2F978-981-13-8437-0.pdf

Keller, M. C. (2018). Evolutionary perspectives on genetic and environmental risk factors for psychiatric disorders. *Annual Review of Clinical Psychology*, *14*(1), 471–493. https://doi.org/10.1146/annurev-clinpsy-050817-084854

Keller, M. C., Howrigan, D. P., & Simonson, M. A. (2011). Theory and methods in evolutionary behavioral genetics. *The Evolution of Personality and Individual Differences*. https://doi.org/10.1093/acprof:oso/9780195372090.003.0010

Keller, M. C., & Miller, G. (2006). Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behavioral and Brain Sciences*, *29*(4), 385–404. https://doi.org/10.1017/S0140525X06009095

Kelly, R. L. (2013). *The Lifeways of Hunter-Gatherers: The Foraging Spectrum*. Cambridge University Press. https://doi.org/10.1017/CBO9781139176132

Kotov, R., Jonas, K. G., Carpenter, W. T., Dretsch, M. N., Eaton, N. R., Forbes, M. K., … Watson, D. (2020). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. *World Psychiatry*, *19*(2), 151–172. https://doi.org/10.1002/WPS.20730

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Krupp, D. B., Sewall, L. A., Lalumière, M. L., Sheriff, C., & Harris, G. T. (2013). Psychopathy, adaptation, and disorder. *Frontiers in Psychology*, *4*(March), 1–5. https://doi.org/10.3389/fpsyg.2013.00139

Laajaj, R., Macours, K., Hernandez, D. A. P., Arias, O., Gosling, S. D., Potter, J., … Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, *5*(7). https://doi.org/10.1126/sciadv.aaw5226

Latzman, R. D., Krueger, R. F., DeYoung, C. G., & Michelini, G. (2021). Connecting quantitatively derived personality–psychopathology models and neuroscience. *Personality Neuroscience*, *4*, e4. https://doi.org/10.1017/PEN.2021.3

Legge, S. E., Santoro, M. L., Periyasamy, S., Okewole, A., Arsalan, A., & Kowalec, K. (2021). Genetic architecture of schizophrenia: A review of major advancements. *Psychological Medicine*, *51*(13), 2168–2177. https://doi.org/10.1017/S0033291720005334

Lightner, A. D., Heckelsmiller, C., & Hagen, E. H. (2021a). Ethnomedical specialists and their supernatural theories of disease. *Review of Philosophy and Psychology*. https://doi.org/10.1007/S13164-021-00589-8

Lightner, A. D., Heckelsmiller, C., & Hagen, E. H. (2021b). Ethnoscientific expertise and knowledge specialisation in 55 traditional cultures. *Evolutionary Human Sciences*, *3*. https://doi.org/10.1017/EHS.2021.31

Lilienfeld, S. O., Waldman, I. D., Landfield, K., Watts, A. L., Rubenzer, S., & Faschingbauer, T. R. (2012). Fearless dominance and the U.S. presidency: Implications of psychopathic personality traits for successful and unsuccessful political leadership. *Journal of Personality and Social Psychology*, *103*(3), 489–505. https://doi.org/10.1037/a0029392

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., … Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1). https://doi.org/10.1038/s41572-019-0138-4

Lukaszewski, A. W., Lewis, D. M. G., Durkee, P. K., Sell, A. N., Sznycer, D., & Buss, D. M. (2020). An adaptationist framework for personality science. *European Journal of Personality*, *34*(6), 1151–1174. https://doi.org/10.1002/PER.2292

Lukaszewski, A. W., & Roney, J. R. (2011). The origins of extraversion: Joint effects of facultative calibration and genetic polymorphism. *Personality and Social Psychology Bulletin*, *37*(3), 409–421. https://doi.org/10.1177/0146167210397209

Lukaszewski, A. W., & von Rueden, C. R. (2015). The extraversion continuum in evolutionary perspective: A review of recent theory and evidence. *Personality and Individual Differences*, *77*, 186–192. https://doi.org/10.1016/j.paid.2015.01.005

Lyon, B. E., & Montgomerie, R. (2012). Sexual selection is a form of social selection. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1600), 2266–2273. https://doi.org/10.1098/rstb.2012.0012

Macfarlan, S. J., & Lyle, H. F. (2015). Multiple reputation domains and cooperative behaviour in two Latin American communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1683). https://doi.org/10.1098/RSTB.2015.0009

Martin, J. S., & Jaeggi, A. V. (2021). Social animal models for quantifying plasticity, assortment, and selection on interacting phenotypes. *Journal of Evolutionary Biology*. https://doi.org/10.1111/JEB.13900

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *The American Psychologist*, *52*(5), 509–516. http://www.ncbi.nlm.nih.gov/pubmed/9145021

McGrath, J., Saha, S., Chant, D., & Welham, J. (2008). Schizophrenia: A concise overview of incidence, prevalence, and mortality. *Epidemiologic Reviews*, *30*(1), 67–76. https://doi.org/10.1093/epirev/mxn001

McNamara, J. M., & Leimar, O. (2010). Variation and the response to variation as a basis for successful cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2627–2633. https://doi.org/10.1098/rstb.2010.0159

McNamara, J. M., Stephens, P. A., Dall, S. R.., & Houston, A. I. (2008). Evolution of trust and trustworthiness: Social awareness favours personality differences. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1657), 605–613. https://doi.org/10.1098/RSPB.2008.1182

Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, *18*(03), 523. https://doi.org/10.1017/S0140525X00039595

Merikangas, K. R., Jin, R., He, J.-P., Kessler, R. C., Lee, S., Sampson, N. A., … Karam, E. G. (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry*, *68*(3), 241–251.

Montiglio, P.-O., Ferrari, C., & Réale, D. (2013). Social niche specialisation under constraints: personality, social interactions and environmental heterogeneity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1618), 20120343. https://doi.org/10.1098/rstb.2012.0343

Mortelliti, A., & Brehm, A. M. (2020). Environmental heterogeneity and population density affect the functional diversity of personality traits in small mammal populations. *Proceedings of the Royal Society B*, *287*(1940), 20201713. https://doi.org/10.1098/RSPB.2020.1713

Mouchet, A., Cole, E. F., Matthysen, E., Nicolaus, M., Quinn, J. L., Roth, A. M., … Dingemanse, N. J. (2021). Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. *Proceedings of the National Academy of Sciences*, *118*(28), 2021. https://doi.org/10.1073/PNAS.2024994118

Nesse, R. M. (2005). Maladaptation and natural selection. *The Quarterly Review of Biology*, *80*(1), 62–70. http://www.ncbi.nlm.nih.gov/pubmed/15884737

Nesse, R. M. (2007). Runaway social selection for displays of partner value and altruism. *Biological Theory*, *2*(2), 143–155. https://doi.org/10.1162/biot.2007.2.2.143

Nesse, R. M. (2019). *Good reasons for bad feelings: Insights from the frontier of evolutionary psychiatry*. Allen Land.

Nettle, D. (2006). The evolution of personality variation in humans and other animals. *American Psychologist*, *61*(6), 622–631. https://doi.org/10.1037/0003-066X.61.6.622

Nettle, D. (2011). Evolutionary perspectives on the five-factor model of personality. In *The evolution of personality and individual differences*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195372090.003.0001

Nettle, D., Gibson, M. A., Lawson, D. W., & Sear, R. (2013). Human behavioral ecology: Current research and future prospects. *Behavioral Ecology*, *24*(5), 1031–1040. https://doi.org/10.1093/beheco/ars222

Nonacs, P., & Kapheim, K. M. (2007). Social heterosis and the maintenance of genetic diversity. *Journal of Evolutionary Biology*, *20*(6), 2253–2265. https://doi.org/10.1111/j.1420-9101.2007.01418.x

Penke, L. (2010). Bridging the gap between modern evolutionary psychology and the study of individual differences. In D. M. Buss & P. H. Hawley (Eds.), *The evolution of personality and individual differences* (pp. 243–279). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195372090.003.0009

Penke, L., & Jokela, M. (2016). The evolutionary genetics of personality revisited. *Current Opinion in Psychology*, *7*, 104–109. https://doi.org/10.1016/j.copsyc.2015.08.021

Plomin, R. (2011). Commentary: Why are children in the same family so different? Non-shared environment three decades later. *International Journal of Epidemiology*, *40*(3), 582–592. https://doi.org/10.1093/ije/dyq144

Polanczyk, G. V, Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. *International Journal of Epidemiology*, *43*(2), 434. https://doi.org/10.1093/IJE/DYT261

Polderman, T. J. C., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. https://doi.org/10.1038/ng.3285

Polimeni, J., & Reiss, J. P. (2002). How shamanism and group selection may reveal the origins of schizophrenia. *Medical Hypotheses*, *58*(3), 244–248. https://doi.org/10.1054/mehy.2001.1504

Power, R. A., Kyaga, S., Uher, R., MacCabe, J. H., Långström, N., Landen, M., … Svensson, A. C. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry*, *70*(1), 22. https://doi.org/10.1001/jamapsychiatry.2013.268

Rogers, W. A., & Walker, M. J. (2017). The line-drawing problem in disease definition. *The Journal of Medicine and Philosophy*, *42*(4), 405–423. https://doi.org/10.1093/JMP/JHX010

Salali, G. D., Chaudhary, N., Thompson, J., Grace, O. M., van der Burgt, X. M., Dyble, M., … Migliano, A. B. (2016). Knowledge-sharing networks in hunter–gatherers and the evolution of cumulative culture. *Current Biology*, *26*(18), 2516–2521. https://doi.org/10.1016/J.CUB.2016.07.015

Sasson, N. J., Lam, K. S. L., Childress, D., Parlier, M., Daniels, J. L., & Piven, J. (2013). The broad autism phenotype questionnaire: Prevalence and diagnostic classification. *Autism Research*, *6*(2), 134–143. https://doi.org/10.1002/aur.1272

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of big five personality traits. *Journal of Cross-Cultural Psychology*, *38*(2), 173–212. https://doi.org/10.1177/0022022106297299

Schniter, E., Wilcox, N. T., Beheim, B. A., Kaplan, H. S., & Gurven, M. (2018). Information transmission and the oral tradition: Evidence of a late-life service niche for Tsimane Amerindians. *Evolution and Human Behavior*, *39*(1), 94–105. https://doi.org/10.1016/J.EVOLHUMBEHAV.2017.10.006

Shuster, S. M. (2010). Alternative mating strategies. In D. F. Westneat & C. W. Fox (Eds.), *Evolutionary behavioral ecology* (pp. 434–450). Oxford University Press. https://doi.org/10.1163/156853984x00290

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: An integrative overview. *The Quarterly Review of Biology*, *79*(3), 241–277. https://doi.org/10.1086/422893

Singh, M. (2017). The cultural evolution of shamanism. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/s0140525×17001893

Singh, M., & Henrich, J. (2020). Why do religious leaders observe costly prohibitions? Examining taboos on Mentawai shamans. *Evolutionary Human Sciences*, *2*. https://doi.org/10.1017/EHS.2020.32

Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., … Daly, M. J. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature 2022*, 1–9. https://doi.org/10.1038/s41586-022-04556-w

Slatkin, M. (1979). Frequency- and density-dependent selection on a quantitative character. *Genetics*, *93*(3), 755–771. https://doi.org/10.1093/genetics/93.3.755

Smaldino, P. E., Lukaszewski, A., von Rueden, C., & Gurven, M. (2019). Niche diversity can explain cross-cultural differences in personality structure. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0730-3

Smith, D., Schlaepfer, P., Major, K., Dyble, M., Page, A. E., Thompson, J., … Migliano, A. B. (2017). Cooperation and the evolution of hunter–gatherer storytelling. *Nature Communications 2017 8:1*, *8*(1), 1–9. https://doi.org/10.1038/s41467-017-02036-8

Smith, E. A. (1985). Inuit foraging groups: Some simple models incorporating conflicts of interest, relatedness, and central-place sharing. *Ethology and Sociobiology*, *6*(1), 27–47. https://doi.org/10.1016/0162-3095(85)90039-1

Smith, J. M. (1976). Evolution and the theory of games. *American Scientist*, *64*(1). https://doi.org/10.1016/0377-2217(83)90101-7

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, *43*(2), 476–493. https://doi.org/10.1093/IJE/DYU038

Sugiyama, L. S., & Sugiyama, M. S. (2003). Social roles, prestige, and health risk. *Human Nature*, *14*(2), 165–190. https://doi.org/10.1007/s12110-003-1002-4

Taylor, R. W., Boutin, S., Humphries, M. M., & Mcadam, A. G. (2014). Selection on female behaviour fluctuates with offspring environment. *Journal of Evolutionary Biology*, *27*(11), 2308–2321. https://doi.org/10.1111/JEB.12495

Thayer, Z. M., Wilson, M. A., Kim, A. W., & Jaeggi, A. V. (2018). Impact of prenatal stress on offspring glucocorticoid levels: A phylogenetic meta-analysis across 14 vertebrate species. *Scientific Reports* 2018 8:1, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-23169-w

Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality*, *58*(1), 17–67. https://doi.org/10.1111/j.1467-6494.1990.tb00907.x

Tooby, J., & Cosmides, L. (1996). Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy*, *88*, 119–143.

Troisi, A. (2005). The concept of alternative strategies and its relevance to psychiatry and clinical psychology. *Neuroscience and Biobehavioral Reviews*, *29*(1 spec. iss.), 159–168. https://doi.org/10.1016/j.neubiorev.2004.06.012

Uchiyama, R., Spicer, R., & Muthukrishna, M. (2021). Cultural evolution of genetic heritability. *Behavioral and Brain Sciences*, 1–147. https://doi.org/10.1017/S0140525X21000893

Ulrich, Y., Saragosti, J., Tokita, C. K., Tarnita, C. E., & Kronauer, D. J. C. (2018). Fitness benefits and emergent division of labour at the onset of group living. *Nature*, *560*(7720), 635–638. https://doi.org/10.1038/s41586-018-0422-6

Verweij, K. J. H., Yang, J., Lahti, J., Veijola, J., Hintsanen, M., Pulkki-Råback, L., … Zietsch, B. P. (2012). Maintenance of genetic variation in human personality: Testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. *Evolution*, *66*(10), 3238–3251. https://doi.org/10.1111/j.1558-5646.2012.01679.x

Von Rueden, C. R. (2014). The roots and fruits of social status in small-scale human societies. In C. Anderson (Ed.), *The psychology of social status* (pp. 179–200). Springer. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2500276

Von Rueden, C. R., Gurven, M., & Kaplan, H. (2008). The multiple dimensions of male social status in an Amazonian society. *Evolution and Human Behavior*, *29*(6), 402–415. https://doi.org/10.1016/j.evolhumbehav.2008.05.001

Von Rueden, C. R., & Jaeggi, A. V. (2016). Men's status and reproductive success in 33 nonindustrial societies: Effects of subsistence, marriage system, and reproductive strategy. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(39), 10824–10829. https://doi.org/10.1073/pnas.1606800113

Von Rueden, C. R., Redhead, D., O'Gorman, R., Kaplan, H., & Gurven, M. (2019). The dynamics of men's cooperation and social status in a small-scale society. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1908), 20191367. https://doi.org/10.1098/rspb.2019.1367

Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin*, *141*(4), 769–785. https://doi.org/10.1037/bul0000017

Wakefield, J. C. (2015). Biological function and dysfunction: Conceptual foundations of evolutionary psychopathology. In *The handbook of evolutionary psychology* (pp. 1–19). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119125563.evpsych242

https://doi.org/10.1017/ehs.2022.23 Published online by Cambridge University Press

Walker, R., Hill, K., Kaplan, H., & McMillan, G. (2002). Age-dependency in hunting ability among the Ache of Eastern Paraguay. *Journal of Human Evolution*, *42*(6), 639–657. https://doi.org/10.1006/jhev.2001.0541

West-Eberhard, M. J. (1979). Sexual selection, social competition, and evolution. *Proceedings of the American Philosophical Society*, *123*(4), 222–234. https://doi.org/10.2307/2828804

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.

Widiger, T., Gore, W., Crego, C., Rojas, S., & Oltmanns, J.. (2017). Five-factor model and personality disorder. In Thomas A. Widiger (Ed.), *The Oxford handbook of the five factor mode* (pp. 449–478). Oxford University Press.

Widiger, T. A., Sellbom, M., Chmielewski, M., Clark, L. A., DeYoung, C. G., Kotov, R., … Wright, A. G. C. (2018). Personality in a hierarchical model of psychopathology. *Clinical Psychological Science*, *7*(1), 77–92. https://doi.org/10.1177/2167702618797105

Willems, E. P., & van Schaik, C. P. (2017). The social organization of *Homo ergaster*: Inferences from anti-predator responses in extant primates. *Journal of Human Evolution*, *109*, 11–21. https://doi.org/10.1016/J.JHEVOL.2017.05.003

Williams, J., & Taylor, E. (2006). The evolution of hyperactivity, impulsivity and cognitive diversity. *Journal of The Royal Society Interface*, *3*(8), 399–413. https://doi.org/10.1098/rsif.2005.0102

Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(46), 23004–23010. https://doi.org/10.1073/pnas.1908430116

Winkelman, M. J. (1990). Shamans and other 'magico-religious' healers: A cross-cultural study of their origins, nature, and social transformations. *Ethos*, *18*(3), 308–352. https://doi.org/10.1525/eth.1990.18.3.02a00040

Winterhalder, B. (1986). Diet choice, risk, and food sharing in a stochastic environment. *Journal of Anthropological Archaeology*, *5*(4), 369–392. https://doi.org/10.1016/0278-4165(86)90017-6

Wolf, M., Doorn, G. S. Van, & Weissing, F. J. (2011). On the coevolution of social responsiveness and behavioural consistency. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1704), 440. https://doi.org/10.1098/RSPB.2010.1051

Wolf, M., & Krause, J. (2014). Why personality differences matter for social functioning and social structure. *Trends in Ecology & Evolution*, *29*(6), 306–308. https://doi.org/10.1016/J.TREE.2014.03.008

Zachar, P., & Kendler, K. S. (2017). The philosophy of nosology. *Annual Review of Clinical Psychology*, *13*(1), 49–71. https://doi.org/10.1146/annurev-clinpsy-032816-045020

Cite this article: Hunt AD, Jaeggi AV (2022). Specialised minds: extending adaptive explanations of personality to the evolution of psychopathology. *Evolutionary Human Sciences* **4**, e26, 1–23. https://doi.org/10.1017/ehs.2022.23

Making Wakefield Workable: the fitness and function framework for taxonomising evolutionary dysfunction

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

Abstract

Jerome Wakefield's 'Harmful Dysfunction Analysis' (HDA) recognises biological and psychological disorder exists wherever two conditions are met: concurrent harm, a subjective value component, and dysfunction, an objective component related to the interruption of evolutionarily selected effects. This is arguably the leading definition of disorder, and is prominently referenced in evolutionary psychiatry, yet suffers various criticisms. These particularly concern the 'dysfunction' component, which is undermined by the indeterminacy of ancestral function and a range of complex counter-examples, particularly relating to mismatch, by-products and extremes of adaptive spectrums. To tackle these criticisms, I provide a definitional framework which allows clear pathways to dysfunction attribution in the cases problematic for Wakefield. The key move is in distinguishing biological objects with fitness, a fundamentally quantifiable variable, from the selective processes of function and dysfunction which lead to that fitness, which rely on qualitative descriptions, and are much harder to exactly specify and demarcate to the level of precision necessary for the HDA. I also note the importance of specifying framing environments and trait descriptions. The resulting framework solves various problems with Wakefield's account and leads to a taxonomy of different classes of dysfunction. This is unique amongst existing evolutionary taxonomies of disorder by offering strict demarcation, exclusivity and greater theoretical completeness. By relying ultimately on fitness effects rather than descriptions of dysfunctional processes and recognising distinct possible classes of evolutionary dysfunction, a more tractable direction for scientific investigation into any trait's dysfunctional status is offered, with the potential to make Wakefield's HDA workable.

1/ Introduction

Despite being central concepts, evolutionary biology lacks precise shared definitions for 'adaptation' or 'function', and clearly distinguishing adaptations from, say, neutral variation, is a longstanding (Williams 1966) and still unresolved issue (Sober, 2000, p85-87). Current mainstream methods rely primarily on analogies with human engineering (Boudry and Pigliucci 2013), and appeals to evident function in solving adaptive problems (Wright 1976). These suffice in identifying self-evident adaptations (e.g. eyes), and stricter definitions generally aren't a scientific concern, although debates over the presence of adaptation or sufficient evidence to prove design are common (Orzack and Sober 1994), especially in

evolutionary psychology (Holcomb 1996). However, the necessity of strictly defining these concepts amplifies dramatically when needing to draw a line between health and disorder.

Jerome Wakefield's (1992, 1997, 2015) Harmful Dysfunction Analysis (HDA) defines 'disorder' by a dual criterion, of concurrent harm (a subjective component, defined by social values) and dysfunction (an objective component, defined by an interruption of evolutionarily selected effects). Whilst widely recognised by scientists, medical professionals and philosophers (Forest and Faucher 2021), it suffers from an array of criticisms which hinder its implementation in medicine and science generally (Faucher and Forest 2021), especially aimed at the 'dysfunction' component. The purpose of this paper is to clarify viable definitions of 'adaptation', 'dysfunction' and related concepts which better captures the complexity of biological evolution and the limits of scientific demarcation of health and disorder.

After briefly introducing previous literature on defining and identifying adaptation, I consider Wakefield's HDA and its major criticisms before suggesting clarifications required to make it workable. Firstly I expand the function/dysfunction dichotomy into the 'fitness and function' framework. The most critical move is distinguishing the objects of selection (e.g. adaptations such as eyes), to which can be ascribed fitness, a quantitative variable, from the processes causing selection (e.g. functions such as seeing), which are not naturally quantifiable, and thus unreliable for precise demarcation. A critical advantage of this distinction is acknowledging dysfunctional processes exist wherever there are costs to fitness, allowing the HDA to overcome major criticisms related to the difficulty of ascertaining precise function. I go on to note two essential 'specifiers' for analyses. One specifies *description* of the target to be categorised, the other specifies a *framing* population and environment. I show why evolutionary definitions without these specifiers are unacceptably imprecise. Combining the fitness and function framework of definitions with the two specifiers, a path is cleared for naturalistic classification of biological entities into these categories. The resulting five classes of dysfunction are more complex than Wakefield's function/dysfunction dichotomy, but necessarily so, and solve critical problems in the HDA, as well as providing an evolutionary taxonomy of classes of disorder which can claim strict demarcation, exclusivity, and greater theoretical completeness.

The purpose of this paper is not to deny or revise Wakefield's HDA as a matter of conceptual analysis of the meaning of 'disorder', but to define critical evolutionary categories more carefully. Ideally, it should be possible for the scientific community to say 'given current evidence we can define X trait as an disorder' or 'to know whether X is dysfunctional we need Y further evidence'. Without clear definitions this is impossible in many important cases, especially of psychiatric interest (Hunt, Abed, and St John-Smith 2022). Wakefield's HDA concentrated on conceptual analysis rather than practical application; my contribution concentrates on stipulating definitions which can be scientifically pursued.

2/ Definitions of Adaptation in the Evolutionary Sciences

After Darwin's (1869) initial proposition of evolution by natural selection, the 20th century brought huge advances in mathematical models, (Fisher 1930) theoretical understanding, (Maynard Smith 1993) and empirical evidence for evolution by natural selection. The 21st century has continued seeing scientific and technological developments relevant to evolutionary theory, especially in genetics. Despite this, somewhat surprisingly, certain fundamental concepts do not share precise definitions: for example, definitions of key concepts such as 'natural selection' and 'fitness' differ between textbooks (Gregory 2009, Table 1) and definitions of 'adaptation' have been quite different depending on the goals of the author (Reeve and Sherman 1993).

George Williams' famous 'Adaptation and Natural Selection' (Williams 1966) was a landmark in outlining the expected characteristics of adaptations[1] (and the fallacy of arguments from group selection). To Williams, adaptation and adaptive design explain complex biological traits, and the necessity of believing adaptive design caused a trait is evident in their efficiency at performing some function or goal. He claims 'any biological mechanism produces at least one effect that can properly be called its goal: vision for the eye' (1966, p.8). Yet he recognises limitations in defining and defending the presence or absence of adaptation. 'It is often easy, in practice, to perceive functional design intuitively, but unfortunately disputes sometimes arise as to whether certain effects are produced by design or merely as by-products of some other function. The formulation of practical definitions and sets of objective criteria will not be easy, but it is a problem of great importance and will have to be faced.' (1966, p.9).

Williams himself relies on informal arguments, especially analogy between organic systems and human technologies, which provide 'indirect evidence of complexity and constancy' (p.10). Mainstream evolutionary biology and psychology has carried this approach of 'reverse engineering' forward (Dawkins 1986; Pinker 1997; Symons 1990). Eyes are complex organs with an array of systems and components perfectly suited for vision (Goldsmith 1990), so vision must be the function of eyes. Whilst this suffices to justify uncontroversial cases such as eyes, it struggles where disputes arise, for example in reverse engineering the function of low mood – it's plausible some function exists, but it is much harder to discern and make an uncontroversial argument. This becomes a serious problem for crisply demarcating health and disorder.

Stricter definitions of adaptation have been proposed in the philosophy of biology. In particular, Elliot Sober's:

[1] 'Adaptation' can refer to the *process* resulting from natural selection by which organisms adapt to their environments or the physical features or adaptive *traits* which arise due to this process.

> Characteristic *c* is an adaptation for doing task *t* in a population if and only if members of the population now have *c* because, ancestrally, there was selection for having *c* and *c* conferred a fitness advantage because it performed task *t.* (Sober, 2000, p.85)[2]

This avoids reliance on analogy or intuition of design, and moves disputes into a different realm of scientific inquiry – if the facts regarding the historical fitness advantages of a trait are revealed, it can be confirmed as an adaptation. The low mood system is recognised as an adaptation not because we can reverse engineer its function, but because there was selection for it and it conferred a fitness advantage. The function of the adaptation of low mood (task *t*) is a different (but related) question. The plausibility of scientific inquiry into ancestral fitness aside, this definition of adaptation avoids relying on analogy to engineering or intuition of design.

Concentration on fitness contradicts traditional adaptationist programs in evolutionary psychology which explicitly care about describing functional mechanisms (Symons 1990), but is aligned with evolutionary biology more broadly (Reeve and Sherman 1993). Referencing fitness effects also allows a clearer framework for distinguishing disorder, both in principle and scientific practice – showing why is the aim of this paper.

3/ Wakefield's Harmful Dysfunction Model

Medical terminology is widely used yet poorly defined in common parlance. Terms such as disease, sickness, pathology, disorder, illness, dysfunction, condition, affliction, malady, and ailment have different connotations and norms of use (Hunt et al. 2022), but are not strictly defined by objectively verifiable criteria. The unifying feature of such terms is simply that they are used to describe biological states deemed to be bad and medically relevant, by the sufferer or society (Murphy 2020).

The DSM-5 (American Psychiatric Association, 2013a) defines mental disorder as: "a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning". It does not further define 'dysfunction'. This allows essentially arbitrary distinction methods (Khoury, Langer, and Pagnini 2014). Past claims that masturbation is a mental disorder, or leads to mental disorders (Engelhardt 1999) and the case of homosexuality being removed from the DSM on a vote (Drescher 2015) exemplify this reliance on social norms rather than empirical facts.

In psychiatry especially, because of past over-pathologisation and a fervent anti-psychiatry movement which attempts to dismiss psychiatry as little more than social control (Crossley 1998), philosophers of medicine have sought ways to 'naturalise' terminology, grounding definitions in objective facts; "a value-free scientific disease concept [is] a bedrock

[2] The description and framing specifiers introduced in section 7 are essentially extensions accounting for the importance of specifically defining 'characteristic', 'population' and 'ancestrally' in Sober's definition.

requirement to block the subversion of medicine by political rhetoric or normative eccentricity" (Boorse, 1997,p.100).

The most influential of these is Jerome Wakefield's definition of disorder as 'harmful dysfunction' (Wakefield, 1992, 2008; Klein, 1999; Spitzer, 1999) in his HDA. This is a hybrid definition with a subjective and objective component. Wakefield proposes we attribute the existence of 'disorder' to cases of concurrent harm, in a broad, value-defined sense, and dysfunction, which is explicitly based in a concept of evolutionary function. A trait is harmful if society or the affected individual says it is harmful, and is a dysfunction when it fails to perform its evolutionarily selected effect. Eyes evolved to see, thus eyes which can't see are dysfunctional. Psychological and physical traits functioned to provide reproductive success to our ancestors, and where that function fails or is interrupted, there is dysfunction. If a trait such as extreme jealousy is considered harmful by a society, then the question of its status as a disorder depends on whether that jealousy was functional ancestrally. Wakefield is explicit that the HDA's primary aim is *conceptual analysis*; an analysis of how the term and concept of 'disorder' is used. If a trait is harmful but not dysfunctional, or dysfunctional but not harmful, we should not consider it a disorder – but if we believe that a trait is both harmful and dysfunctional, we should call it a disorder.

This belongs to a class of models of health and disorder variously called 'backwards looking', 'etiological' or 'selected effects' (Millikan 1989; Neander 1991; Wright 1976). Closely related to concepts of function and normativity in the philosophy of biology and the life sciences (e.g. Varner, 1998), backwards looking models of disease explicitly reference the theory of evolution by natural selection in discerning function and dysfunction, and thus health and disease. Although one might *say* that atoms, molecules and all physical processes are 'functioning' to the extent that they play specific roles in physical systems, this use of 'function' does not carry the goal-directedness with which 'function' can be used in the life sciences. If the sun is warming a puddle and making it evaporate, we could *say* that the photons' function are evaporating the puddle – but we cannot say the photons exist *to* evaporate the puddle; the evaporation of the puddle is an incidental occurrence resulting from the properties of water and photons. If instead a fox starts drinking the puddle, the sense in which the fox is functioning in draining the puddle is different because it is *supposed* to drink water. Biological systems are products of natural selection – their present existence, form and processes exist because they served a historical function, or goal, of achieving reproductive success. Although evolution and life may not have an end-goal in a classical teleological sense, it is nevertheless goal-defined: biological entities and processes have been repeatably selected by their ability to survive and reproduce (Williams (1966, p.258) used the term 'teleonomy'). Thus, the fox can be said to be functioning in a way which the photons cannot. And thus, dysfunction in the fox can be based in objective fact, whilst it cannot in the photon.

Despite challenges from the 'causal role' school, who characterise function by current causal role in accomplishing assumed capacities of a containing system (Ariew, Cummins, and

Perlman 2002; Cummins 1975), even Wakefield's critics (Faucher and Forest 2021) largely agree that the concept of evolution-based, backwards-looking approaches are conceptually superior as naturalistic approaches in defining disorder. One criticism of causal role approaches is that they are descriptive rather than normative – they lack a necessary 'should' (Neander 1991). Without historical reference grounding an objectively defined normativity, definitions can be swayed by fluctuating social norms. It is also unclear how causal role accounts could provide a naturalistic differentiation between traits of psychiatric interest, for example distinguishing normal low mood from depression as a disorder. This paper shall follow evolutionary psychiatry and evolutionary medicine (Hunt et al. 2022) in considering this the account of most importance and potential. However, this fundamental conceptual strength has not yet led to the widespread adoption of evolutionary principles of defining disorders in mainstream medicine (Hidaka et al. 2015) and psychiatry (Abed & St John-Smith, 2022).

3.1/ Criticisms of Wakefield's 'Dysfunction'

As the most prominent definition of disorder in the philosophy of medicine, and with substantial recognition outside of philosophy, the HDA has understandably been the target for much criticism; indeed, a recently published book "Defining Mental Disorder: Jerome Wakefield and His Critics" (Faucher and Forest 2021) contained 13 chapters written by critics, and 15 lengthy responses by Wakefield – a Herculean effort and unparalleled resource for analysis of the state of the HDA thirty years after its inception. The most numerous criticisms are against the 'dysfunction' component of the HDA.

These criticisms generally follow two strategies. On the one hand, they argue that Wakefield's evolutionary approach to defining dysfunction does not accurately fit our perception of what is and is not a disorder – they claim the conceptual analysis fails (Faucher 2021; Kincaid 2021; De Vreese 2021); a case of autism is still seen as a disorder even if proven to result from adaptive cognitive specialisation (Hunt and Jaeggi 2022). This is not a criticism this paper is particularly concerned with – I think Wakefield's retorts are relatively sound, usually by claiming that the author's understanding of how the dysfunction component would apply in their particular examples is incorrect (e.g. (Wakefield, 2021j)) and re-affirming that we (and he) would update our label of 'disorder' around precisely harmful dysfunctions. The second criticism more commonly raised – and to my mind the one that makes the current HDA unworkable – is the question of whether the dysfunction component is actually scientifically usable. Can the HDA be applied to science and medicine and reveal which conditions are disorders and which are not? This problem is recognised outside of philosophy; Leif Kennair, a psychologist working on clinical psychology and evolutionary psychology has written "despite Wakefield's definition having the greatest promise of providing mental health research with a theoretically and scientifically based overarching definition of psychopathology, it currently lacks the research needed to be able to categorise mental states (harmful dysfunction and non-harmful dysfunction, as well as harmful function and non-harmful function) and create a more valid nosology." (Kennair, 2010, p.453). Wakefield

includes the objective component as protection against anti-psychiatrists and pure value theorists, who can claim psychiatric labels are arbitrary. However, if the objective component cannot be clearly identified – if it is often hard, or indeed impossible, to decide whether a particular condition is functioning or dysfunctioning from an evolutionary perspective, then the objective component of the HDA is essentially useless. Various authors believe this is indeed the case, and thus the HDA should be abandoned or revised (De Block and Sholl 2021; Faucher 2021; De Vreese 2021). Their arguments come in many forms.

Faucher (2021) labels this general criticism of the dysfunction component the 'epistemic objection'. This argues that "because the harmful dysfunction analysis holds that whether one has a disorder depends on facts about internal mechanisms and their evolutionary history, and we are largely ignorant of these facts, therefore the analysis implies that it is impossible to know at this time whether conditions are disorders or non-disorders" (Bergner, 1997, p. 255). This is a general claim about the scientific plausibility of discovering the relevant facts to discern dysfunction when couched in evolutionary terms. Although theoretically objective by being historical and not altered by contemporary norms, traits can often not be conclusively defined as functional or dysfunctional because history is unobservable. This is a continuous problem for backward-looking approaches (Holm 2012), and faces all evolutionary hypotheses, but especially those in psychology and psychiatry. We lack relevant repeatable, observable evidence. We can't replay human evolutionary history in a lab, or ever in the future. The backward looking account's greatest strength, of being historical and unalterable, and therefore entirely objective, is thus also a serious weakness.

The general epistemic problem with the dysfunction component is exaggerated by several more specific criticisms:

1) *Extreme tails are selected* (Kincaid 2021). It is possible that evolution has selected for dimensions of traits, but we mistakenly label the tails of the distribution as disorder. If so, these are cases of disorders which have been selected in some sense – but there is no way to verify or falsify this, by knowing which parts of the dimension were functional.
2) *Evolutionary by-products* (Murphy 2021). The HDA defines dysfunction as a failure to achieve selected effects; however, many disorders may occur in by-products or accidental genetic 'hitch-hikers', which themselves have no selected effects, and so are beyond Wakefield's definition of dysfunction, despite seeming clearly dysfunctional.
3) *Mismatch* (Garson 2021). Evolutionary mismatch occurs when adaptive traits become maladaptive in modern environments. This seems to be a case of failing to achieve selected effects. However, this means that social values can cause dysfunction (e.g. if aggression becomes punished rather than admired), undermining the objectivity of the dysfunction component.
4) *Indeterminacy of function* (Thornton 2021). There are many ways of describing a trait's function. The function of the heart could be to pump blood; or to pass oxygen around

> the body; or to enable the individual to survive and reproduce; or given a multiplicity of other, seemingly quite different framings. These are all selected effects of sorts – where and how should dysfunction be identified? This becomes particularly complex when considering psychological traits such as language, and having to account for its many positive effects in the human and pre-human lineage. Function may also change from generation to generation (e.g. the function of eyes changed with the invention of writing). If indeterminacy of function is rife and empirically unsolvable, defining dysfunction by failure of selected effects becomes unworkable.

Of Wakefield's many lengthy responses to his critics (Wakefield 2021a, 2021h, 2021f, 2021i, 2021o, 2021d, 2021k, 2021j, 2021g, 2021b, 2021m, 2021n, 2021e, 2021c, 2021l), his primary line of argument is defending the HDA as conceptual analysis – therefore he *doesn't have to* answer concerns of how to discern dysfunction – all he maintains is that when we do, we will update our conception of whether the trait is a disorder. In cases where critics claim the HDA misses dysfunction or attributes dysfunction to non-disorders, Wakefield often responds by re-analysing the examples given and finding that the HDA can justify our intuitions (e.g. Wakefield, 2021g). In one case (responding to Garson (2021) on mismatch) he actually tweaks his definition of dysfunction to identify the dysfunction as existing in the internal mechanism and under the appropriate (ancestral-like) circumstances: "*a dysfunction is the inability of a mechanism to perform its function for internal reasons even under the appropriate circumstances for which it was selected to perform that function."* (Wakefield 2021j). The efficacy of this alteration is returned to in section 7.2; for now, suffice to say that this alteration does not improve the scientific tractability of the dysfunction component because identifying the 'appropriate circumstances' is problematic.

In defending the strength of the HDA's conceptual analysis, I consider Wakefield generally successful, yet missing the bigger, more crucial point made by critics of the dysfunction component – how can we apply this concept of dysfunction to identify disorder in science and medicine? Here Wakefield is on (self-admitted) less certain ground (Wakefield 2021d, 2021h). He makes two general points. Firstly, he reflects on the growth of science in relevant areas, and gives examples of how, for instance, new evidence from genetics can be used to decide whether a trait is dysfunctional or not (e.g. if we find alleles predisposing to autism or ADHD have been positively selected or confer advantages). He emphasises that it will be science's job to identify dysfunction, and expresses optimism that this job is possible, despite the scepticism of critics: "what seems epistemically challenging now may not be in the future" (Wakefield 2021d). Secondly, he argues that until scientific clarity is reached, appeals to intuition are a valid initial method of assessment as to whether a trait is dysfunctional. We can speak of dysfunction without precisely knowing what it consists of – which he calls a 'black-box' view (Lemoine 2021) – and can trust that intuitions and imprecise analysis of function and dysfunction are at least somewhat accurate. Just as we don't need knowledge of the inner mechanics of a car to understand whether it is broken or working, we should similarly accept that we can assume that many mental disorders are indeed dysfunctions

without precisely understanding their cause. Although this reliance on intuition has been called a 'dangerous game' (Faucher 2021), especially in the case of mental disorders, which are seriously open to cultural interpretation (consider homosexuality), Wakefield recognises this weakness, and leans into the conceptual analysis aspect of the HDA, emphasising these intuitive assessments of dysfunction are entirely at the mercy of scientific developments. "There are simply no conditions— not severe depression, not schizophrenia, not psychopathy—that are so indefeasibly considered disorders that no new information about their etiology could persuade us differently." (Wakefield, 2021c, p359). This retort doesn't, however, help us discern dysfunction.

4/ Making Wakefield Workable – The Fitness and Function Framework of definitions

This paper's aim is not further conceptual analysis of 'disorder', but to tackle this problem of a clear and scientifically tractable definition of the 'dysfunction' component used in the HDA. I agree with various naturalistic accounts (Millikan 1989; Neander 1991; Wright 1976) that argue definitions of terms of health and disease require evolutionary theory; evolutionary theory can make medical terminology scientifically sound. However, I also agree with critics that previous attempts at evolution-informed definitional frameworks, including Wakefield's, have troubling flaws (Faucher and Forest 2021). Evolutionary processes lead to more complexity than Wakefield's dysfunction component currently accounts for.

My contribution is stipulative in suggesting refining definitions, yet broadly follows common usage, hopefully allowing smooth integration into modern science, medicine and public discourse. Indeed, the proposed alterations may meet common usage and intuition more closely than Wakefield's HDA, especially in complex and borderline cases recognised as problematic. For brevity I cannot engage with the lengthy debate on defining 'function' historically (Wouters 2005), beyond noting my account generally follows use by evolution-referencing theorists (Godfrey-Smith 1994; Sober 2000) but is honed to allow scientific attribution of disorder-related terminology.

The 'fitness and function' framework (FAFF) of definitions supplied here aims to avoids many of the HDA's definitional flaws, although stops short of providing an adequate scientific method – the FAFF simply makes such science *possible* in a way the HDA does not. To do this, it expands Wakefield's functional/dysfunctional dichotomy into a set of six definitions, composed of two related trios defined by their negative, neutral and positive effects on reproduction. On the one hand, negative, neutral and positive *fitness*[3] can be observed in *biological objects*[4] (biological systems, structures, characteristics, states, traits; phenotypes

[3] I use 'fitness' and 'reproductive success' interchangeably.

[4] Note that although 'object' is being used here, processes and systems (e.g. the limbic system, photosynthesising systems, cellular transport systems) qualify; the key qualification is a biological apparatus or entity which is an object of selection – that which is being selected – in contrast to the reasons for that selection.

and genotypes; biological things; for example, eyes, depression, anger, schizophrenia). On the other hand, we can observe the negative, neutral and positive *selection processes* of *functioning* (reasons for survival and reproduction; fitness-beneficial and costly actions and events; for example, seeing, disengaging from conflict, threatening others, paranoid socialisation).[5] Biological objects are characterised by their fitness, and that fitness results from selection processes, which involve functions (and dysfunctions).

Distinguishing fitness and function and applying the negative/neutral/positive possibilities, the six definitions can be formulated as below. [6]

Biological objects can be categorised by fitness thus:

Adaptations increase reproductive success.

Passengers have no effect on reproductive success.[7]

Disruptions reduce reproductive success.

Selection processes can be categorised by function thus:

Functions are processes which increase reproductive success.

Neutral have no effect on reproductive success.

Dysfunctions are processes which reduce reproductive success.[8]

As shall be explained in detail in section 7, for this definitional framework to be reliably applied, the two specifiers of *description* and *framing* are required. These essentially force a FAFF analysis to exactly *describe* the set of biological objects in question (e.g. are you describing 'depression' as individuals who have been diagnosed by a psychiatrist using the Hamilton Rating Scale for Depression, or by self-report of people who answer 'yes' to 'are you feeling depressed right now?') and also *frame* the analysis within a particular time and environment (e.g. are you trying to characterise whether depression is dysfunctional within industrialised, urban, Western societies, or whether it was dysfunctional over an evolutionary

[5] Note that although 'processes' is being used here, the term 'processes' in biology may refer to objects and systems (e.g. metabolic processes). Processes in the sense meant here are the activities affecting reproductive success: they require situating over a period of time and observation of how they affect the organism's fitness.
[6] Note that common parlance today uses 'adaptive' and 'maladaptive' as verbs to imply direction of effects; although it is more common to specify the process which is adaptive as the function. In the FAFF, the meaning of 'adaptive' and 'maladaptive' as verbs are unspecified, and adaptation, passenger and disruption are explicitly nouns. Function, neutral and dysfunction are the only specified verbs. The best specific terminology here may be debated; the definition of the concepts is the critical point.
[7] 'Passengers' and 'neutral' effects are included for theoretical completeness but will be mostly left aside.
[8] Wakefield's original definition of dysfunction was 'interruption of selected effects' (1992, 1997, 2015). Selected effects to Wakefield are those effects which increased fitness – functions in the FAFF. Note a key difference, that Wakefield defined dysfunction by *interruption* of function – the FAFF instead defines dysfunction as processes reducing fitness, without referencing the specific function being interrupted (the importance of this alteration is made clear in sections 5 and 6).

history of hundreds of thousands of years?). No precise analysis can be done without specifying the described object and the framing environment.

Once description and framing are specified, we can ask questions of fitness and function. Start with fitness: if depression (a 'biological object' in the FAFF) increases fitness, it's an adaptation. If it has a negative effect on fitness, it's a disruption. We can then ask questions around the precise selection effects: depression might cause disengagement from hopeless goals and unfavourable social conditions (Nesse 2000), or cause useful rumination (Andrews and Thomson Jr 2009); positive effects on fitness which then count as functions. Alternatively, depression might have dysfunctional effects in restricting socialising, engagement in relationships, childcare, working, and so on. Depression could classify as an adaptation (if it increases fitness) but show both functional and dysfunctional effects simultaneously (e.g. disengaging from conflict is protective, which is functional, but also causes the loss of friends, which is dysfunctional). This complexity is discussed later, and allows the taxonomising of distinct forms of dysfunction. Specifiers also account for various important complexities: if depression is redefined to be extremely severe, or extremely mild and akin to normal low mood, the analysis will change with the description specifier. If depression is adaptive in communal living settings similar to human evolutionary history but useless and harmful in urbanised, industrialised environments, the framing specifier accounts for the flipping of adaptation into disruption. Again, this complexity is discussed later and affects the taxonomising of dysfunction.

The basic process of defining with the FAFF is to describe a trait, frame it in a particular environment, and investigate its fitness and the processes by which it achieves them. Any biological object or trait (including psychological or psychiatric conditions) can be analysed in this way. Their positive, neutral or negative fitness effect define their status as adaptation, passenger or disruption, and the specific processes affecting their fitness are defined as functional, neutral or dysfunctional.

The FAFF's distinction between fitness and function serves an important role in making analyses more scientifically tractable. The empirical programs required to show biological objects' effects on reproductive success and fitness are relatively straightforward, because they rely on a fundamentally *quantifiable* variable, related to an organism's success in producing and supporting offspring. Two sexually reproducing organisms can become four, or ten, or one, or zero. Investigation and agreement on whether a trait has a negative, neutral or positive effect on fitness is more attainable than describing exactly *why* reproductive success was gained or lost – those effects fundamentally revolve around qualitative description. We must agree on whether the process involved in organism reproduction fits a statement such as "eyes are functioning to see", "fear is functioning to protect" or "depression is functioning to disengage". Agreeing on function descriptions is uncontroversial for canonical examples such as eyes functioning to see, but more controversial examples such as depression can be much more challenging. If an interlocutor argues that the function description 'depression is functioning to disengage' is not appropriate, the empirical program

to decide the correct assessment is complicated, even in modern, directly observable environments, let alone historically. What exactly do you mean by disengage – how should we operationalise it in observing behaviour or designing questionnaires?[9] How does disengagement lead to improved survival and reproduction? What about cases where depression doesn't cause disengagement in that sense? What about other forms of disengagement which don't involve depression? How does this align or compete with the alternative hypothesis of "depression functions to encourage rumination"? Compare the complexity of the necessary scientific program in sorting out this debate to one which seeks to identify quantitative fitness effects of biological objects – if depression is observed or inferred to increase or decrease fitness, its status as adaptation or disruption is confirmed, without needing to precisely describe the functional or dysfunctional processes involved.

Application of Wakefield's HDA revolved around identifying the original functions ('selected effects') of a trait and observing their interruption. Depression is functional if acting in a way which was fitness-enhancing in the past, dysfunctional if it is interrupting systems from performing their ancestral function. Application of the FAFF can instead revolve around evidence of fitness effects, with specification of exact processes of function or dysfunction *possible* to describe, but not *essential* to categorisation. Depression is an adaptation wherever it increases fitness, a disruption where reducing fitness; if it increases fitness by encouraging disengagement from conflict, that is a function – if it decreases fitness by causing disengagement from relationships, that is a dysfunction. Importantly, if we observe or infer the fitness effects of depression *without fully understanding them*, we can still classify it as an adaptation or disruption, and know that functional or dysfunctional processes are in play – there must be *some* functional or dysfunctional process explaining the fitness effects. Thus, the central debate moves from being qualitative to being quantitative, with very different evidence admissible – where the HDA *needed* to describe the functional processes being interrupted, which requires some complex qualitative story, the FAFF only needs to appeal to evidence of reproductive success – which may be inferred from selection on genes, facts of epidemiology or demographic characteristics, or directly measured in modern reproductive success. When defining depression with a broad description specifier, it is common, early onset, predictably environmentally reactive, complex and visible across human societies and other taxa. This could be taken as evidence that it increased reproductive success and was an adaptation, with some sort of functional effects over extended framings of recent hundreds of thousands of years. If we changed the framing specifier of the analysis to concentrate specifically on industrialised societies, and measured it as associated with lower fitness, the same depression would be defined as a disruption. Nuances, justifications and limitations of this definitional framework follow below, but a key initial justification of the FAFF to note is

[9] This operationalization is a key place for subjectivity and disagreement to creep in. Behaviour needs to be parsed into quantifiable elements for analysis – drawing a line between disengaging and non-disengaging – but behaviour is (mostly) not so naturally quantifiable. This contrasts to fitness.

the circumvention of the need to engage in qualitative storytelling as to the exact functional or dysfunctional processes involved.

5/ Investigating fitness to discern adaptations, passengers and disruptions

Within the FAFF, the empirical program to define a biological object as adaptation, passenger or disruption requires, essentially, identification of the object or trait (using the description specifier) and measurement or inference of the reproductive success of organisms in its presence or absence. Crucially, the deciding factor is in principle quantifiable and exclusive – biological objects can only classify as adaptation, passenger or disruption, with a strict dividing line between positive, negative and neutral effects. Fitness can be observed directly by studying living organisms' inclusive fitness, or historical fitness can be inferred.

This relies upon quantifying individual or absolute fitness, rather than relative fitness (Orr 2009). Relative fitness of competing genotypes and phenotypes in a population is the more common measure in evolutionary genetics and biology (Reeve and Sherman 1993), providing insight into how populations evolve, but isn't appropriate for applying definitions useful in distinguishing health and disorder, for various reasons. Firstly, adaptations can be outcompeted without becoming 'dysfunctional' or being disorders; if a more effective eye arises, the other eyes are not suddenly dysfunctional. Secondly, adaptations and disruptions should maintain their status once reaching fixation, despite no longer affecting relative fitness; eyes are adaptations and cancerous cells are disruptions even though they occur in every individual. Adaptation and disruption (and health and disorder) are features of the individual organism which affect its pursual of fitness, regardless of the comparable population – if an individual enters a state of depression, the facts about whether that depression is functional or dysfunctional relate to the facts about whether that individual did better or worse due to the depression. The individual fitness effect provides what biological objects 'should' be doing.

Quantifying fitness results in a continuous variable, so utilising this clearly demarcated trio of adaptation/passenger/disruption still allows biological objects different degrees of positive or negative fitness: traits can be confirmed as 'a better adaptation than X' or 'a worse disruption than X' by comparing their effects on reproductive success in relation to X. We can seek evidence that eyes are more functional than hands, which are more functional than feet; or that schizophrenia is less of a disruption than Down's syndrome – experimental setups testing these hypotheses require measuring or inferring the effects of eyes, hands, feet, schizophrenia and Down's syndrome on fitness, and comparing them.

Obvious bodily adaptations are eyes, organs and lactase persistence (in certain populations; discussed in §7.2); obvious psychological adaptations are hunger, sexuality and language. Scientific confirmation of their fitness benefits can be directly observed by comparison to individuals lacking the traits, or inferred from various biological features such as universal prevalence, complex and positively selected genetics and mechanisms, early and complexly integrated developmental history, and any other signs expected from adaptations (although

by-products must be distinguished; see §5.1). Bodily and psychological passengers are any objects which apparently make no difference to fitness. Slight differences in bodily hair, certain taste and smell receptors and minor differences in personality would likely classify as passengers, with no discernible or inferable effects on reproductive success. Obvious disruptions include classic targets of medicine: cancer, wounds, toxins, infections, Down's syndrome and so on. Categorisation as disruption could involve direct observation of modern day fitness effects or inference from various sources (e.g. discovering causal de novo mutations). Most crucially, demarcation by fitness effects allows the parsing of complex cases (given sufficient evidence) with a single quantified metric, offering a tractable scientific research direction. Diagnosing disorder at a single point along a spectrum of low mood blending into depression becomes possible – wherever the state negatively effects fitness.

5.1/ By-products

Although the FAFF provides a simple quantifiable definition of adaptation, adaptations must cause fitness increases themselves, which makes analysis more complicated. Associating reproductive success with a trait is not sufficient to prove that trait is an adaptation, because reproductive success can be random, caused by genetic or population effects, or be observed in by-products which are not themselves adaptations – biological objects can be reproductively successful without themselves causing the reproductive success. Major depression may have been reproductively successful because low mood and mild depression are adaptive, but bring a vulnerability to major depression as an occasional by-product. Evolutionary medicine and psychiatry recognises various by-products of adaptations causing disorder. This may come via costly trade-offs; mismatch to novel environments; extremes of adaptive spectrums; maladaptive outcomes of developmentally plastic adaptive systems; pleiotropic effects over the lifespan; sexually antagonistic pleiotropy; heterozygote advantages; and more (Abed and St John-Smith 2022).

Distinguishing non-adaptive alleles reproductively successful by chance is a task for specific methods in genetics (e.g. identifying genetic hitchhiking (Schlötterer 2003); or population effects (Peter and Slatkin 2013)). By-products of adaptive alleles which gain fitness via related adaptations without lending positive fitness themselves are more conceptually and scientifically complicated to distinguish. They (by definition) are not functional in themselves (although they may be necessary to functions e.g. metabolic expenditure on cognition), and may be neutral or dysfunctional in themselves. They may be inevitable and permanent features of adaptations (e.g. calorie consumption) which are thus consistently associated with identical fitness as the adaptation. In such cases, they can only be differentiated from adaptations by analysing function and selected effects (e.g. recognising that the calorie consumption of speaking is not functional in itself; if speaking didn't require calories, it would be equally functional), but cannot be directly experimentally distinguished (we cannot ever test the fitness effects of calorie-free speaking). These are considered further in section 6.3. On other occasions (and especially in cases likely to be diagnosed as disorder) by-products are not inevitable and permanent but occasional, occurring in specific individuals or periods

during an individual's life – perhaps the case for major depression. These cases can have different measurable fitness than the related adaptation, but are ultimately caused by the same alleles, which are adaptations at a population level. Pointing out their existence is important, because they are causally linked to adaptations and are complex cases for Wakefield's HDA. They can be dealt with as distinct cases of dysfunction with description specifiers (§7.1), again integrated into a novel taxonomy going beyond the function/dysfunction dichotomy.

6/ Investigating function to discern function, neutrality and dysfunction

There is an inextricable relation between fitness and function. Differential fitness of biological objects *necessarily* occurs because of selection processes, so adaptations and function or disruptions and dysfunction theoretically co-exist *by necessity* (e.g. eyes are adaptations with functions of seeing; if they didn't have a function, they wouldn't be an adaptation). All adaptations have functions (processes which increase their reproductive success) and all disruptions have dysfunctions (processes which decrease their reproductive success). However, quantifying a trait's fitness is a distinct scientific problem from knowing *why* it affects fitness. Ascertaining fitness is relatively simple; ascertaining effects of function or dysfunction is not.

A fundamental issue is that whilst scientific assessments of fitness are quantifiable, assessments of selection effects are stated qualitatively. Qualitative scientific statements describing function have great intuitive appeal and commonplace application; being able to say "eyes evolved to see; seeing is their function" is useful and desirable, a quintessential aim and example of evolutionary explanation (Williams 1966, p6), and must be true to some extent, but is more scientifically problematic to define and test than "eyes increase fitness". Human eyes also contain white irises, which didn't evolve to see, but likely for some social signalling function (Kano, Kawaguchi, and Hanling 2022). Part of the eye's function is blinking and self-cleaning. The function of seeing is itself multifaceted (e.g. identifying ripe food; potential mates; dangers; grounding circadian rhythms, and much more) and hard to neatly describe in its relationship to attaining reproductive success. The complexity of cases beyond paradigmatic 'seeing eyes' balloons substantially. We could perhaps quantify observer agreement on whether a particular case fits a functional or dysfunctional description, or try to agree on the proportion of cases of e.g. 'depression' which fit such a description, but this is a far cry from the strict demarcation desired in distinguishing complex cases of health and disorder.

Here I illustrate difficulties ascertaining function, especially in comparison to ascertaining fitness, making points in line with several of the criticisms made of Wakefield's HDA (Faucher and Forest 2021). I then go on to show how function and dysfunction *can* be meaningfully attributed and used in Wakefield's HDA if the FAFF is adopted.

6.1/ Difficulties in discerning function

Imagine reaching the end of your life and being asked by an inquisitive evolutionary researcher to describe how your hands functioned to achieve your fitness. The answer is overwhelmingly complex. It essentially requires your life story, recounting endless details of hand-use – many of which are so trivial and multifaceted you won't be able to remember or explain them. Even perfect reporting here wouldn't suffice, because you don't know the counter-factual of your handless life. Compare this daunting task (needing to be repeated across a population before being summarised into a description of general hand-function) to the question of whether hands were positive, negative or neutral for your fitness. On this, you could pretty clearly conclude they were positive; and at a population level, this becomes a fairly simple question to investigate. The different complexities of these questions (hard for hands; harder for psychological traits) gets to the crux of why fitness and function must be conceptually separated to make the HDA workable, and is blatant in any number of examples.

Take lactase persistence into adulthood, an adaptation prominent in Northern European and other specific human populations, allowing the consumption of milk beyond infancy, but not seen in the majority of living humans or other mammals. Genetic evidence indicates lactase persistence has evolved multiple times in different pastoralist populations (Ségurel and Bon 2017). Evidence of its status as an adaptation in these populations is fairly incontrovertible because of evidence of positive selection for the associated alleles (Ranciaro et al. 2014). Despite comfortably being categorisable as an adaptation, and being fairly biologically simple (the pre-existing enzyme is produced persistently rather than 'turned off' after infancy), describing the precise function or selected effects of lactase persistence is surprisingly hard (Ségurel and Bon 2017). Although undoubtedly associated with pastoralism and milk and milk-product consumption, potential functions vary. Possibilities include helping make animal milk safe to drink for its water content; helping to avoid harmful symptoms of lactase intolerance; or increasing the calorific and nutritional content of milk (or both to some degree). These effects may have been functional at different times between different populations. In some cases, lactose digestion may have supported mothers breast-feeding their children; in others it could have allowed travel over long distances with scarce food and water; or provided an alternative food source in times of famine and disease; or allowed fattening up to survive harsh winters; or increased muscle mass, which in turn was useful in aggressive interactions – or tool creation – or hunting. Only recently have these possibilities been narrowed down, in a study finding short bursts of positive selection coinciding with periods of population decline – but the exact reason lactase-persistence enabled higher survival rates during those periods can still be debated (Evershed et al. 2022). In comparison to the fairly straightforward analysis allowing us to classify lactase persistence as an adaptation in certain populations, clarifying its exact function is much more complicated, perhaps impossible, depending on how much we can ever know about the evolutionary history of the relevant populations.

For an example of complex dysfunction, take schizophrenia. Measuring schizophrenic fitness in modern populations, it reduces reproductive success in comparison to non-affected

individuals (Bundy, Stahl, and MacCabe 2011; Haukka, Suvisaari, and Lönnqvist 2003) and thus classifies as a disruption. Comparing fertility in schizophrenics and non-schizophrenics results in this quantifiable fact. By definition within the FAFF (assuming causality, which is likely justified in this case), we know dysfunctional processes are at play – selection effects are evident in the fitness differences. But exactly what dysfunctional processes are causing schizophrenics' reduced fitness? Reduced lifespan? Increased drug use? Stigmatisation? Infertility? Are so-called positive or negative symptoms more important? Delusions of grandeur or paranoia? Is the dysfunction primarily caused by social reactions to schizophrenics or regardless of social context? Investigating the effect of each of these plausible dysfunctional processes is not impossible (Bowie et al. 2010; Rabinowitz et al. 2012), but is highly complex – and results will likely differ between individuals, families, societies, and generations, so the goal of qualitatively describing the dysfunctional processes of schizophrenia is highly intimidating.

Thornton (2021) nicely summarised these sort of problems in his criticism of Wakefield regarding the indeterminacy of function. A classic example of indeterminacy of function considers whether the heart's function is to pump blood or pass oxygen around the body. The answer, of course, is both – but depending on how the function is stated, slightly different criteria for dysfunction arise, as the selected effects which can be interrupted are slightly different – and the function of the heart can, of course, be stated in many more ways. Wakefield recognises this complexity as highly problematic (Wakefield, 2021l). For his definition of dysfunction to work, a selected effect needs to be described, and then shown to be interrupted. However, because function and selected effects are described qualitatively and part of complex biological life, they are extremely hard to specify exactly and completely. Sufficiently describing the function of hands or hearts so that interruptions of that function can be reliably observed and classified as dysfunction without counter-example is hard-going-on-impossible. The fact these functions are mostly qualitatively described semantically is only one source of imprecision. Scientifically, such selected effects are extremely hard to parse out and study, as seen in the lactase tolerance and schizophrenia examples; and there may not even be a single true answer, given the fact that the processes which lifeforms engage in, both in biotic and abiotic interactions, are highly idiosyncratic. Every individual organism lives a technically unique life because of its unique position in spacetime (considered further in section 7.1). The precise physical process through which it achieves fitness will be unique in turn. Finding a suitably general statement of a trait's function which can then be used to derive a trustworthy strict criteria for its dysfunction is complex to the point of implausibility. Beyond essentially tautological statements like 'the function of hands is to function as hands', or 'eyes function to see' (whilst the definition of eyes is organs which see) complete description of function seems impossible.

Difficulty specifying function is not simply a failure of our analysis, but can result from adaptive biological features. Genotypes and phenotypes which are highly flexible and plastically alter depending on the environment are often preferable to traits which are fixed

(West-Eberhard 2003), especially in behavioural traits which should alter to optimise around local ecologies, but also in bodily features which differ in growth depending on resources and stressors. Not only do biological objects alter in specific function over generations and between individuals, they *should* alter to maximise fitness given their unique environment. If you could specify a precise function of eyes, hands or low mood, they would be less useful, because they would be less flexible. Stating plasticity as itself a functional effect is possible, but adds another layer of complexity to asserting functional stories in a sufficiently specific manner.

Intuitive examples of 'eyes seeing' and 'lungs respirating' make precise stipulation of selected effects seem far more tractable than in reality. Under philosophical scrutiny or scientific investigation, functional effects are hard to pin down in sufficient detail. Statements such as "depression causes useful disengaging" or "schizophrenia causes dysfunctional socialisation" may be true, but only if they are not supposed to be exclusive and complete descriptions of processes affecting the fitness of depression and schizophrenia, applicable in every instance, individual and generation. The problem this causes for the HDA is that it *is* supposed to be universally applicable – we should be able to take any given biological trait and say whether it is functional or dysfunctional, at least in principle. When the definition of dysfunction involves interrupted selected effects, the path to making those judgements is unclear, because functions are hard to specify with sufficient precision.

6.2/ Discerning dysfunction within the FAFF

The discussion above emphasises difficulties ascribing function – presenting problems for the HDA. However, whilst Wakefield's specific definition of dysfunction as interruption of selected effects is critically affected if unable to precisely describe selected effects, the FAFF can sidestep this problem. The core advantage the FAFF has over Wakefield's definition of dysfunction is in allowing the attribution of dysfunctional *status* without specifying dysfunctional *effects*.

As noted, disruptions *must* involve dysfunction – *some* process is causing the fitness reduction. It is describing dysfunctional processes qualitatively that is problematic. However, we know schizophrenia *is* dysfunctional – involves dysfunctional processes – *without specifying them*, simply citing evidence of its effect in reducing fitness. Wakefield's definition of dysfunction requires such effects to be specified, but within the FAFF, the specific effects are tertiary. "There is dysfunction – we know not what" becomes a meaningful statement, so *designation as disorder can avoid the problems of describing functional effects*. The simplicity and tractability of measuring fitness is ported into the HDA.

This might feel like it's missing the point – surely the reason we can define non-pumping hearts or non-seeing eyes as dysfunctional is precisely *because* the function of hearts is to pump blood and the function of eyes is to see; stating fitness alone misses mentioning dysfunctional biology. Granting this point, note that the possibility to ascribe functional status via fitness measurements does not preclude the possibility for attributing that dysfunction to

particular effects. In asked why un-pumping hearts or blindness are dysfunctional, the FAFF allows the very same qualitative statements describing dysfunctional effects that Wakefield referred to, except the FAFF also resorts back to ultimate justification via fitness effects. The dysfunctional process of blindness is caused by the effects of interrupting seeing, and everything that involves, and those effects are dysfunctional because they reduce fitness. There is a relatively negligible difference here from a Wakefield rendition of 'blindness is dysfunctional because it interrupts seeing, and seeing is a selected effect'.

Benefits of ultimately justifying dysfunction with fitness effects become particularly evident in defining edge cases. Instead of outright blindness, imagine a thousand individuals with different forms of steadily worsening eyesight and ask where the dysfunction exists. In Wakefield's definition of dysfunction, this is a hard task – all eyesight below perfection seems to be interrupting the selected effect of seeing – so are essentially all eyes dysfunctional? In the FAFF, poor eyesight is still an adaptation (some eyesight is better than none; this is the importance of specifying description, see §7.1) and the various physical traits which worsen eyesight (e.g. cataracts, various degrees of myopia) become disruptions only when they have a measurably negative effect on fitness – likely to start at some much more intuitive point on the worsening scale. The qualitative description of the dysfunctional process of 'interrupting seeing' is still applicable, but finds more sensible results, not having to define all except the most healthy eyes as dysfunctional (the same reason relative fitness is an inappropriate measure for disorder). In cases of psychological spectrums where low mood and depressive tendencies blend into major depression, or autism spectrum disorder becomes more debilitating, such an analysis will similarly result in the drawing of a line which somewhat intuitively fits the definition of 'disorder' – the trait has reached a level of harm where it is interrupting the individual's reproductive goals.

Paradigmatic diseases negatively affect fitness and thus categorise as disruptions within the FAFF; their dysfunction visible in their myriad of processes harming fitness. More interesting, and potentially problematic, are cases where biological ailments and suffering exist without clear pathological cause, and perhaps without negative effects on fitness. Headaches, fatigue, muscle aches and back pain may fall into this category. If such conditions don't negatively affect fitness[10] we can't label them as disruptions. Wakefield's definition was more amenable to classifying such ailments as dysfunctional – we can plausibly say that headaches are interrupting the selected effect of thinking clearly; that fatigue interrupts the selected effect of acting energetically, and so on. This is possible precisely because the qualitative nature of describing function is more flexibly applied than fitness measurement, allowing functions and their interruption to be widely ascribed – however, as a consequence of this flexibility we could equally argue that sexual desire interrupts the selected effect of thinking clearly, and tiredness interrupts the selected effect of acting energetically, and thus both are dysfunctional. This conundrum is widely problematic for Wakefield's definition of dysfunction

[10] Although they might, depending on how they are described (see §7.1) – if the description strictly includes the most severe cases, they are likely to be correlated with reduced fitness.

– many adaptations or normal responses interrupt selected effects of other sorts, and thus we can portray all sorts of apparently normal or healthy phenotypes as interrupting selected effects. In cases such as sexual desire such ascriptions of interrupting selected effects are intuitively absurd (although social values may cause disagreement – a particularly pious religious individual may make all sorts of lengthy descriptions of the selected effects which sexual desire interrupts). In controversial cases such as low mood or autism, this problem is critical – interlocutors can easily argue that they interrupt selected effects of socialising and behaving optimally, so are dysfunctional, even when not particularly harmful and potentially beneficial ancestrally. Relying on quantifiable fitness effects prevents this problem.

6.3/ Functional and dysfunctional by-products and trade-offs

There is one nuance to the FAFF which allows ascription of dysfunction terminology in a similarly flexible manner to Wakefield, where biological objects are associated with apparently opposite effect processes i.e. where clear disruptions cause functions (e.g. blind individuals 'compensating' with improved senses in other areas) or clear adaptations cause dysfunction (e.g. seeing with eyes in wanton procrastination). Like by-product objects (§5.1), these by-product selection processes are complicated cases needing consideration. Due to emphasis here on defining disorder, dysfunctional by-products are most important to consider.

Dysfunctions in the FAFF are negatively fitness-affecting *processes* rather than the fitness result itself, and such processes can occur without associating with overall negative fitness when counter-acted by other fitness-enhancing processes. Wherever we see an adaptation associated with an apparent negative effect on its overall reproductive success (e.g. highly calorifically costly brain maintenance) yet without an overall negative effect (human brain size and processes are an adaptation and not a disruption), it is likely a dysfunctional by-product. If the dysfunctional by-product were able to be removed or reduced without affecting the trait, the organism would have higher fitness. Adaptations are never perfect; they always include trade-offs and costs. Those trade-offs and costs are bad for the organism, negatively effecting its fitness – but because they are constantly outweighed by associated positive, functional aspects, they will still measurably associate with higher fitness and be classified as adaptations.

Various disorders which evolutionary medicine and psychiatry recognise (Nesse 2005) may fall into this category. These should clearly retain categorical distinction from true disruptions. Note, however, that dysfunctional by-products can be widely ascribed, arguably too widely to be useful for strictly ascribing disorder. Every biological phenotype has dysfunctional by-products, and is involved in processes which in themselves are harmful to fitness. Most obviously, consuming energy or being susceptible to damage or somatic mutation are necessary trade-offs of every adaptation. The advantage of labelling them as dysfunctional by-products is recognising these biological costs as costly, especially for occasions of harm which medicine might prevent – but distinguishing these from dysfunctional disruptions is

necessary. Wakefield's definition of dysfunction allowed them equal status, ignoring this important heterogeneity by relying on the oversimplified dichotomy of function/dysfunction. These trade-offs form another class within the more complete taxonomy of dysfunction (§8.2).

7/ Necessary Specifiers

The FAFF as a definitional framework covers the three possible fitness and function states of biology, but for categorisation into these definitions, two additional considerations are persistently required. We can call these *specifiers*. The first specifier involves clearly identifying the trait to be studied and categorised. Fitness and function must be measured *of* something. We can call this *description*. The second specifier clearly identifies a reference environment. Reproductive success only occurs due to fitness within a particular environment (including the population in that environment). We can call this *framing*. Both of these specifiers are not just desirable but genuinely necessary for any sound application of evolution-based definitions. Evolutionary classification of depression is impossible without specifying what 'depression' means and in which environment it is being assessed. When these specifiers are ignored or left implicit, the clarity of the resulting analysis suffers, especially in cases where alterations of description and framing alter facts of fitness and function: changing the definition of depression or the environment of analysis could easily change facts about whether it increases or decreases fitness.

Importantly, in this altering of description and framing specifiers, cases of disruption and dysfunction can be observed to arise due to adaptations. Their recognition allows common targets of evolutionary medicine and psychiatry (e.g. mismatches and trade-offs) to be separately categorised as more complex forms of dysfunction than e.g. cancers and infections, with a level of specificity impossible in the HDA or with a simple function/dysfunction dichotomy.

7.1/ Specifying Description

To study anything, it must first be identified. Problems describing specific effects of function and dysfunction aside, even identifying a set of biological objects to ascribe fitness to is a non-trivial task (Zachar and Kendler 2017). Measuring the success with which depressed individuals reproduce is easier than describing the selection processes on depression – but still requires you to begin by defining depression. To do this, token instances must be identified and categorised together. This process of drawing lines for classifying biological objects together is the process of specifying description.

The complexity of physical biology makes it technically unique in every instance, which means every description specifier will contain somewhat heterogenous manifestations of what is ostensibly the same thing. Approximations in description specifiers are justified – although each cytosine will have minute physical differences at atomic and sub-atomic levels, they are

comfortably assumed to have similar enough functional[11] properties to be categorisable as one of four discrete human nucleobases. Equating sequences of genetic code in ancestors with 'identical' sequences in descendants is justified, because their effect is assumed to be essentially identical; we could extend this argument to red blood cells, hearts, and so on (with varying validity). The only way to situate biological objects within an evolutionary history of selection with fitness over multiple generations is to equate them physically; although replication and reproduction of biological genotypes and phenotypes isn't perfect, it can be functionally equivalent. This heterogeneity is a necessary limitation to biological analysis.

Decisions in specifying description follow theoretical goals (e.g. in the HDA the goal is classifying disorder) in combination with scientific practicality (cost, time and technological restrictions exist). Psychological traits such as 'anger' or 'depression' are less reliably defined than many biological objects, with most modern psychological studies relying on single questionnaires (which may differ between studies) and Likert scales (Norman 2010), often with arbitrary cut off points on continuous distributions to create categorical types – particularly where the goal is to identify a type requiring medical treatment. The evolutionary relevance of these labels have not been a concern for mainstream psychiatry; the primary aim is clustering together individuals needing help who broadly share symptomology, and may respond to the same treatment. To further this aim, psychiatric diagnostic manuals often apply some criterion for only capturing debilitating cases. For example, beyond core symptoms of inattention, hyperactivity and impulsivity, a DSM-5 ADHD diagnosis requires "clear evidence that the symptoms interfere with, or reduce the quality of, social, academic, or occupational functioning" (American Psychiatric Association 2013, p60). Similarly, along with five of nine possible core symptoms of depression, a major depressive disorder diagnosis requires that "the symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning" (American Psychiatric Association 2013, p161). These criteria make sense for justifying medical intervention, but also mean that individuals with essentially identical phenotypes (and potentially genotypes) can differ in diagnosis purely due to functional fit to their particular environment. In practise today, different descriptions of depression are applied by pharmaceutical companies researching treatment or psychiatrists seeing patients (Bech 2022); and standards differ between individuals, organisations and nations (Simon et al. 2002). These differences in description specifiers are justified practically, but altering description specifiers can also serve a useful purpose in understanding evolutionary cause and discerning different types of dysfunctional processes – particularly in identifying by-products.

Essentially all common psychiatric conditions are genetically and phenotypically related to subclinical variation presenting with similar traits to diagnosable disorders (Kotov et al. 2017). Description specifiers designed for clinical application exclude this subclinical variation, despite likely shared cause (Hunt and Jaeggi 2022). For example, Wakefield references

[11] Note that here the 'causal role' definition of function of Cummins (1975) and others is appropriate, rather than backwards-looking approaches.

findings that certain alleles associated with autism have been under positive selection (Wakefield, 2021h). In specifying that we are describing and assessing these alleles alone, we categorise them as adaptations because of their positive effects on fitness (presuming their positive selection is not due to other genetic effects such as linkage disequilibrium). However, when changing description specifier to assess the individuals who classify as autistic because of those alleles, reduced reproductive success could be found – the same alleles may have functional cognitive benefits in certain individuals (Crespi 2016) but cause dysfunctional cognition in others. Thus, in analysing autism-related phenomena with different descriptive specifiers we can find both adaptation (the alleles) and disruption (instances of autism which those alleles cause as a harmful phenotype). Importantly, in recognising relationships between different described biological objects (especially in cases of shared genetics and overall fitness of causal alleles) a picture can emerge accounting for evolutionary persistence – findings of functional and dysfunctional manifestations of autism are expected in 'balancing selection', where combined positive effects and negative effects of alleles maintain variation in the population (Hunt and Jaeggi 2022). Concentrating on individuals with diagnosable disorder may miss the relevant explanatory picture (Figure 1, case A). This is a sort of 'distributed disruption' – reduced fitness may be caused by a trait which is evolutionarily explained by related traits which confer increased fitness. Alleles which confer advantages to heterozygotes but disease to homozygotes are simple examples, such as in malaria protection and sickle cell disease. These forms of dysfunction are by-products of adaptive processes and not simple disruptions. Altering description specifiers appropriately allows them to come into view.

Initial descriptions of targets of analysis may also capture harmful trade-offs without noticing related adaptations, a slightly different problem which altering description specifiers may solve. For example, definitions of autism are negatively framed and concentrate on social and behavioural difficulties, whilst ignoring associated cognitive strengths which may explain the persistence of autism in the population as a by-product of selection for high-systemising individuals (Baron-Cohen 2020). Similar dynamics might be seen from within the developing field of network approaches to clinical psychology (Borsboom, Cramer, and Kalis 2019; Fried et al. 2017), considering systems of psychological interactions perpetuating mental disorder. Individual elements in a network of interacting psychological modules could be adaptations which create a system with vulnerability to a clinically diagnosable state. So long as the overall selection on the adaptive components is still neutral or positive, natural selection can allow dysfunctional trade-offs to persist. For example, it is possible that the stress-response system, learning system, mood regulation system and steroidal system (however appropriately described) are adaptations, generally conferring fitness benefits, but in certain instances or environments are the primary systems interacting in networks maintaining a state of highly debilitating major depression. Recognising this by-product trade-off effect lends clearer understanding of the situation, and could lead to more satisfying evolutionary accounts (again, perhaps by explaining why such depression states are common and persist despite being harmful (Keller and Miller 2006)). When we propose a description of major depression

(or any psychiatric condition), we could be ignoring the full picture of the relevant psychological systems (Figure 1, Case B). Another simple example from bodily medicine is in fever – concentrating on the negative experiential aspects of fever misses the positive effects on fighting infection. In these cases of trade-offs, altering description specifiers to recognise linked traits with functional effects balancing the dysfunctional effects is necessary to appreciate that the initial trait being considered is not a simple disruption, but a by-product of adaptation.

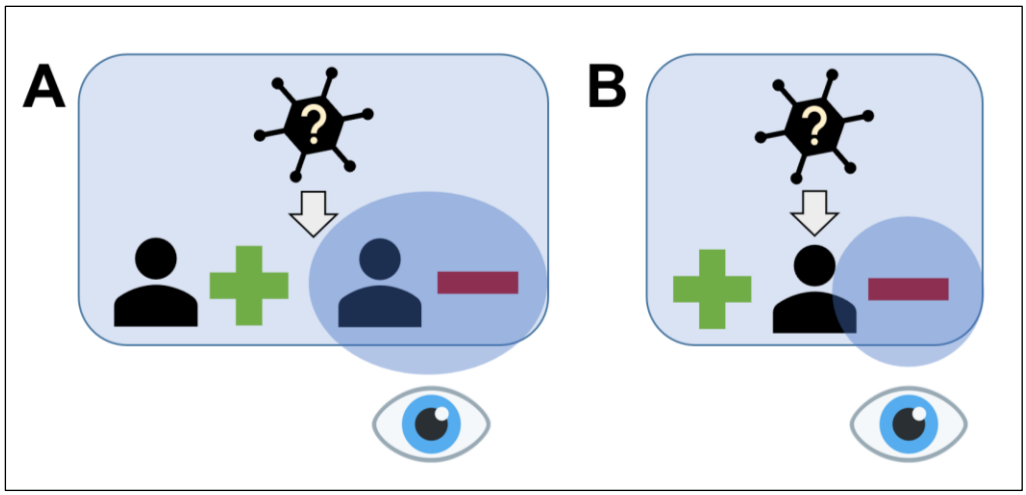

Figure 1: Description specifiers can identify dysfunctional by-products of adaptations. In case A, a trait or allele is fitness enhancing in one individual but fitness reducing in another individual, and the description specifies the disruption. In case B, the description specifier identifies the dysfunctional trade-offs of adaptation.

Essentially all described biological systems are both part of larger networks and contain smaller components (this is an observation of system theory, which inspired George Engel's (1977) biopsychosocial model of health, and also justifies an 'evobiopsychosocial' approach (Hunt, St-John Smith, and Abed 2023)). Importantly, when altering, expanding or narrowing descriptions ('drawing lines' differently; zooming in or out to different scales), the facts of fitness effects and functional or dysfunctional processes can change. In doing so, we can discover disruption, passenger and adaptation components of systems which are constituted by and constituents of differently categorizable systems in interesting ways. When we zoom out from sickle cell disease and see heterozygote malaria protection, we are changing our descriptive specifier to better understand sickle cell disease. Cases visible via altering description specifiers are more complex cases for ascribing dysfunction to justify disorder attribution – they may be explained by adaptive processes, not simple disruptions. Whether by-products of adaptations are dysfunctional in the sense implied by 'disorder' may depend on different intuitions or between cases. This also complexifies Nesse's warning against 'viewing disease as adaptation' (Nesse 2023) – it is possible to view dysfunction in trade-offs and particular individuals despite adaptation being the ultimate cause of the viewed traits. The function/dysfunction dichotomy clearly can't capture the appropriate complexity here.

The value of forcing description specifiers is in clarifying exact targets of analysis; the value of altering description specifiers comes from enabling recognition of these complex cases.[12]

7.2/ Specifying Framing

Once a trait is clearly described, its fitness can be measured and function assessed, but asking the fitness of e.g. depression is impossible without referencing for whom, when, and where. We need to specify an environment for 'framing' the analysis. Explicitly recognising framing adds further necessary nuance to definitions, because different framings, like different descriptions, can result in different diagnoses of fitness and function. This also allows the FAFF to be meaningfully applied by evolutionary researchers who care about either historical or contemporary adaptation (Reeve and Sherman 1993).

Theoretically integrating framing into the FAFF is fairly straightforward. Once describing something, be it a particular allele, enzyme, limb, brain system, psychiatric condition, and so on, it can be assessed for its fitness effects in certain environments. The requisite scientific methodology may be complicated or limited, especially if framing within historical environments, but in principle this allows quantitative differentiation of adaptations, passengers and disruptions. Functional, neutral and dysfunctional processes within the specified framing can be proposed. Categorisation must be recognised as relative to that framing. Lactase persistence is a useful example here, because it is so well understood: "lactase persistence is an adaptation" should be framed with "in certain pastoralist populations since 10kya" for it to be correct; lactase persistence was not an adaptation beforehand or in other, non-pastoralist populations. This is more precise than Sober's (2000, p85) account of 'adaptation' which simply stipulates that adaptations are products of ancestral positive selection. The FAFF framing specifier allows an important element of flexibility – for adaptations to become disruptions in different environments, where their presence begins to harm fitness, a recognised problem with Sober's account (Reeve and Sherman 1993). This is desirable when defining disorder naturalistically, because adaptations clearly do become biologically harmful (both for fitness and subjective suffering) in different environments – the halting of lactase production in adulthood was an adaptation, but became a disruption in pastoralist populations.[13] Similarly, dysfunctional trade-offs of adaptations may be exaggerated in different environments (Figure 2).

[12] Note that the genetic level of description is the only one at which clear-cut adaptations, passengers and disruptions can be reliably distinguished in a way representing fundamental evolutionary causation: at the phenotypic level, it is possible (and in cases of questioning disorder status, likely) for by-products to be described. In such cases, the by-product may be a disruption in itself, but caused by adaptive alleles. As the fundamental unit of selection, gene descriptions will never be by-products of another, perhaps oppositely FAFF defined process – although note non-adaptive genetic effects (e.g. linkage disequilibrium and hitchhiking) may make it difficult to assess the fitness effects of specific alleles. To some extent, this recognition encourages description specifiers to hone in on the gene-eyed view as the proper unit of selection (Dawkins 1978).

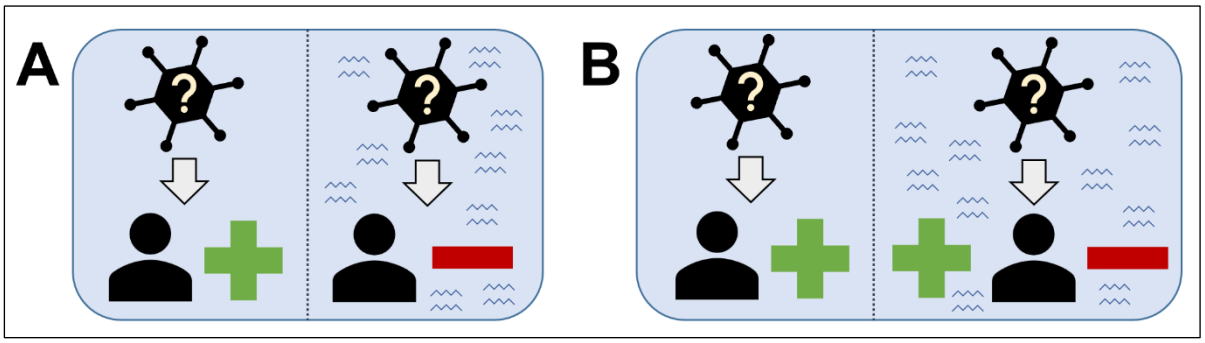

Figure 2: Framing specifiers can identify dysfunction resulting from adaptations changing environment. In case A, a trait or allele is fitness enhancing but becomes fitness reducing in another environment. In case B, a trait or allele is fitness enhancing but has additional dysfunctional trade-offs in another environment.

Note that, similarly to altering description specifiers, altering framing can provide useful information as to cause. Placed within the wide framing of the whole human population for the past two hundred thousand years, lactase persistence will be categorised as a disruption (we can infer, because it assumedly arose multiple times yet was selected out, until arising in specific groups of pastoralists). Reframed within specific pastoralist populations since 10kya, it is an adaptation. Reframing even more specifically within such populations could imply its function – as seen in recent genetic evidence of episodes of strong positive selection concurrent with periods of population decline, potentially implying a function of increasing survival in the face of famine or disease (Evershed et al. 2022).

Wakefield's reaction to the counter-examples provided by Garson (2021) of evolutionary mismatch was to redefine dysfunction as a failure to perform a selected effect 'in an ancestral environment with internal similarity' (Wakefield 2021j). This is a problematic move, not only because difficulties identifying functions are multiplied when necessarily assessed outside of modern environments, but because there is no such thing as a specific ancestral environment within which a trait evolved – every generation experiences evolutionary selection pressures. There can be generations of stronger, weaker and even alternating selecting pressures; and moments of inception, fixation or elimination of traits in populations; but none of these provides a single identifiable ancestral environment. What Wakefield seems to be trying to get at, however, is a definition of dysfunction impervious to cultural norm shifts. The addition of framing specifiers solves this problem more precisely.

Every biological object can be framed at any scale, from inception to present day. Framings focussing on extended evolutionary causation are those Wakefield thinks we should utilise for disorder attribution (Wakefield, 2021j); if ADHD or depression were once functional, they should not classify as a disorder. Framings for uniquely human traits would presumably attempt to infer fitness and function over the recent hundred thousand years, perhaps honing in on certain framings in particular populations and periods; assessing mammal-wide traits

[13] Note that description specifier of the phenotype matters here: lactase production is always an adaptation (for every individual, it adds reproductive success in comparison to having no lactase); lactase *persistence* past childhood is a disruption in most mammals and non-pastoralists (it reduces fitness by being excessively costly) but an adaptation in pastoralists.

may benefit from much longer framing timescales. However, this approach excludes cases of evolutionary mismatch from being attributed as dysfunctional, which is arguably too heavy-handed to align with modern conceptions of appropriate 'disorder' attribution (e.g. for dyslexia and dyscalculia). Framing specifiers can account for complexity in such attributions.

Take a classic example of evolutionary mismatch, fat-storing capacity leading to obesity (Power 2012). Individuals in developed societies with calorific excess can experience reduced fitness because of excessive fat storage (Best and Bhattacharya 2015), but shifting framing to pre-industrial societies with common calorific shortages, fat-storing capacity indubitably increases fitness. Depending on the specified framing, efficient fat-storage can be an adaptation, passenger or disruption. Indeed, in considering the effects of altering framing specifiers, we better understand the functional and dysfunctional effects of fat-storage.[14] Such analyses are more comprehensive than requiring situation within ancestral environments: instead of claiming that we cannot diagnose dysfunction in morbidly obese individuals because they are living in unusual environments, we can state that their fat-storing capacities are dysfunctional disruptions in modern societies, but adaptations in environments where available calorific content sits in a range representative of human evolution.

The additional precision of specifying framing (beyond being a genuine theoretical necessity) can be socially appealing, by alleviating the heavy-handed divisiveness which a simple function/dysfunction line encourages. Statements such as 'ADHD is dysfunctional and thus a disorder' or 'ADHD is not dysfunctional and thus not a disorder' are tempered by specifying framing – more nuanced analyses resulting in statements such as 'ADHD was not dysfunctional throughout most of human history but is in modern society' become coherent. Trying to apply a simple, single overarching binary categorisation as functional or dysfunctional can lead to backlash against unintuitive categorisations and arguments over cases which differ in status between environments. Recognising framing specifiers provides a clause of 'within these contexts', potentially serving to prevent disagreement over controversial cases, whilst avoiding value relativism. The only relativism is relativism of framing, a necessary consideration to make sense of fitness at all, and entirely naturalistic (even if social norms can affect fitness, and thus naturalistic diagnosis). This even somewhat aligns with modern advocacy perspectives encouraging a social model of disability (Oliver 2013) or an 'ecological' model of neurodiversity (Chapman 2021), claiming social factors are responsible for both pathologising natural differences and alleviating disability. Although such models generally deny the usefulness of naturalistic definitions of disability and dysfunction, specifying framing in the FAFF allows evolution-derived naturalistic definitions to account for the effects of altered social contexts in some limited extent, whilst avoiding vulnerabilities of pure value approaches.

[14] Note that we could also alter description specifiers to pick up morbid obesity alone, which would always be a disruption, but simply didn't exist (presumably) prior to modernity. However, this would not provide evolutionary insight into the cause of obesity, unlike altering framing specifiers.

8/ Taking the FAFF forward

Combining the FAFF definitional framework with description and framing specifiers, a tractable system for assessing dysfunctional status accounting for complex cases should become possible. One major issue remains, that evolutionary history is obscured by time – scientific assessment of the fitness of a given trait over a specific framing of human evolutionary history may be difficult. I echo Wakefield in saying that this is a matter for scientific advance, but reiterate that the FAFF requires more easily attainable evidence, and has clarified less challenging steps for dysfunction attribution: concentration on evidence implying fitness, with recognition of the complexity of stipulating function. Functions and selection change from generation to generation, environment to environment. Arguably almost every ancestral trait no longer functions for its selected effects, and every trait can be argued to interrupt some kind of selected effects. We should not expect to pin down an exact function, a perfectly consistent selected effect. Genes arise, phenotypes form, and individuals struggle their way to reproductive success whilst living unique lives, some more successfully than others, thus causing evolution by natural selection. The idea that there should be some specifiable, isolatable, empirically definable selected effect functioning in a particular body part or brain activity, and that we can precisely specify it, and then specify when it is not working, is a dream inspired by engineering and human design (Nesse and Stearns 2008), which does not match the messy reality of biological evolution. In requiring specification of selected effects, Wakefield's definition of dysfunction suffered from this often-unassailable complexity.

8.1/ Problems with emphasising fitness

One possible criticism of the FAFF is of overreliance on fitness values. Fertility, and thus fitness, can now apparently be almost totally disconnected from biological dysfunction. There is a trend of steady reduction in fertility rates in industrialising cultures, with total fertility of females now around 1.5 in Europe and Japan (Skakkebæk et al. 2021), in comparison to natural fertility of 5-6 (Kelly 2013, p193-200). Various odd consequences come from directly referencing fitness: couples who choose not to have children are dysfunctional; high education and income (associated with lower fertility) is dysfunctional within developed societies; societies where people are safer, richer, more hygienic, better educated and have better health on many metrics carry more dysfunction than societies where people are unsafe, poor, less hygienic, less educated and at much higher risk of many other diseases, yet retain natural fertility. Using contraceptives is to cause dysfunction. Such examples imply fertility and fitness are now so disconnected from health as to be inadequate defining factors for dysfunction.

The primary response to this criticism requires recognising the necessity of controlling for other effects on fitness; achieved fitness is an eventual result of many complex traits affecting an individual over their lifetime. *Where all else remains equal*, fitness effects of traits should still generally appropriately define biological dysfunction. The scientific task is to try and parse

out the fitness effects of the trait itself. First and foremost, this requires accounting for the effects of culture. Comparing *within* similar demographics, most counter-examples disappear – parasites harm fitness within traditional societies, even though those individuals have higher fitness than non-parasite infected Westerners. Certain counter-examples such as contraceptive pills will remain; and here, the FAFF defender can bite the bullet and recognise medical interventions can cause evolutionary dysfunction – and indeed, we often *want* to cause such dysfunction in accordance with social values.

Modern medicine and social structures often basically eliminate a predictable connection between biological disruption and negative effects on fitness. This is an interesting phenomena, but doesn't mean the FAFF is inadequate in principle. It might mean that we tend towards utilising a pre-modern-medicine framing for defining disorder, as Wakefield suggests.

8.2/ A taxonomy of evolutionary dysfunction

Considering the added complexity of the FAFF's landscape of dysfunction attribution is a final necessity. The FAFF and specifiers provide a framework for attributing dysfunction which is widely applicable (it can be applied to any biological trait or process) and exclusive and strict (no analysis can result in fitting into more than one category, because fitness effects can only be positive, neutral or negative). It also allows the attribution of dysfunction to complex cases which troubled Wakefield's dichotomy, where altered specifiers describe by-products (the criticisms of Kincaid and Murphy) or frame traits in environments which exacerbate dysfunction (the criticism of Garson). However, this necessitates going beyond the function/dysfunction dichotomy, and recognising distinct classes of dysfunction. Dysfunctional processes of disruptions (in every framing) are the most obvious and simple. Description specifiers can identify by-products of adaptations including dysfunctional trade-offs or occasional disruptions. Framing changes can elicit dysfunction and in some cases make adaptations or passengers into disruptions. Given these possibilities, we can identify five common classes of dysfunction in the FAFF[15]:

1. **Simple disruption**: A disruption's dysfunctional process.
2. **Distributed disruption**: An adaptation's by-product which is a disruption in another individual.
3. **Trade-off**: An adaptation's dysfunctional by-product in the same individual.
4. **Mismatch trade-off**: Changing framing increases dysfunctional by-products of adaptation.[16]

[15] Passengers are ignored in these definitions, but 'adaptation' could be replaced with 'passenger' and the same possibilities exist. These are much less likely, due to the requirement of perfect balancing of fitness effects.

[16] Trade-off dysfunctions via mismatch could potentially be given description specifiers which make them disruptions if specified closely enough, in which case they become dysfunction via mismatch e.g. if blue light sensitivity increases wakefulness but can lead to sleeplessness in a society with electronic screens, it would be

5. **Mismatch disruption**: Changed framing makes adaptation into disruption.

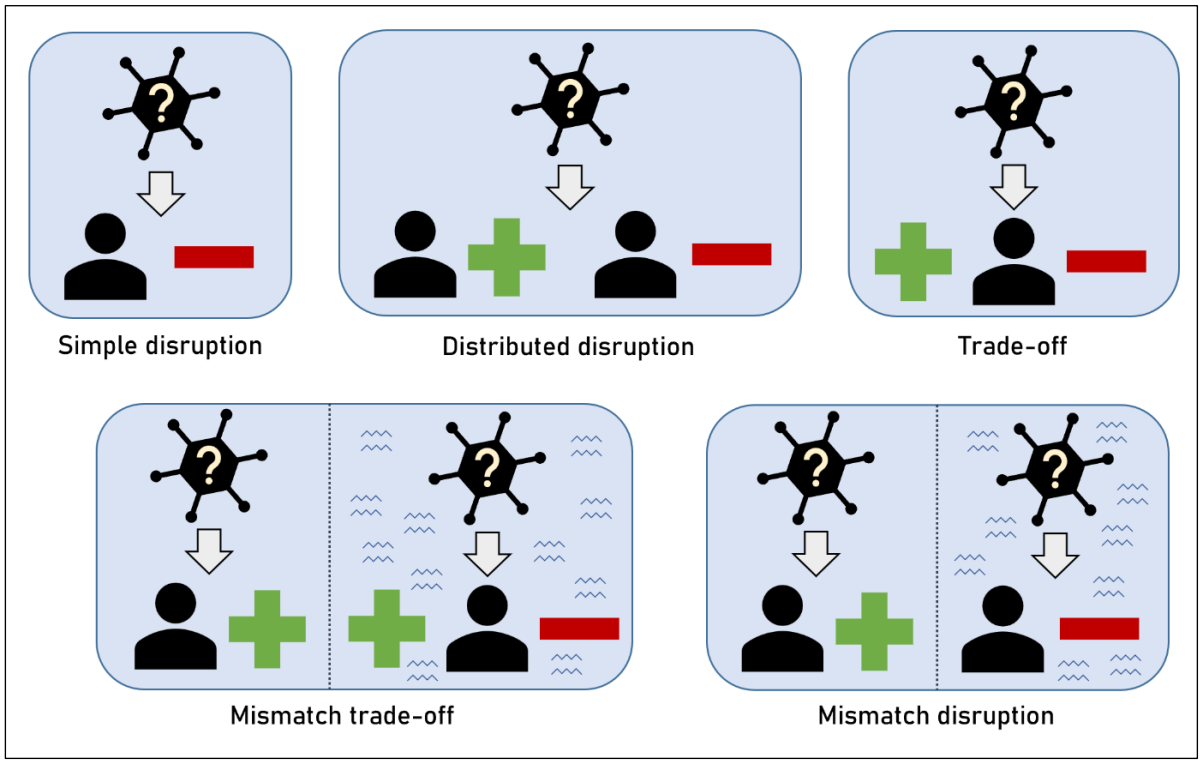

Figure 3: Five distinct classes of dysfunction in the FAFF given the effects of specifiers

Crucially, these classes are *quantifiably* distinguishable by effect on reproductive success, without stipulating processes: fitness is either generally reduced (simple disruption), reduced in confined cases (distributed disruption), positive but with negative by-products (trade-off), reduced due to environment but still positive (mismatch trade-off), or reduced to negative due to environment (mismatch disruption). Altering description and framing specifiers makes these complex cases visible, and with their application, the FAFF analysis can place any biological trait, including any existing disease or disorder, into these distinct classes (Table 1), assuming sufficient scientific investigation.

Class of dysfunction	**Likely examples**
Simple disruption	Cancer; broken bones; infection
Distributed disruption	Autoimmune diseases; birth complications and mortality; sickle cell disease
Mismatch disruption	Obesity; smoking-related illnesses; eating disorders
Mismatch trade-off	Blue-light sensitivity interrupting sleep; workaholism; myopia
Trade-off	Suffering associated with pain and fever; inappropriate instances of emotional reactions

Table 1: Possible examples of different classes of dysfunction. Fitness effects of each case are assumed.

possible to describe the sleeplessness itself, which would then classify as a disruption. It's useful to separate them for cases where the important biological trait is still clearly an adaptation but has gained additional dysfunctional processes due to mismatch.

This fundamentally quantitative definition means they theoretically provide comprehensive coverage of possibilities, strict demarcation, and exclusivity[17]. This contrasts to existing taxonomies of disorder currently available in evolutionary psychiatry, evolutionary medicine and the philosophy of biology. These are plentiful, and vary. Nesse's (2005) six reasons are mismatch, host-pathogen co-evolution, trade-offs, constraints, traits which increase fitness at the expense of health and defences. Al-Shawaf et al., (2020) additionally mention intragenomic conflict and adaptively biased mechanisms, and Abed and St John-Smith (2022b) also include life history factors, sexual selection and its consequences, balancing selection and heterozygote advantage, demographic history, deleterious alleles and extremes of adaptations. Matthewson and Griffiths (2017) suggest four biological criteria for disease, of mechanism failure, abnormal environments, inhospitable environments, and heuristic failures.

Such taxonomies contain broad categories of *processes* of dysfunction; they are filling the necessary void in qualitative explanation, just as statements such as 'eyes function to see' or 'depression functions to disengage'. However, they carry complexities of the indeterminacy of function (here, the indeterminacy of dysfunction), and are not mutually exclusive. For example, mismatched defences can have trade-offs (e.g. if social anxiety is overactivated by social media), and mechanism failures can result from abnormal and inhospitable environments. Mismatches of various types may occur (Bourrat and Griffiths 2021). There are no strict definitions of explanatory categories such as 'defence'. The FAFF's primary utilisation of quantifiable fitness leads to five classes of disorder within a *complete* and *strictly demarcated* framework; there will never be a biological object or process which doesn't fit into only one category of affecting fitness positively, neutrally or negatively; and description and framing specifiers allow the parsing of complexities by shifting focus of analysis or environment. This makes for a crisper (but less qualitatively informative) taxonomy of disorder, which is especially suitable for porting into the HDA, where a clear line must be drawn between health and disorder. Other taxonomies are useful for describing common processes of dysfunction, but inevitably imprecise.

When considering how the FAFF can be incorporated into the HDA, all four classes of dysfunction beyond 'simple dysfunction' are complex and interesting examples. Defying dichotomous categorisation, they make 'dysfunction' attribution necessarily complex. Wakefield's HDA demanded dysfunction to attribute disorder, but these cases aren't simply biology 'going wrong'. Most obviously, in cases associated with adaptation i.e. trade-offs and mismatch trade-offs, we may hesitate to diagnose true 'disorder', although such conditions are often considered medically relevant, and dysfunction in the sense of harmful processes

[17] Although note that descriptions are often vague enough to include phenomena which classify differently e.g. 'anger' may be loosely defined enough that it includes cases caused by brain tumours and normal anger – in which case average fitness would still be positive and classify as an adaptation, but important inner heterogeneity would remain, requiring closer description specifiers. Strict demarcation and exclusivity refers to every analysis having a single positive, neutral or negative fitness. Different description and framing specifiers can alter this result.

exists. We likely reconsider whether depression is a 'disorder' if believing it has modern functional aspects and would have been adaptive ancestrally. However, cases of mismatch disruption are serious enough to affect an individual's fitness, and likely involve novel biological manifestations which meet most naturalistic intuitions for justifying diagnosing disorder. If major depression occurs in a new, truly debilitating form, despite the responsible mood system being ultimately caused by an adaptive history, we may still consider this a case of naturalistic 'dysfunction' justifying disorder attribution. Distributed disruption, similarly, likely meets intuitions of biological processes sufficient to diagnose disorder, again despite being ultimately caused by adaptation. Still, these are certainly not simple diseases, and an evolutionary perspective on their differential status seems warranted.

Perhaps attaching other existing health terminology such as 'ailment', 'disease' or 'condition' to specific classes of dysfunction may be appropriate. This is not my concern here, and I leave this matter open to future conceptual analysis and social consensus. What's important is that scientific programs to discern the class of dysfunction are more strict and tractable (if methodologically complex) than in Wakefield's HDA, making his contribution workable.

Bibliography

Abed, Riadh, and P. St John-Smith, eds. 2022. *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge: Cambridge University Press.

Abed, Riadh, and Paul St John-Smith. 2022. "Introduction to Evolutionary Psychiatry." Pp. 1–18 in *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*, edited by P. St John-Smith and R. Abed. Cambridge: Cambridge University Press.

Al-Shawaf, Laith, David M. G. Lewis, Nicole Barbaro, and Yzar S. Wehbe. 2020. "The Products of Evolution: Conceptual Distinctions, Evidentiary Criteria, and Empirical Examples." *The SAGE Handbook of Evolutionary Psychology* 70–95. doi: 10.4135/9781529739442.N5.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Association.

Andrews, Paul W., and J. Anderson Thomson Jr. 2009. "The Bright Side of Being Blue: Depression as an Adaptation for Analyzing Complex Problems." *Psychological Review* 116(3):620.

Ariew, Andre, Robert Cummins, and Mark Perlman, eds. 2002. *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford University Press.

Baron-Cohen, Simon. 2020. *The Pattern Seekers: How Autism Drives Human Invention*. Basic Books.

Bech, Per. 2022. "Rating Scales in Depression: Limitations and Pitfalls." *Https://Doi.Org/10.31887/DCNS.2006.8.2/Pbech* 8(2):207–15. doi: 10.31887/DCNS.2006.8.2/PBECH.

Bergner, Raymond M. 1997. "What Is Psychopathology? And so What?" *Clinical Psychology: Science and Practice* 4(3):235–48.

Best, Damian, and Siladitya Bhattacharya. 2015. "Obesity and Fertility." *Hormone Molecular Biology and Clinical Investigation* 24(1):5–10. doi: 10.1515/HMBCI-2015-0023/MACHINEREADABLECITATION/RIS.

De Block, Andreas, and Jonathon Sholl. 2021. "Harmless Dysfunctions and the Problem of Normal Variation."

Defining Mental Disorder 495–510. doi: 10.7551/MITPRESS/9949.003.0031.

Boorse, Christopher. 1997. "A Rebuttal on Health." Pp. 1–134 in. Humana Press, Totowa, NJ.

Borsboom, Denny, Angélique O. J. Cramer, and Annemarie Kalis. 2019. "Brain Disorders? Not Really: Why Network Structures Block Reductionism in Psychopathology Research." *Behavioral and Brain Sciences* 42:1–54. doi: 10.1017/S0140525X17002266.

Boudry, Maarten, and Massimo Pigliucci. 2013. "The Mismeasure of Machine: Synthetic Biology and the Trouble with Engineering Metaphors." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44(4):660–68. doi: 10.1016/J.SHPSC.2013.05.013.

Bourrat, Pierrick, and Paul Edmund Griffiths. 2021. "The Idea of Mismatch in Evolutionary Medicine." *Https://Doi.Org/10.1086/716543*. doi: 10.1086/716543.

Bowie, Christopher R., Colin Depp, John A. McGrath, Paula Wolyniec, Brent T. Mausbach, Mary H. Thornquist, James Luke, Thomas L. Patterson, Philip D. Harvey, and Ann E. Pulver. 2010. "Prediction of Real-World Functional Disability in Chronic Mental Disorders: A Comparison of Schizophrenia and Bipolar Disorder." *American Journal of Psychiatry* 167(9):1116–24. doi: 10.1176/APPI.AJP.2010.09101406/ASSET/IMAGES/LARGE/APPI.AJP.2010.1116B.JPEG.

Bundy, H., D. Stahl, and J. H. MacCabe. 2011. "A Systematic Review and Meta-Analysis of the Fertility of Patients with Schizophrenia and Their Unaffected Relatives." *Acta Psychiatrica Scandinavica* 123(2):98–106. doi: 10.1111/j.1600-0447.2010.01623.x.

Chapman, Robert. 2021. "Neurodiversity and the Social Ecology of Mental Functions." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 16(6):1360–72. doi: 10.1177/1745691620959833.

Crespi, Bernard. 2016. "Autism as a Disorder of High Intelligence." *Frontiers in Neuroscience* 10(JUN):300. doi: 10.3389/fnins.2016.00300.

Crossley, Nick. 1998. "R. D. Laing and the British Anti-Psychiatry Movement: A Socio–Historical Analysis." *Social Science & Medicine* 47(7):877–89. doi: 10.1016/S0277-9536(98)00147-6.

Cummins, Robert. 1975. "Functional Analysis." *Journal of Philosophy* 72(November).

Darwin, Charles. 1869. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. 5th ed. London: John Murray.

Dawkins, R. 1986. *The Blind Watchmaker*. Longman Scientific and Technical.

Dawkins, Richard. 1978. *The Selfish Gene, The Selfish Gene*.

Drescher, Jack. 2015. "Out of DSM: Depathologizing Homosexuality." *Behavioral Sciences* 5(4):565–75. doi: 10.3390/bs5040565.

Engel, George L. 1977. "The Need for a New Medical Model: A Challenge for Biomedicine." *Science* 196(4286):129–36. doi: 10.1126/science.847460.

Engelhardt, T. 1999. "The Disease of Masturbation: Value and the Concept of Disease." in *Meaning and Medicine: A Reader in the Philosophy of Health Care*, edited by J. L. Nelson and Jh. L. Nelson. Routledge.

Evershed, Richard P., George Davey Smith, Mélanie Roffet-Salque, Adrian Timpson, Yoan Diekmann, Matthew S. Lyon, Lucy J. E. Cramp, Emmanuelle Casanova, Jessica Smyth, Helen L. Whelton, Julie Dunne, Veronika Brychova, Lucija Šoberl, Pascale Gerbault, Rosalind E. Gillis, Volker Heyd, Emily Johnson, Iain Kendall, Katie Manning, Arkadiusz Marciniak, Alan K. Outram, Jean-Denis Vigne, Stephen Shennan, Andrew Bevan, Sue Colledge, Lyndsay Allason-Jones, Luc Amkreutz, Alexandra Anders, Rose-Marie Arbogast, Adrian Bălăşescu, Eszter Bánffy, Alistair Barclay, Anja Behrens, Peter Bogucki, Ángel Carrancho Alonso, José Miguel Carretero, Nigel Cavanagh, Erich Claßen, Hipolito Collado Giraldo, Matthias Conrad, Piroska Csengeri, Lech Czerniak, Maciej Dębiec, Anthony Denaire, László Domboróczki, Christina Donald, Julia Ebert, Christopher Evans, Marta Francés-Negro, Detlef Gronenborn, Fabian Haack, Matthias Halle, Caroline Hamon, Roman Hülshoff, Michael Ilett, Eneko Iriarte, János Jakucs, Christian Jeunesse, Melanie Johnson, Andy M. Jones,

Necmi Karul, Dmytro Kiosak, Nadezhda Kotova, Rüdiger Krause, Saskia Kretschmer, Marta Krüger, Philippe Lefranc, Olivia Lelong, Eva Lenneis, Andrey Logvin, Friedrich Lüth, Tibor Marton, Jane Marley, Richard Mortimer, Luiz Oosterbeek, Krisztián Oross, Juraj Pavúk, Joachim Pechtl, Pierre Pétrequin, Joshua Pollard, Richard Pollard, Dominic Powlesland, Joanna Pyzel, Pál Raczky, Andrew Richardson, Peter Rowe, Stephen Rowland, Ian Rowlandson, Thomas Saile, Katalin Sebők, Wolfram Schier, Germo Schmalfuß, Svetlana Sharapova, Helen Sharp, Alison Sheridan, Irina Shevnina, Iwona Sobkowiak-Tabaka, Peter Stadler, Harald Stäuble, Astrid Stobbe, Darko Stojanovski, Nenad Tasić, Ivo van Wijk, Ivana Vostrovská, Jasna Vuković, Sabine Wolfram, Andrea Zeeb-Lanz, and Mark G. Thomas. 2022. "Dairying, Diseases and the Evolution of Lactase Persistence in Europe." *Nature 2022 608:7922* 608(7922):336–45. doi: 10.1038/s41586-022-05010-7.

Faucher, Luc. 2021. "Facts, Facts, Facts: HD Analysis Goes Factual." *Defining Mental Disorder* 47–70. doi: 10.7551/MITPRESS/9949.003.0006.

Faucher, Luc, and Denis Forest, eds. 2021. *Defining Mental Disorder: Jerome Wakefield and His Critics*. The MIT Press.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford: The Clarendon Press.

Forest, Denis, and Luc Faucher. 2021. "Introduction." in *Defining Mental Disorder: Jerome Wakefield and His Critics*, edited by L. Faucher and D. Forest. The MIT Press.

Fried, Eiko I., Claudia D. van Borkulo, Angélique O. J. Cramer, Lynn Boschloo, Robert A. Schoevers, and Denny Borsboom. 2017. "Mental Disorders as Networks of Problems: A Review of Recent Insights." *Social Psychiatry and Psychiatric Epidemiology* 52(1):1–10. doi: 10.1007/s00127-016-1319-z.

Garson, Justin. 2021. "The Developmental Plasticity Challenge to Wakefield's View." *Defining Mental Disorder* 335–52. doi: 10.7551/MITPRESS/9949.003.0021.

Godfrey-Smith, Peter. 1994. "A Modern History Theory of Functions." *Noûs* 28(3).

Goldsmith, Timothy H. 1990. "Optimization, Constraint, and History in the Evolution of Eyes." *The Quarterly Review of Biology* 65(3):281–322. doi: 10.1086/416840.

Gregory, T. Ryan. 2009. "Understanding Natural Selection: Essential Concepts and Common Misconceptions." *Evolution: Education and Outreach* 2(2):156–75. doi: 10.1007/S12052-009-0128-1/FIGURES/3.

Haukka, Jari, Jaana Suvisaari, and Jouko Lönnqvist. 2003. "Fertility of Patients with Schizophrenia, Their Siblings, and the General Population: A Cohort Study from 1950 to 1959 in Finland." *American Journal of Psychiatry* 160(3):460–63. doi: 10.1176/APPI.AJP.160.3.460/ASSET/IMAGES/LARGE/L46T2.JPEG.

Hidaka, Brandon H., Anila Asghar, C. Athena Aktipis, Randolph M. Nesse, Terry M. Wolpaw, Nicole K. Skursky, Katelyn J. Bennett, Matthew W. Beyrouty, and Mark D. Schwartz. 2015. "The Status of Evolutionary Medicine Education in North American Medical Schools." *BMC Medical Education* 15(1):1–9. doi: 10.1186/S12909-015-0322-5/FIGURES/3.

Holcomb, Harmon R. 1996. "Just so Stories and Inference to the Best Explanation in Evolutionary Psychology." *Minds and Machines* 6(4):525–40. doi: 10.1007/BF00389657/METRICS.

Holm, Sune. 2012. "Biological Interests, Normative Functions, and Synthetic Biology." *Philosophy and Technology* 25(4).

Hunt, Adam D., Riadh Abed, and Paul St John-Smith. 2022. "The Biopsychosocial Model Advanced by Evolutionary Theory." in *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*, edited by R. Abed and P. St John-Smith. Cambridge: Cambridge University Press.

Hunt, Adam D., and Adrian V. Jaeggi. 2022. "Specialised Minds: Extending Adaptive Explanations of Personality to the Evolution of Psychopathology." *Evolutionary Human Sciences* 4:e26. doi: 10.1017/EHS.2022.23.

Hunt, Adam D., Paul St-John Smith, and Riadh Abed. 2023. "Evobiopsychosocial Medicine." *Evolution, Medicine, and Public Health* 11(1):67–77. doi: 10.1093/EMPH/EOAC041.

Kano, Fumihiro, Yuri Kawaguchi, and Yeow Hanling. 2022. "Experimental Evidence That Uniformly White Sclera

Enhances the Visibility of Eye-Gaze Direction in Humans and Chimpanzees." *ELife* 11. doi: 10.7554/ELIFE.74086.

Keller, Matthew C., and Geoffrey Miller. 2006. "Resolving the Paradox of Common, Harmful, Heritable Mental Disorders: Which Evolutionary Genetic Models Work Best?" *Behavioral and Brain Sciences* 29(4):385–404. doi: 10.1017/S0140525X06009095.

Kelly, Robert L. 2013. *The Lifeways of Hunter-Gatherers: The Foraging Spectrum*.

Kennair, Leif Edward. 2010. "The Problem of Defining Psychopathology and Challenges To Evolutionary Psychology Theory." Pp. 451–79 in *The Evolution of Personality and Individual Differences*, edited by D. M. Buss and P. H. Hawley.

Khoury, Bassam, Ellen J. Langer, and Francesco Pagnini. 2014. "The DSM: Mindful Science or Mindless Power? A Critical Review." *Frontiers in Psychology* 5:602. doi: 10.3389/fpsyg.2014.00602.

Kincaid, Harold. 2021. "Doing without 'Disorder' in the Study of Psychopathology." *Defining Mental Disorder* 123–32. doi: 10.7551/MITPRESS/9949.003.0010.

Klein, Donald F. 1999. "Harmful Dysfunction, Disorder, Disease, Illness, and Evolution." *Journal of Abnormal Psychology* 108(3):421–29. doi: 10.1037/0021-843X.108.3.421.

Kotov, Roman, Robert F. Krueger, David Watson, Thomas M. Achenbach, Robert R. Althoff, R. Michael Bagby, Timothy A. Brown, William T. Carpenter, Avshalom Caspi, Lee Anna Clark, Nicholas R. Eaton, Miriam K. Forbes, Kelsie T. Forbush, David Goldberg, Deborah Hasin, Steven E. Hyman, Masha Y. Ivanova, Donald R. Lynam, Kristian Markon, Joshua D. Miller, Terrie E. Moffitt, Leslie C. Morey, Stephanie N. Mullins-Sweatt, Johan Ormel, Christopher J. Patrick, Darrel A. Regier, Leslie Rescorla, Camilo J. Ruggero, Douglas B. Samuel, Martin Sellbom, Leonard J. Simms, Andrew E. Skodol, Tim Slade, Susan C. South, Jennifer L. Tackett, Irwin D. Waldman, Monika A. Waszczuk, Thomas A. Widiger, Aidan G. C. Wright, and Mark Zimmerman. 2017. "The Hierarchical Taxonomy of Psychopathology (HiTOP): A Dimensional Alternative to Traditional Nosologies." *Journal of Abnormal Psychology* 126(4):454–77. doi: 10.1037/abn0000258.

Lemoine, Maël. 2021. "Is the Dysfunction Component of the 'Harmful Dysfunction Analysis' Stipulative?" *Defining Mental Disorder* 199–212. doi: 10.7551/MITPRESS/9949.003.0016.

Matthewson, John, and Paul Griffiths. 2017. "Biological Criteria of Disease: Four Ways of Going Wrong." *The Journal of Medicine and Philosophy* 42. doi: 10.1093/jmp/jhx004.

Maynard Smith, John. 1993. *The Theory of Evolution*. Cambridge University Press.

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

Murphy, Dominic. 2020. "Concepts of Disease and Health" edited by E. N. Zalta. *The Stanford Encyclopedia of Philosophy*.

Murphy, Dominic. 2021. "Function and Dysfunction." *Defining Mental Disorder* 251–66. doi: 10.7551/MITPRESS/9949.003.0018.

Neander, Karen. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science* 58(2):168–84. doi: 10.1086/289610.

Nesse, Randolph M. 2000. "Is Depression an Adaptation?" *Archives of General Psychiatry* 57(1):14–20.

Nesse, Randolph M. 2005. "Maladaptation and Natural Selection." *The Quarterly Review of Biology* 80(1):62–70.

Nesse, Randolph M. 2023. "Evolutionary Psychiatry: Foundations, Progress and Challenges." *World Psychiatry* 22(2):177–202. doi: 10.1002/WPS.21072.

Nesse, Randolph M., and Stephen C. Stearns. 2008. "The Great Opportunity: Evolutionary Applications to Medicine and Public Health." *Evolutionary Applications* 1(1):28–48. doi: 10.1111/j.1752-4571.2007.00006.x.

Norman, Geoff. 2010. "Likert Scales, Levels of Measurement and the 'Laws' of Statistics." *Advances in Health*

Sciences Education : Theory and Practice 15(5):625–32. doi: 10.1007/S10459-010-9222-Y.

Oliver, Mike. 2013. "The Social Model of Disability: Thirty Years On." *Https://Doi.Org/10.1080/09687599.2013.818773* 28(7):1024–26. doi: 10.1080/09687599.2013.818773.

Orr, H. Allen. 2009. "Fitness and Its Role in Evolutionary Genetics." *Nature Reviews. Genetics* 10(8):531–39. doi: 10.1038/nrg2603.

Orzack, Steven Hecht, and Elliott Sober. 1994. "Optimality Models and the Test of Adaptationism." *The American Naturalist* 143(3):361–80.

Peter, Benjamin M., and Montgomery Slatkin. 2013. "Detecting Range Expansions from Genetic Data." *Evolution; International Journal of Organic Evolution* 67(11):3274–89. doi: 10.1111/EVO.12202.

Pinker, Steven. 1997. *How the Mind Works.* New York, NY, US: W W Norton & Co.

Power, Michael L. 2012. "The Human Obesity Epidemic, the Mismatch Paradigm, and Our Modern 'Captive' Environment." *American Journal of Human Biology* 24(2):116–22. doi: 10.1002/AJHB.22236.

Rabinowitz, Jonathan, Stephen Z. Levine, George Garibaldi, Dragana Bugarski-Kirola, Carmen Galani Berardo, and Shitij Kapur. 2012. "Negative Symptoms Have Greater Impact on Functioning than Positive Symptoms in Schizophrenia: Analysis of CATIE Data." *Schizophrenia Research* 137(1–3):147–50. doi: 10.1016/J.SCHRES.2012.01.015.

Ranciaro, Alessia, Michael C. Campbell, Jibril B. Hirbo, Wen Ya Ko, Alain Froment, Paolo Anagnostou, Maritha J. Kotze, Muntaser Ibrahim, Thomas Nyambo, Sabah A. Omar, and Sarah A. Tishkoff. 2014. "Genetic Origins of Lactase Persistence and the Spread of Pastoralism in Africa." *The American Journal of Human Genetics* 94(4):496–510. doi: 10.1016/J.AJHG.2014.02.009.

Reeve, Hudson Kern, and Paul W. Sherman. 1993. "Adaptation and the Goals of Evolutionary Research." *The Quarterly Review of Biology* 68(1):1–32.

Schlötterer, Christian. 2003. "Hitchhiking Mapping – Functional Genomics from the Population Genetics Perspective." *Trends in Genetics* 19(1):32–38. doi: 10.1016/S0168-9525(02)00012-4.

Ségurel, Laure, and Céline Bon. 2017. "On the Evolution of Lactase Persistence in Humans." *Https://Doi.Org/10.1146/Annurev-Genom-091416-035340* 18:297–319. doi: 10.1146/ANNUREV-GENOM-091416-035340.

Simon, G. E., D. P. Goldberg, M. Von Korff, and T. B. Üstün. 2002. "Understanding Cross-National Differences in Depression Prevalence." *Psychological Medicine* 32(4):585–94. doi: 10.1017/S0033291702005457.

Skakkebæk, Niels E., Rune Lindahl-Jacobsen, Hagai Levine, Anna Maria Andersson, Niels Jørgensen, Katharina M. Main, Øjvind Lidegaard, Lærke Priskorn, Stine A. Holmboe, Elvira V. Bräuner, Kristian Almstrup, Luiz R. Franca, Ariana Znaor, Andreas Kortenkamp, Roger J. Hart, and Anders Juul. 2021. "Environmental Factors in Declining Human Fertility." *Nature Reviews Endocrinology 2021 18:3* 18(3):139–57. doi: 10.1038/s41574-021-00598-8.

Sober, Elliott. 2000. *Philosophy of Biology*. Routledge: Routledge.

Spitzer, R. L. 1999. "Harmful Dysfunction and the DSM Definition of Mental Disorder." *Journal of Abnormal Psychology* 108(3):430–32.

Symons, Donald. 1990. "Adaptiveness and Adaptation." *Ethology and Sociobiology* 11(4):427–44. doi: https://doi.org/10.1016/0162-3095(90)90019-3.

Thornton, Tim. 2021. "Naturalism and Dysfunction." *Defining Mental Disorder* 469–84. doi: 10.7551/MITPRESS/9949.003.0028.

Varner, Gary E. 1998. *In Nature's Interests: Interests, Animal Rights, and Environmental Ethics*. Vol. 109. Oxford University Press.

De Vreese, Leen. 2021. "Against the Disorder/Nondisorder Dichotomy." *Defining Mental Disorder* 97–108. doi:

10.7551/MITPRESS/9949.003.0008.

Wakefield, Jerome. 2021a. “Are Cognitive Neuroscience and the Harmful Dysfunction Analysis Competitors or Allies? Reply to Philip Gerrans.” *Defining Mental Disorder* 415–32. doi: 10.7551/MITPRESS/9949.003.0025.

Wakefield, Jerome. 2021b. “Are There Naturally Selected Disorders? Supplementary Reply to Rachel Cooper.” *Defining Mental Disorder* 577–92. doi: 10.7551/MITPRESS/9949.003.0035.

Wakefield, Jerome. 2021c. “Biological Function Hierarchies and Indeterminacy of Dysfunction: Supplementary Reply to Justin Garson.” *Defining Mental Disorder* 385–96. doi: 10.7551/MITPRESS/9949.003.0023.

Wakefield, Jerome. 2021d. “Can a Nonessentialist Neo-Empiricist Analysis of Mental Disorder Replace the Harmful Dysfunction Analysis? Reply to Peter Zachar.” *Defining Mental Disorder* 177–96. doi: 10.7551/MITPRESS/9949.003.0014.

Wakefield, Jerome. 2021e. “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. doi: 10.7551/MITPRESS/9949.003.0019.

Wakefield, Jerome. 2021f. “Can the Harmful Dysfunction Analysis Distinguish Problematic Normal Variation from Disorder? Reply to Andreas De Block and Jonathan Sholl.” *Defining Mental Disorder* 511–36. doi: 10.7551/MITPRESS/9949.003.0032.

Wakefield, Jerome. 2021g. “Do the Challenges of Autism and Neurodiversity Pose an Objection to the Harmful Dysfunction Analysis? Reply to Denis Forest.” *Defining Mental Disorder* 449–68. doi: 10.7551/MITPRESS/9949.003.0027.

Wakefield, Jerome. 2021h. “Do the Empirical Facts Support the Harmful Dysfunction Analysis? Reply to Luc Faucher.” *Defining Mental Disorder* 71–96. doi: 10.7551/MITPRESS/9949.003.0007.

Wakefield, Jerome. 2021i. “Do the Works of Carl Craver or Marcel Weber Explain How Causal Role Functions Can Provide Objective Medical Judgments of Dysfunction? Supplementary Reply to Dominic Murphy.” *Defining Mental Disorder* 317–34. doi: 10.7551/MITPRESS/9949.003.0020.

Wakefield, Jerome. 2021j. “Does Developmental Plasticity Pose a Challenge to the Harmful Dysfunction Analysis? Reply to Justin Garson.” *Defining Mental Disorder* 353–84. doi: 10.7551/MITPRESS/9949.003.0022.

Wakefield, Jerome. 2021k. “From Ribot and Dupré to Spitzer and RDoC: Does the Harmful Dysfunction Analysis Possess Historical Explanatory Power? Reply to Steeves Demazeux.” *Defining Mental Disorder* 27–46. doi: 10.7551/MITPRESS/9949.003.0005.

Wakefield, Jerome. 2021l. “Is Indeterminacy of Biological Function an Objection to the Harmful Dysfunction Analysis? Reply to Tim Thornton.” *Defining Mental Disorder* 485–92. doi: 10.7551/MITPRESS/9949.003.0029.

Wakefield, Jerome. 2021m. “Is the Harmful Dysfunction Analysis Descriptive or Stipulative, and Is the HDA or BST the Better Naturalist Account of Dysfunction? Reply to Maël Lemoine.” *Defining Mental Disorder* 213–50. doi: 10.7551/MITPRESS/9949.003.0017.

Wakefield, Jerome. 2021n. “Must Social Values Play a Role in the Harm Component of the Harmful Dysfunction Analysis? Reply to Rachel Cooper.” *Defining Mental Disorder* 553–76. doi: 10.7551/MITPRESS/9949.003.0034.

Wakefield, Jerome. 2021o. “Quinian Qualms, or Does Psychiatry Really Need the Harmful Dysfunction Analysis? Reply to Harold Kincaid.” *Defining Mental Disorder* 133–54. doi: 10.7551/MITPRESS/9949.003.0011.

Wakefield, Jerome C. 1992. “The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values.” *American Psychologist* 47(3):373–88. doi: 10.1037/0003-066X.47.3.373.

Wakefield, Jerome C. 1997. “When Is Development Disordered? Developmental Psychopathology and the Harmful Dysfunction Analysis of Mental Disorder.” *Development and Psychopathology* 9(02):269–90. doi: 10.1017/s0954579497002058.

Wakefield, Jerome C. 2015. “Biological Function and Dysfunction: Conceptual Foundations of Evolutionary Psychopathology.” Pp. 1–19 in *The Handbook of Evolutionary Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

West-Eberhard, M. J. 2003. *Developmental Plasticity and Evolution.* Oxford, UK: Oxford University Press.

Williams, George C. 1966. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Wouters, Arno. 2005. “The Function Debate in Philosophy.” *Acta Biotheoretica* 53(2):123–51. doi: 10.1007/S10441-005-5353-6/METRICS.

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

Zachar, Peter, and Kenneth S. Kendler. 2017. “The Philosophy of Nosology.” *Annual Review of Clinical Psychology* 13(1):49–71. doi: 10.1146/annurev-clinpsy-032816-045020.

The DCIDE Method: evolutionary explanations beyond just-so storytelling, exemplified with autism

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

Abstract

Evolutionary explanations of mental disorders are a longstanding aim of evolutionary psychiatry, but suffer from complexities including within-disorder phenotypic and etiological heterogeneity and environmental effects obscuring relevant adaptive processes, exaggerating the 'just-so storytelling' critique levelled against evolutionary psychology. To address this, we present a novel five-stage method of systematic review, the 'DCIDE method', exemplified with autism. Description identifies a trait to explain, Categorisation excludes non-adaptive cases, Integration hones a target for adaptive explanation by including relevant correlated traits visible to selection. Evolutionary hypotheses are then Depicted and Evaluated for their ability to explain all the evidence at hand. Each stage uses standardised areas of evidence and theoretically motivated analytical principles to increase rigour. In the DCIDE review of autism, when Described with current diagnostic criteria up to 20% of cases Categorise as non-adaptive, primarily caused by de novo mutations and environmental trauma. Remaining cases are eligible for adaptive explanation. Integrating subclinical familial and camouflaged female cases, which share genetic components, is necessary. Competing Depictions contrast a high intelligence by-product hypothesis with social niche specialisation for high 'systemising' cognition. In Evaluation, broad evidence including age of onset, duration, sex differences, environmental plasticity and prevalence support the social niche hypothesis while the intelligence by-product hypothesis fails to predict various lines of evidence. The DCIDE method allows improved systematic evolutionary analysis across psychiatric conditions, and may also strengthen evolutionary psychology more generally, providing robust defence against just-so storytelling and cherry-picking critiques.

1/ Introduction

Major outstanding questions pervade the science of psychopathology. Psychiatric conditions such as autism and schizophrenia are generally of high cost, unknown etiology and incomprehensible pathophysiology (Alawieh et al., 2012). Unsuccessful efforts towards explanations or treatments have been plentiful, with advances in neuroscience and genetics generally failing to identify pathology and blurring boundaries between disease states and normal variation (Geschwind & Flint, 2015). Thomas Insel, who led the National Institute of Mental Health between 2002 and 2015 and oversaw around $20 billion of research spending (Rogers, 2017) said, 'Whatever we've been doing for five decades, it ain't working

... Maybe we just need to rethink this whole approach ... With no validated biomarkers and too little in the way of novel medical treatments since 1980 ... it is time to rethink mental disorders' (G. Greenberg, 2013).

In tandem with burgeoning evidence in genetics, epidemiology and neuroscience in mainstream psychiatric research, the fields of evolutionary psychiatry and evolutionary medicine have been developing theoretical frameworks attempting to make sense of disorder at the evolutionary level (Abed & St John-Smith, 2022; Hunt et al., 2023; Nesse, 2019; Nesse & Williams, 1996). Evolutionary approaches are appealing partly because they offer uniquely objective criteria for classifying biological traits as healthy or disordered, functional or dysfunctional (Hunt et al., 2022; Klein, 1999; Spitzer, 1999; Wakefield, 1992, 1997), by asking about traits' positive or negative effects on reproductive success (Hunt, 2023). A further fundamental motivation for evolutionary psychiatry is driven by a paradox recognised for many decades (Huxley et al., 1964) – how is it possible that mental disorders are obviously harmful, including often for reproductive success, yet still common and heritable? Why has natural selection not removed them from the population? Keller and Miller (2006) framed this most sophisticatedly, yet their proposed solution relying upon mutation-selection balance has only been partially supported by recent genetic data (Keller, 2018). The paradox remains, and researchers continue to present possible solutions (Abed & St John-Smith, 2022; Del Giudice, 2018a; Hunt & Jaeggi, 2022; Nesse, 2023) ranging from adaptationist perspectives which reframe disorders as functional adaptations misinterpreted as diseases; as novel products of adaptive systems which are mismatched to modern environments and so produce harmful phenotypes; as unavoidable consequences of mutation or stochastic developmental processes; as costly extremes of adaptive ranges which fall off a 'cliff-edge'; and more.

Despite evolutionary approaches offering a novel paradigm for understanding mental disorders, evolutionary theorising, especially of psychological traits, has long been criticised for making speculative untestable hypotheses which represent mere 'just-so storytelling' (Dupré, 2003; Gould & Lewontin, 1979; Smith, 2019). Evolutionary psychologists typically counter this critique by noting that only selection can produce complex design in nature, building on the classic definition of adaptation by George Williams (1966); they proceed to study psychological adaptations by predicting specific design features given a hypothesized function and testing for evidence of such design in laboratory experiments, sometimes contrasting different adaptive and non-adaptive hypotheses (Cosmides & Tooby, 2013; Lewis et al., 2017). Nonetheless, criticism persists due to perceived difficulty in objectively defining functional specificity and complexity, lack of sufficient testable predictions or general principles for deriving predictions, and reliance on modern observations of functioning (Smith, 2019).

Evolutionary psychiatry is arguably troubled even more than evolutionary psychology by these criticisms (Zachar & Kendler, 2017). Disorder categories are heterogenous and so may contain functional and dysfunctional subtypes, and complex design may be harder to ascertain due to uncertainty of functional role or mismatch with modern environments (Hunt, 2023). Relevant subclinical manifestations of a particular disorder may be disregarded from analysis due to clinical focus on disease categories rather than quantitative traits, despite genetic relationships and shared causality between subclinical and clinical forms (Smoller et al., 2018). The prima facie appeal of explaining psychological and psychiatric traits with reference to our evolutionary history is thus troubled by several limitations. This article aims to help overcome such limitations by providing a systematic approach to formulating and evaluating evolutionary hypotheses.

2/ The DCIDE Method

The DCIDE method is a novel approach to analysis and systematic review aimed at structuring evolution-informed explanations making principled use of a holistic picture of empirical evidence. The 'DCIDE' acronym stands for the five stages of the method (Figure 2). Briefly, a trait is Described; subtypes within it are Categorised as requiring adaptive explanation or not; for those requiring adaptive explanation, relevant genetically related traits are Integrated to try and capture the phenotype visible to selection; various hypotheses are Depicted; and those hypotheses are Evaluated in a standardised manner.

A vast array of relevant evidence from genetics, neurobiology, epidemiology and psychology exists, but attempts to integrate them currently proceed without a theoretically-justified structure or standardised analytical principles, allowing criticisms of just-so storytelling and cherry-picking. The DCIDE method is introduced to provide a better structure for formulating and comparing explanatory adaptive hypotheses. Its framework helps distinguish evolution-informed kinds of disorder (Figure 1), allowing for a single condition's heterogeneity to be parsed out (e.g. different cases of autism could be primarily caused by vulnerability, mutation, by-product or adaptation).

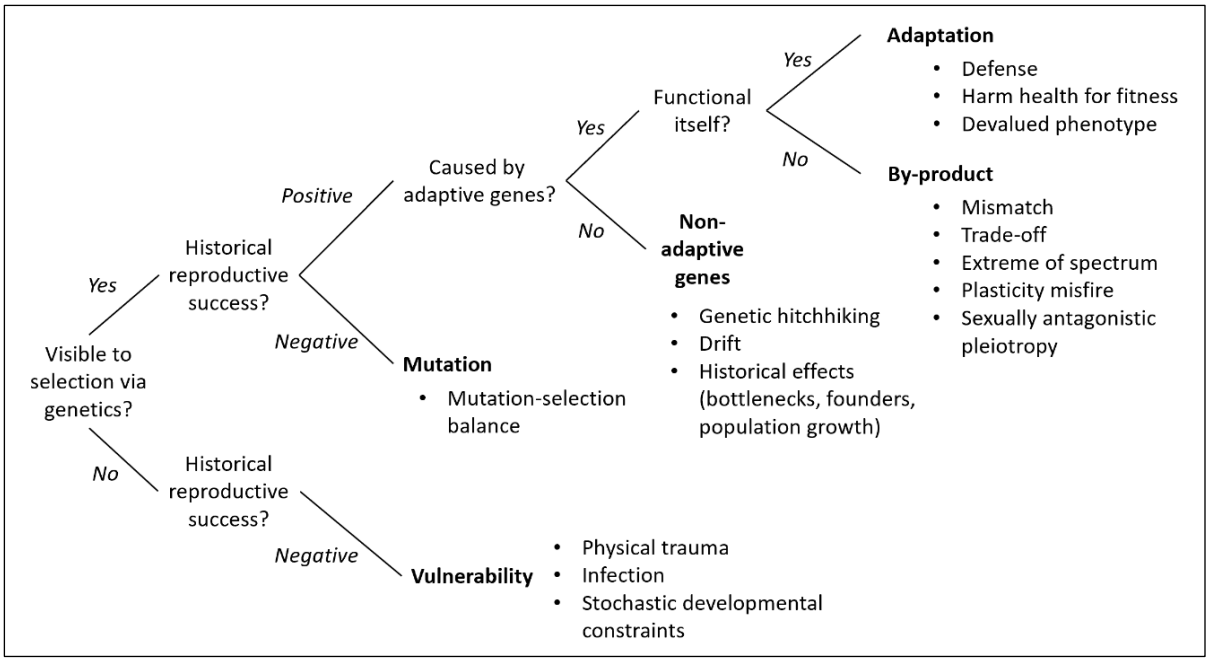

Figure 1: Common kinds of explanations for medically-relevant conditions. Bold text indicates distinct kinds of explanation. Bullet points list common examples from evolutionary medicine, which are not mutually exclusive with each other. For a closer breakdown of mutually exclusive evolutionary explanations, see Hunt (2023). Note that mutation-selection balance is easily identifiable in cases implying historical negative reproductive success, but non-adaptive genes can also be mutations awaiting negative selection, positively selected due to other circumstances.

When causing mental disorders, mutation, vulnerability and non-adaptive genes are likely fitness-reducing, so can be defined as disruptions leading to dysfunctional processes (Hunt 2023): fairly plain cases of disorder. However, adaptations and by-products of adaptations may also cause traits we diagnose as mental disorders via various pathways. We broadly call these 'adaptive explanations' – they are explanations requiring reference to positive selection on the causal alleles, although note that many traits requiring adaptive explanation may be harmful by-products, not adaptive in themselves (considered further in §5).

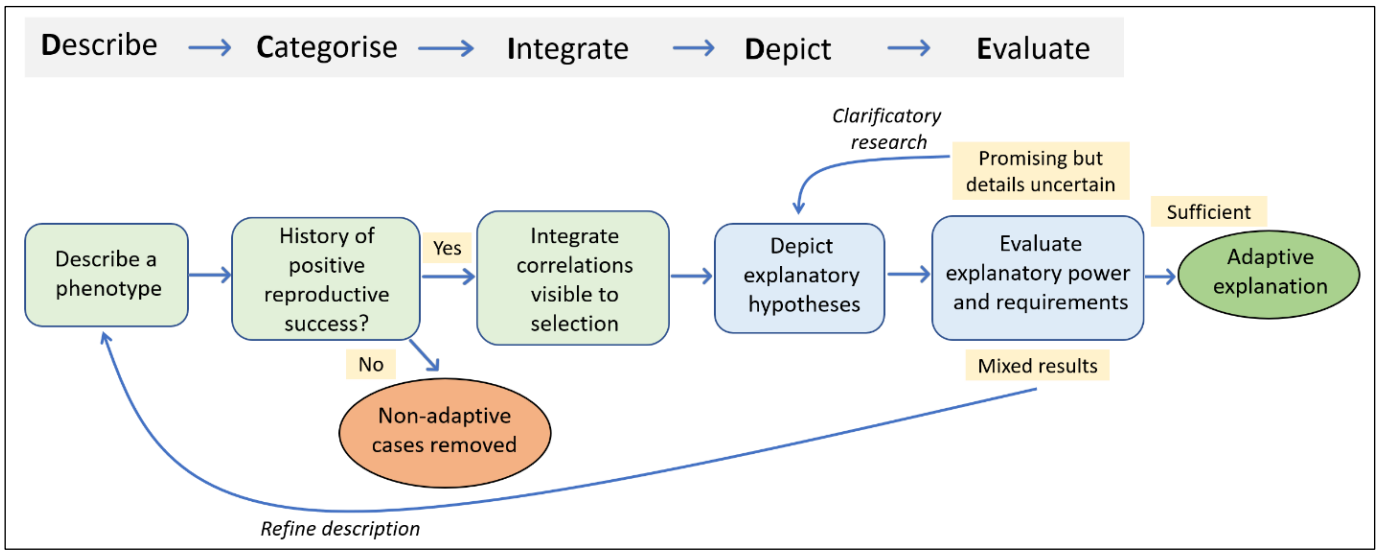

Figure 2: The DCIDE Method. Green rectangles are steps for honing target of explanation. Blue rectangles are steps attempting adaptive explanation. Yellow text indicates possible outcomes of analysis. Ovals indicate sufficient explanation.

Despite limited space preventing extensive exemplification of its application here, the basic principles and format of the DCIDE method could flexibly adapt around new evidence, analytical methods, or even entirely new areas of evidence (e.g. microbiome research). Its structure is suitable for reviews of literature of various lengths – from full systematic reviews to structuring introductions in empirical papers. It could be used to design large scale research programs attempting to explain a target phenotype. Fundamentally it is a way of organising scientific evidence in a way that assists adaptive explanations of complex biological phenomena by integrating multiple areas of incommensurable relevant evidence. With minor adjustments it could help illuminate adaptive explanations of healthy psychological or physiological traits, or vulnerability to simple diseases. Its usefulness in enhancing evolutionary psychology will be shown in conclusion, although the overall method's primary scientific value is in assessing the phenomena of heterogenous etiology that remain confusing to psychiatry.

This paper primarily exemplifies the DCIDE method with autism ('autism spectrum disorder' in current diagnostic manuals; 'autism' and other de-pathologised terminology (Bury et al., 2023) used here where possible). Autism is exemplified for several reasons. Firstly, it is highly heterogenous, in both phenotypic presentation and known etiologies (Lord et al., 2020), which is highly problematic (Mottron, 2021), and the DCIDE method is particularly suitable for untangling such heterogeneity. Secondly, autism is widely researched, providing a large pool of available evidence. Thirdly, autism presents a profound evolutionary paradox (Keller & Miller, 2006), as one of the most highly heritable mental disorders (Baselmans et al., 2021), and is the subject of various evolutionary and non-evolutionary hypotheses (Del Giudice, 2018b), which can be assessed to exemplify the method's potential in discerning between hypotheses. Fourth, autism is a central concern of the 'neurodiversity' movement (Kapp et al., 2020) which seeks its reconceptualization as normal difference rather than disease, often emphasising associated cognitive strengths. Critics of this movement argue that highly debilitating cases cannot be construed as mere 'difference' and warrant separate categorisation (Singer, 2022). Previous authors note that this clash can be somewhat resolved by recognising heterogeneous causation in the autism spectrum (Baron-Cohen, 2019). The DCIDE method provides a formal analysis offering some resolution to this debate. This speaks to the importance of resolving scientific questions in an area of such critical social importance as psychiatry.

3/ DCIDE: Describe

Science necessarily starts by stipulating its object of study and empirical inclusion criteria. This is equivalent to Description, which identifies a phenotype to analyse. Methods of Describing could and do vary depending on theoretical goals, available technology and practicality. In psychiatry and psychology, questionnaires or behavioural checklists are

usually used, although the difficulty with presuming that these identify coherent targets of explanation is widely recognised (Zachar & Kendler, 2017), and a fundamental problem for the science of psychopathology. The DCIDE method aims at adaptive explanations, so Described phenotypes should ideally be considered plausible candidates for a shared adaptive explanation – however, later stages of Categorisation (§4), Integration (§5) and Evaluation (§7) provide specific methods for honing coherent targets of explanation. If an inappropriate target of explanation is Described, Evaluation will encourage re-Description (Figure 2), and in an iterative process, coherent targets should be identified. Hunt (2023) considers nuances of altering trait descriptions and downstream effects on evolutionary analysis in more detail.

Initial Description could identify a trait or cluster of traits forming a syndrome; or an emotional state; or theoretically any biological or psychological characteristic defined by the investigator. Modern mental disorder manuals provide descriptivist criteria based on clinical experience and expert consensus (Stein et al., 2010) to define conditions such as autism. Research on such conditions may use questionnaires, confirmed diagnosis, self-reported diagnosis or even self-diagnosis as inclusion criteria. Other methods of Describing psychological traits could come from factor analysis, as used by personality psychology (McCrae & Costa Jr., 2008) and proposed as appropriate for psychopathology, e.g. in the 'HiTOP' model (Kotov et al., 2017). In medical research outside psychiatry, specific biomarkers are often the preferred method of Description. The associated improvement in reliability with specific biomarkers may lead to simpler analysis of causation if reducing causal heterogeneity within the Described trait, although later stages of the DCIDE method do address this.

3.1/ Describe: Autism

In the case of autism, a DSM-5 (American Psychiatric Association, 2013) diagnosis (essentially equivalent to an International Classification of Disease (ICD) diagnosis) requires persistent deficits in social communication and social interaction across multiple contexts (e.g. deficits in social-emotional reciprocity; deficits in nonverbal communicative behaviors used for social interaction; deficits in developing, maintaining, and understanding relationships) and restricted, repetitive patterns of behavior, interests, or activities (e.g. stereotyped or repetitive motor movements, use of objects, or speech; insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal/nonverbal behavior; highly restricted, fixated interests that are abnormal in intensity or focus; hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment). These symptoms must be present in the early developmental period and cause clinically significant impairment in social, occupational, or other important areas of current functioning, and not be better explained by another intellectual disability or

developmental delay disorder – although can be diagnosed alongside intellectual disability if social communication is below expectation. These criteria can be taken to Describe autism, although research inclusion criteria differs in specifics from study to study, which is a limitation for a DCIDE review. Ideally, studies from genetics, epidemiology, and psychometrics would share inclusion criteria, making their results reliably mergeable.

It's notable that autistic symptoms are negatively framed – Baron Cohen has restated these core features in more neutral or even positively-charged language (e.g. 'the child may be fascinated by systems'; 'the child notices and recalls things other people may not' (Baron-Cohen, 2002)). Diagnostic criteria also don't mention positive traits associated with autism, such as special skills and perceptual sensitivity (Meilleur, Jelenic and Mottron, 2015) or enhanced logical reasoning (Brosnan et al., 2014; De Martino et al., 2008; South et al., 2014). Pioneers in the field of autism research recognised these talents as commonly accompanying autism (Asperger, 1944, translated by Frith 1991; Kanner, 1971; Rimland, 1978, 1988). If a requirement for some such strengths was added to diagnostic criteria, it would change the Described trait cluster; as it would if presence of such strengths was an exclusion criterion.

Diagnostic norms shift due to a variety of cultural reasons, including political goals such as access to healthcare and accommodations (Fletcher & O'Brien, 2008). Official psychiatric manuals have changed the definition of autism over time. The DSM-5 combined Asperger's syndrome, Autism and Post-Developmental-Disorder Not-Otherwise-Specified (PDD-NOS) into a more general 'Autism Spectrum Disorder', undoubtedly combining multiple etiologies under a single diagnosis (Lai et al., 2013); researchers now recognise they need to explain the 'autisms' (Geschwind & Levitt, 2007; Happé et al., 2006). This change has accompanied an apparent rise in autism prevalence as the diagnosis widens to encompass behaviours closer to 'normality' (Idring et al., 2015; Rødgaard et al., 2019), with loosening interpretations of core behavioural and social features (Zaroff & Uhm, 2012) and leading to much lower rates of accompanying intellectual disability (Lyall et al., 2017).

Although this has been reported by some as implying the 'autism' diagnosis will eventually become meaningless (Knapton, 2019), there is in principle no loosening, tightening, subtyping or removing of subtypes that makes a Described trait beyond the scope of the DCIDE method – although it affects the process of gathering and analysing the evidence and producing sufficient hypotheses. However, loosening criteria can increase causal heterogeneity. Tackling this begins in Categorisation.

4/ DCIDE: Categorise

The purpose of Categorisation is distinguishing a major kind of heterogeneity in causation, separating cases primarily explained by vulnerability, mutation or non-adaptive genetic factors from cases requiring adaptive explanation (Figure 1). Cases of dysfunctional

disruptions can be removed as sufficiently explained within a biomedical paradigm, whilst cases requiring reference to fitness-enhancing adaptations and their by-products are taken forward to the next stage of analysis.

Discovering verifiable dysfunction causing specific cases or correlations has led to claims that all autism (Keller and Miller 2006) and personality differences (Verweij et al., 2012) are simply maladaptive variation. The DCIDE method avoids this deduction. That *de novo* mutations cause phenotypes diagnosable as autism doesn't mean every instance of autism is caused by *de novo* mutations, or even that every instance of autism is due to non-adaptive processes. Down's syndrome is associated with high reported happiness (Skotko et al., 2011), traumatic brain injury with aggression (Rao et al., 2009) and Williams syndrome with gregariousness (Järvinen et al., 2013) – happiness, aggression and gregariousness cannot therefore be assumed as non-adaptations, and nor are these verifiable dysfunctions relevant to adaptive explanations of the wider functional traits (albeit potentially valuable for understanding proximate mechanisms).

Individuals' phenotypes arise downstream of an array of adaptations and disruptions – essentially every trait is affected by both positively selected alleles and rare or de novo alleles, and both adaptive and stochastic environmental reactions. Increased gregariousness of Williams syndrome exists amongst cognitive adaptations causing neurotypical cognition. Neurotypical gregariousness is affected by a mix of adaptations and disruptions. A perfect causal explanation would account for how these factors interact to lead to the individual phenotype. Such perfect explanations are currently implausible. More plausible and desirable is separating cases where an evolutionary process of adaptation plays a major role in explaining a trait from cases where disruption plays the major role.

Categorisation asks whether the major factor explaining Described trait cases is adaptation, vulnerability, mutation or non-adaptive genetic factors (Figure 1). If clearly non-adaptive cases are removed, evolutionary explanations of any adaptive causal elements of the Described trait become plausible. This utilises familiar biomedical variables from neuroscience, genetics and environmental evidence (Table 1). Identifying trait cases without a history of positive reproductive success allows identification of vulnerability and mutation, and molecular genetics can identify non-adaptive alleles. After Categorisation, cases of certifiable disruption are considered explained, and excluded from the ongoing analysis. Cases remaining are candidates for adaptive explanation, and taken forward. Details on Categorising each area now follow, exemplified with autism.

Area of Evidence	***Indications of non-adaptive cause***
Proximate system (Brain)	Differences not caused by inherited ontogenetic processes (e.g. physical damage, tumours, lesions).
Genes	DNA which cannot have a functional history (e.g. de novo mutations of SNPs, CNVs or other chromosomal abnormalities); rare genetic variation with traceable history of persistent harmful effects. Non-adaptive alleles identified as successful due to population history, linkage disequilibrium or drift.
Environment	Damage with direct pathological effects (e.g. toxins, physical traumas, infections) not mediated by genetic pathways.

Table 1: Biomedical variables allowing Categorisation as confirmable disruption and thus exclusion from further analysis. SNP = single nucleotide polymorphism. CNV = copy number variation.

4.1/ Brain Categorisation

Neuroscience can discover proximate causes of mental activity by studying brain function and structure. However, the assumption that finding brain differences associated with diagnosed disorder proves the presence of a brain disease is a serious (if tempting) error (Protopapas & Parrila, 2019). In Categorisation, the question asked of such findings is whether they result from a history of positive selection or vulnerabilities (Figure 1). Asking whether broadly similar biological systems (e.g. brain states) are visible in relatives and seem to result from normal heritable processes is a sufficient heuristic. We can clearly Categorise severe disruptions of an individual's normal ontogenetic developmental process (e.g. toxins, physical traumas, tumours) in this way – we can confirm they aren't inherited. Stochastic differences occurring in neurodevelopment are plausible future findings to allow Categorisation as disruption, but require development of specific methods discerning normal ontogenetic variation occurring due to generally functional alleles from dysfunctional stochastic errors.

4.1.1/ Brain categorisation: Autism

Common examples of non-adaptive cases of autism with clear neuropathology are seen in various cases of so-called 'syndromic autism'. For example, Rett syndrome meets an autism diagnosis between 40-97% of the time (Lyst & Bird, 2015). Caused by mutations in the MECP2 gene, it is associated with reduced brain volume, abnormally small and densely packed neurons with reduced dendritic complexity and spine density, and shorter pyramidal neurons in the motor and frontal cortices (Ip et al., 2018). Categorising these cases is possible by observing their significant contrast with relatives. In cases when syndromes with gross morphological and functional brain differences are inherited, Categorising could

require reference to genetics (see below). Evidence converges in such cases – Rett syndrome could be Categorised as a disruption on both neurological and genetic grounds.

Although autism of all forms is undoubtedly associated with brain differences, rarely does neuroscience discover autism caused by tumours or brain trauma (partly because criteria require symptoms from an early age). Certain syndromic cases show exceptional neuropathology but the majority of cases, particularly without intellectual disability, show diffuse and complex differences in brain growth and connectivity (Muhle et al., 2018), plausibly visible to selection via genetics and positively selected (Figure 1). Most cases thus cannot be Categorised as explained by disruption at the neuroscientific level.

4.2/ Genetic Categorisation

The growth of genetics as a field provides opportunity for unprecedented depth of evolutionary analysis: we can use genetics to infer basic facts about a trait's evolutionary history (Schraiber & Akey, 2015; Wilson, 1998). Most common mental disorders have a genetic component (Baselmans et al., 2021). Such apparently harmful genetic predispositions are explainable as positively selected (e.g. caused by antagonistic pleiotropy, mismatch, life history trade-offs or balancing selection) or awaiting negative selection (e.g. under mutation-selection balance or genetic drift) (Durisko et al., 2016).

Genetic evidence is critical in Categorisation. Alleles are usually classified as common, rare or *de novo* (although researchers differ on the prevalence cut-off between common and rare; either setting 'common' alleles as having frequencies above 1% (Raychaudhuri, 2011) or 5% (Belmont et al., 2005; Glinton & Elsea, 2019) of the population). Large effect *de novo* variants, including single-nucleotide polymorphisms (SNPs) or copy number variants (CNVs) are the simplest cases explainable as non-adaptive mutations (Figure 1).

Rare and especially common variants are more complicated, as they have been reproductively successful, perhaps because they are adaptive. Their individual effects on fitness may be neutral and drifting, positive and selected, or weakly negative but persisting due to inefficient selection (Figure 1). Approximately 130,000 variants (~4% of the genome) are shared with 0.5-5% of the population (Bourgeron, 2015). It's possible these result from positive or balancing selection (see §5.5), so ideally they require individual analysis. Specific techniques from population genetics (Adrion et al., 2020) may potentially be used to analyse whether such alleles have increased in frequency due to hitchhiking, population growth or bottlenecks, or simple drift and inefficient negative selection, and Described traits explained by these processes can be Categorised as non-adaptive. We suggest a placeholder heuristic that rare alleles (W. D. Hill et al., 2018) present in less than 0.5% of the population are likely non-adaptive, except in three possible scenarios: a) the allele is common and functional in

another population, but is diluted in the study population b) the allele arose recently and provides adaptive benefits but has not yet spread through the population c) the allele was more common and functional at some time in the past, but has undergone recent negative selection. Traits primarily caused by rare alleles not explained by these three possibilities are likely non-adaptive.

For other inherited traits, adaptive explanations become plausible. Ubiquitous or common genetic variants are evidence of significant historical reproductive success (inferred from their increase in frequency since arising *de novo*) and an apparent lack of negative selection (Uher, 2009). This also applies to cases of 'missing heritability', where broad sense heritability has not yet been explained by discovery of specific contributing alleles (Manolio et al., 2009). Despite their potential to discover additive genetic variation, genome wide association studies (GWAS) lack the power to detect specific causal alleles if there are complex gene-by-gene or gene-by-environment interactions (Parker & Palmer, 2011), expected of many adaptations, especially of highly plastic (e.g. psychological) phenotypes. Trait heritability observed in population and quantitative genetics needn't be fully explained by molecular genetics to be meaningful and imply adaptive explanation is required.

4.2.1/ Genetic Categorisation: Autism

Autism has high heritability – approximately 80-90% (Baselmans et al., 2021). Subgrouping the autism spectrum by genetic evidence is crucial (Geschwind & State, 2015; Lai et al., 2013). In a large Swedish study, Gaugler *et al.*, (2014) estimated approximately 6% of cases of autism are caused by a *de novo* or rare genetic variant (excluding confirmable syndromes such as Williams or Downs). Glinton and Elsea (2019) estimate over 400 forms of 'monogenic' autism exist, caused by *de novo* copy number variants (CNVs) and *de novo* single nucleotide variants (SNVs), and made a higher estimate that these forms account for 10-20% of all autism cases. These are often highly penetrant and syndromic, with various other cognitive and physical effects, often associated with intellectual disability (Iossifov et al., 2015; Robinson et al., 2014; Ronemus et al., 2014). These cases can be Categorised as disruptions and considered explained. Authors such as Mitchell (2015) develop Keller and Miller's (2006) mutation-selection balance account to propose all unexplained cases of autism may similarly be caused by disruptive rare mutations, potentially of digenic or oligogenic causality, avoiding detection. If so, Categorisation of autism will continuously reduce the variance eligible for adaptive explanation. However, further developments in molecular genetics have failed to find sufficient causal variants (Keller, 2018), and indeed find distinct genetic underpinnings for cases of autism with distinct phenotypic signatures: common SNPs being associated with higher intelligence, less developmental delays and more core autism features whilst de novo mutations associate with lower intelligence and more developmental delay (Warrier et al., 2022). This supports the longstanding recognition

that familial cases associate with normal or higher intelligence (DeLong & Dwyer, 1988), and implies that explaining autism-causing alleles requires reference to both non-adaptive and adaptive explanations for different alleles.

4.3/ Environmental Categorisation

Categorising a Described trait as a disruption by environmental evidence requires certifying that an environmental factor is causal without acting via an adaptive mechanism. Reactions to environments can often be fitness-enhancing examples of adaptive plasticity. Evidently fitness-reducing environmental factors (e.g. bullet wounds) can be causal of adaptations (e.g. blood clotting and healing), so identifying harmful environmental causes is not sufficient to prove any downstream effects are fitness-reducing disruptions. It is seductive to believe that environmental influences we see as 'bad', or pressures such as stress and trauma which cause harmful phenotypes, are self-evidently diseases, but that needn't be the case (Belsky, 2012). Even in cases harming health and fitness in a particular individual (possible through chronic exposure (e.g. to stress) (Beauchaine et al., 2011) or unusual underexposure (Seery et al., 2010)), such maladaptive manifestations can be by-products of adaptation and require adaptive explanation.

The critical question is whether the precise aspect of the phenotype being Described (for further discussion see Hunt, 2023) and resulting from the environmental factor is ontogenetically mediated and the mechanism could plausibly result from a history of positive reproductive success (bleeding from a bullet wound could not; blood clots could). Cases of environmental causes which Categorise as disruption include physical trauma, infections and toxins, and various widely recognised extrinsic causes of disease. Categorising by environment therefore ideally requires identifying environmental factors (likely correlating with exposure prior to Described trait appearance) and assessing the pathway of causation.

4.3.1/ Environmental Categorisation: Autism

Modabbernia, Velthorst and Reichenberg (2017) reviewed meta-analyses and systematic reviews of environmental risk factors for autism. Perinatal and neonatal risk factors (Gardener et al., 2011) include birth injury or trauma (relative risk (RR): 4.90), maternal haemorrhage (RR: 2.39) and umbilical cord complications (RR: 1.5). There is some evidence that heavy metals, most importantly inorganic mercury and lead, play a role in some cases. Valproate use during pregnancy causes foetal valproate syndrome, which regularly meets autism diagnostic criteria (hazard ratio of valproate use during pregnancy: 2.9-5.2; Christensen *et al.* 2013). These types of cause can be comfortably Categorised as disruptions; they involve identifiable pathological processes with large individual effects and no potential functional history or direct inheritance. There are, to our knowledge, no good

summary estimates of the proportion of autistic cases which would be Categorised via these mechanisms, but they are small in comparison to genetic factors.

Another key environmental factor associated with autism is increasing parental age: every 10 year increase in age increases ASD risk by about 18% for mothers and 21% for fathers (Wu et al., 2017). Fathers over fifty have more than twice the chance of having an autistic child to fathers under thirty (Hultman et al., 2011). This affects autism without intellectual disability as well (Croen et al., 2007; Tsuchiya et al., 2008). There have been attempts to find disease processes behind this effect. Older parents give birth to children with more mutations, but *de novo* mutations cannot account for most of the increased risk (Gratten et al., 2016; Taylor et al., 2017). Alongside other subtler risk factors, such as maternal infection during pregnancy and maternal obesity, the obscurity of these pathways and the possibility they result via positively selected heritable mechanisms means these additional cases can't be Categorised as non-adaptive without further research.

4.4/ Concluding Categorisation

Categorisation sorts out a major fundamental form of causal heterogeneity, distinguishing Described trait variance explainable by dysfunction and non-adaptive processes from variance requiring adaptive explanation and requiring further analysis.

In the case of autism, broad diagnostic criteria capture disparate phenotypes and etiologies. Social and behavioural capacities are clearly highly sensitive to interruption via various pathways. Genetic syndromes linked to specific CNVs or rare or *de novo* mutations, and cases caused by early life traumas or prenatal toxin exposure are obvious cases of autism which Categorise as disruptions. Less common individual cases of unusual brain development or other discoverable pathologies may exist and could Categorise as disruptions. Confirmable disruptive cases are more often associated with intellectual disability and co-occurring health conditions such as epilepsy (Lord et al., 2021). They likely account for between 5-20% of currently diagnosable autistic cases, although this is highly dependent on inclusion criteria, which has recently been expanding to include less intellectually-disabled cases (Lyall et al., 2017). The majority of cases of autism, however, Categorise as requiring adaptive explanation of some sort. This is the task of the remainder of the DCIDE method.

5/ DCIDE: Integrate

Integration aims to hone Description upon traits which are adaptations themselves. Reproductively successful heritable traits are not necessarily adaptations. Evolutionary medicine and psychiatry recognise various by-products of adaptations causing harm and diagnosed as disorder. Adaptive alleles can cause various harmful by-products: via costly

trade-offs; mismatch to novel environments; maladaptive extremes of adaptive spectrums; maladaptive outcomes of developmentally plastic adaptive systems; pleiotropic effects including sexual antagonism; heterozygote advantages; and more (Abed and St John-Smith 2022; Hunt 2023).

By-products gain fitness via related adaptations without lending positive fitness themselves. They can be common, heritable, and in themselves harmful, but persist and be positively selected nonetheless, especially if the benefits of a related adaptation historically outweighed their harms and selection couldn't (or randomly didn't) mitigate them. They could be caused by common alleles, or universal amongst a species. Their costs may incline us towards diagnosing them as simple disorder, whilst showing no traditionally recognised signs of pathology – they can be direct products of healthy function – confusing normal biomedical investigation.

	Area of Evidence	**Indications of high visibility to selection**	**Indications of very low visibility to selection**
Phenotype features	Age of Effect	Active before and in prime years of fertility and child-raising	Effects in old age, especially after grandparental support declines
	Duration	Long duration	Duration too insignificant to affect reproductive success
	Sex Differences	Present in both sexes	Low visibility in one sex; none in the other
Causal factors	Environmental plasticity	Appears and reacts under evolutionarily conserved environmental features	Contingent on evolutionarily novel environmental features
	Prevalence of genetic propensity	High prevalence	Too rare to exist in most hunter-gatherer agglomerated-band social groups

Table 2: Visibility variables with their implications of high or very low visibility to selection. Note that environmental visibility variables differ substantially from environmental biomedical variables in their lack of reference to biological mechanisms. Phenotype features are combinatorial (e.g. longer duration means more visibility to selection, but not if only in old age).

Adaptationist accounts are necessary for explaining by-products. If anxiety disorders are by-products of adaptive anxiety, explaining anxiety disorders requires referencing adaptive anxiety. Explaining obesity requires understanding the adaptation of fat storage. Explaining sickle cell disease requires understanding malaria resistance. Where disorder is by-product, the related adaptation needs identification for explanation. Problematically, disorder-focused Description may include by-products but exclude relevant adaptations (e.g. Describing sickle cell disease ignores heterozygous individuals), making explanation impossible.

To mitigate the possibility of Description criteria containing by-products and excluding relevant adaptations, Integration assesses a trait's *visibility to selection* (the extent to which it can affect reproductive success) via various areas of evidence (Table 2). These 'visibility variables' allow inference of selective pressures applied to a given phenotype. Very low visibility in any area implies adaptations could be missed, and so an Integrated trait including genetic and phenotypic correlations with greater visibility to selection should be analysed along with the originally Described trait (Figure 3). The phenotype predominantly visible to selection needs consideration for effective evolutionary explanations.

Figure 3: Genes have multiple effects across the visibility variables, and Integrating phenotypes most visible to selection provides the appropriate target of adaptive explanation. Dark coloured ellipses indicate targets of high visibility – see Table 2 for details.

The precise process and extent of Integration depends on the case (and practical aspects such as available evidence and resources) but aims to recognise the causal genes' phenotypic traits major visibility to selection. This could utilise multi-trait genetic correlations (e.g. Integrating schizophrenia, bipolar, schizotypal and subclinical cases into a general psychosis spectrum (Guloksuz & Van Os, 2018)), shared cellular mechanisms (e.g. Integrating obesity and normal fat), or correlations over the life course (e.g. Integrating cholesterol and lipid levels throughout life with Alzheimer's and heart disease risk in later life, as related to APOE4 mutations (Garcia et al., 2021)). This is not arbitrary expansion: if alleles predispose to various phenomena, their evolutionary explanation is necessarily related, so the Integrated trait is the proper object of explanation. Not every correlation needs Integration, only correlated traits with significant visibility.

It is possible that even with attempted Integration, Described traits do not meet minimum threshold to be visible to selection – for example, some heritable rare condition occurring in 1/10,000 individuals in the population (Hunt & Jaeggi, 2022), without evidence allowing Categorisation as disruption or an identifiable genetic or phenotypic relationship with other traits visible to selection. Unless it is a recently arising adaptation, the implication is that the trait is neither by-product nor adaptation, but a disruption with mysterious aetiology. Rare diseases and various neutral or stochastic processes could lead to such phenotypes. As in Categorisation, these can be excluded from eligibility for adaptive explanation.

5.1/ Integrating by Age of Effect

Natural selection acts on reproductive success, so the ages leading up to and during reproductive potential are the most important i.e. youth and early-to-mid adulthood (G. C. Williams, 1957) are where visibility to selection is highest. The relative importance of early life is so high that deterioration in aging could be a by-product of genes increasing early life fitness, i.e. antagonistic pleiotropy (Austad & Hoffman, 2018). The longer life goes on, the higher the likelihood of prior reproduction and cumulative extrinsic mortality risk, reducing that age's visibility to selection.

Generally, seeking earlier correlates of late-onset traits optimises analysis for visibility to selection. The likelihood of a Described trait being an adaptation drops the further it appears after reproductive age. Grandparents can increase their inclusive fitness, so traits appearing in post-reproductive age may be adaptations in exceptional cases (e.g. reproductive senescence allowing grandmothering (K. Hill & Hurtado, 1991)), but Integration of effects during youth and prime reproductive age is usually necessary (Brown & Athena Aktipis, 2015).

5.1.1/ Integrating by Age of Effect: Autism

Autism is an early-onset developmental condition, present essentially from birth, with lifelong duration (Lord et al., 2021). Relevantly to later Evaluation of function and by-product effects, symptoms often change over the lifespan (Waizbard-Bartov & Miller, 2023), but as age of effect makes it almost completely visible to natural selection, it doesn't require Integration of earlier correlates.

5.2/ Integrating by Duration

Integrating correlates by duration encourages the inclusion of longer-lasting phenotypes. Described traits could be invisible to selection if they are sufficiently short-lasting (depending on the phenotype this may be minutes or months). Stochastic temporary bodily or psychological phenomena may occasionally arise as by-products. For example, adaptive capacities for fear can result in startles when random harmless sights or sounds are briefly misinterpreted as representing dangers. Integrating correlated traits which last long enough to have significant effects on reproductive success increases the likelihood of including relevant adaptations.

5.2.1/ Integrating by Duration: Autism

Autism is associated with cognitive and behavioural effects of various durations (e.g. bursts of behaviour in meltdowns, phases of obsession, lifelong difficulties in certain specific social domains (Lord et al., 2021); also see §7.3.1). Overall, this makes it highly visible to selection by duration criteria and doesn't need Integration of longer-lasting correlates.

5.3/ Integrating by Sex Differences

Selection upon (non-Y chromosome) genes occurs in both sexes. Identical DNA can lead to similar or different phenotypes between sexes, and constraints can result in manifestation of sexually selected traits in both sexes despite positive selection in only one sex (Wittman et al., 2021) (when the manifestation in the other sex is costly, we label it 'sexually antagonistic'). Sexual selection is unlikely to optimise phenotypes to the extent that the sexually selected trait only has effects in one sex (Lande, 1980; T. M. Williams & Carroll, 2009), so by-product effects are expected.

Sex difference Integration forces consideration of whether sex-biased traits are the result of genes with different effects in the opposite sex, and attempts to include the manifestation between sexes. If single-sex traits with low visibility (e.g. appearing late in life in females) are Described, possible higher visibility in the other sex should be carefully assessed (e.g. it's possible late-life female manifestation is a by-product of early-life male adaptations).

5.3.1/ Integrating by Sex Differences: Autism

Autism shows a pronounced sex ratio difference, with estimates of a 3:1 male to female ratio (Solmi et al., 2022). However, females may be more likely to 'camouflage' and show different symptom profiles which are then underdiagnosed (Hull et al., 2020). Indeed, females require a higher polygenic score to meet autism diagnostic criteria (Warrier et al., 2022). Selection on autism acts on the sum of the male and female manifestation, even if male presentation meets diagnostic criteria more regularly, so Integrating camouflaged female cases would better reflect total phenotypic impact of autism genes for informing adaptive explanations.

5.4/ Integrating by Environmental Plasticity

Environmental factors leading to trait manifestation imply trait visibility to selection if they are common and ancestrally relevant (e.g. stress, diet, infection). Traits appearing specifically in evolutionarily novel environments (e.g. high sugar diets, reading materials) cannot be an adaptation for that specific environmental factor. Disorders appearing in these specific circumstances are products of mismatch (Chaudhary & Salali, 2022; Lea et al., 2023). When affected by novel environments, Integrating the relevant adaptation requires identifying the system's manifestation as visible to ancestral selection: its ancestral reaction norm. Obesity only exists in high numbers in evolutionarily novel contexts, so fat tissue growth in response to excess calories is the relevant plausible adaptation.

5.4.1/ Integrating by Environmental Plasticity: Autism

The question of whether autism exists or presents differently outside of novel environments is essentially unresolved because of lack of research (Davison & Gurven, 2021) in small-scale societies. Between industrialised societies autism prevalence differs quite significantly (Solmi et al., 2022), but this is likely due mainly to diagnostic uncertainty and cultural differences in condoned behaviour (Zaroff & Uhm, 2012). Further cross-cultural research would help in confirming whether and how autistic traits show visibility across cultures (and, presumably, evolutionary history). Otherwise, advancing parental age is a known prominent relevant factor (see §4.3.1), but as hunter-gatherer adults do have children into late life and autistic children are born to young couples, this doesn't imply autism is a product of mismatch or force Integration.

5.5/ Integrating by Prevalence of Genetic Propensity

Adaptive genotypes often reach fixation and become species-wide, which are clearly visible to selection. However, important counter-examples exist of adaptive strategies via individual differences, especially in cases of social selection, social niche specialisation and negative-frequency dependency (Hunt & Jaeggi, 2022; Martin et al., 2023), which can be

maintained at the genetic level by balancing selection (Wolf et al., 2007). Such strategies cannot be arbitrarily rare: an approximate 'minimum adaptive prevalence' is of one individual per social group (Hunt & Jaeggi, 2022).

Integration already seeks relevant within-individual effects of genotypes by considering duration and age of effect, but in cases of heritable individual differences, between-individual effects are also important. Selection acts on the total effects of genes across a population (inclusive fitness), so their visibility to selection depends on the sum of their common and rare manifestation. Higher prevalence indicates higher visibility to selection, so Integrating more common correlates of rare traits is a useful strategy for enabling adaptive explanations – Description criteria may be excluding adaptations. This may be particularly relevant for explaining rare psychopathological traits with more common heritable subclinical correlates. Simplistically, an allele with certain effects in 5% of individuals and different (perhaps more extreme or harmful) effects in 1% of individuals is five times more visible to selection in the 5% (although specific phenotypic effects may mean stronger selection in the rarer cases).

5.5.1/ Integrating by Prevalence of Genetic Propensity: Autism

Recent epidemiological studies of autism find a prevalence of around 1% (Fombonne et al., 2021) although rates differ widely by study, ranging from 0.01% to 4.36% (Zeidan et al., 2022). In the UK, rates increased 787% between 1998 and 2018 (Russell et al., 2021), primarily because inclusion criteria loosened. Cases included now are disproportionately without intellectual disability (Lyall et al., 2017), and more likely to Categorise as requiring adaptive explanation.

At the 1% rate, autism meets minimum visibility criteria – given observed group size of approximately 165 (Hamilton et al., 2007), about one or two of these autistic individuals are on average expected in every extended hunter-gatherer social group (considered further in §7.6). However, although adaptations can plausibly persist in around 1% of the population, they are more likely to do so at higher prevalence, implying Integration of more common genetically correlated traits is necessary. This could include cases incorporated in more permissive inclusion criteria and subclinical autistic traits distributed throughout the population, especially in family members, which has led to the identification of a 'Broad Autism Phenotype' (BAP) in 5-9% of the population (Morrison et al., 2018; Sasson et al., 2013). The genetic basis for autism and the BAP is shared, so capturing the relevant selected phenotype requires Integrating this wider autism spectrum – it is possible that adaptation visible in the BAP explains autism as a by-product.

5.6/ Concluding Integration

Categorisation distinguishes non-adaptive cases from traits requiring adaptive explanation using biomedical variables. Integration expands Described traits to incorporate relevant correlated phenotypes using visibility variables. After Integration the paradox of common, harmful, heritable mental disorders is refined into the paradox of mental disorders with early onset, long-lasting, ancestrally-activated, common, heritable correlates, with no discoverable pathology.

In autism's case, its age of effect and duration imply high visibility to selection. Environmental evidence doesn't suggest complete novelty. However, its prevalence and sex differences imply current criteria may be missing relevant related phenotypes, so the Integrated trait should include subclinical cases – both camouflaging cases in females and broader autism phenotype traits in family members. Their shared genetics require shared evolutionary explanation.

Unfortunately, prior research into autism largely hasn't Integrated autism's subclinical or female manifestations, which are under-researched (Hull et al., 2020). Studies utilising diagnostic manual criteria (even with more recent permissive inclusion criteria) may exclude adaptive phenotypes, obscuring explanations. This makes analysis more difficult. Nonetheless, to exemplify the DCIDE method, Depiction and Evaluation will proceed, recognising the limitations of available evidence.

6/ DCIDE: Depict

The fundamental evolutionary question is why biology has its particular form. Differential fitness of varying heritable phenotypes is Darwin's (1859) answer. Differential fitness can be a by-product of other evolutionary processes, or it can result from functions of those heritable phenotypes (Figure 1). The major aim of evolutionary explanations is describing the processes of function which led to the form of the present-day adaptation. Description, Categorisation and Integration have honed the target of explanation to optimise for this analysis. In Depiction, explanatory hypotheses are presented.

There is a rich history and varied methods across evolutionary biology, psychology and anthropology in developing functional hypotheses for observed phenomena, often with different goals in mind. For example, a significant aim for some evolutionary psychologists is to describe the evolved cognitive architecture of the mind in terms of information-processing mechanisms shaped for specific tasks (Cosmides & Tooby, 1994). The DCIDE method doesn't have an explicit aim beyond providing a structure for incorporating available evidence in assessing adaptive hypotheses. As a minimum criterion for a Depiction, some functional processes (of gene-, individual- or group-dynamics) explaining the reproductive success of the Described and Integrated traits are sought. When medical or

psychiatric conditions are under consideration, functional effects should plausibly justify any by-product dysfunctional effects.

Hitherto evolutionary hypothesising, on psychiatric traits and elsewhere, has often referenced varied evidence in an unprincipled way, potentially cherry-picking and becoming open to 'just-so storytelling' critiques. The DCIDE method tackles this by systematically referencing 'direct evidence' in Depiction and 'circumstantial evidence' in Evaluation. Direct evidence is supposed to be of the function in action; the eye seeing, the wing flying, the jealousy causing mate guarding. This evidence is often sought by evolutionary psychology (e.g. in experiments designed to reveal the mechanics of cognitive faculties such as anger (Sell et al., 2017)), or evolutionary biology (e.g. in observing that peppered moth wing colour affects survival (Majerus, 2008)). The key aim of Depictions is to reference this direct evidence of the functioning adaptation, proposing it contributed to positive fitness and thus caused the observed traits via natural selection. Varying relevance of potential forms of direct evidence and the circumstances of evidence collection are noted in Figure 4. In Evaluation, Depicted hypotheses will be systematically compared and assessed using circumstantial evidence, potentially including biomedical and visibility variables, which have specific inferential connotations. Hypotheses are eventually judged on their sufficiency to explain this totality of evidence. Restricting cited evidence to direct and circumstantial evidence helps parse the potentially infinite studies characterising phenotypes to retain those most useful for adaptive hypotheses (e.g. avoiding including every survey ever conducted on autistics).

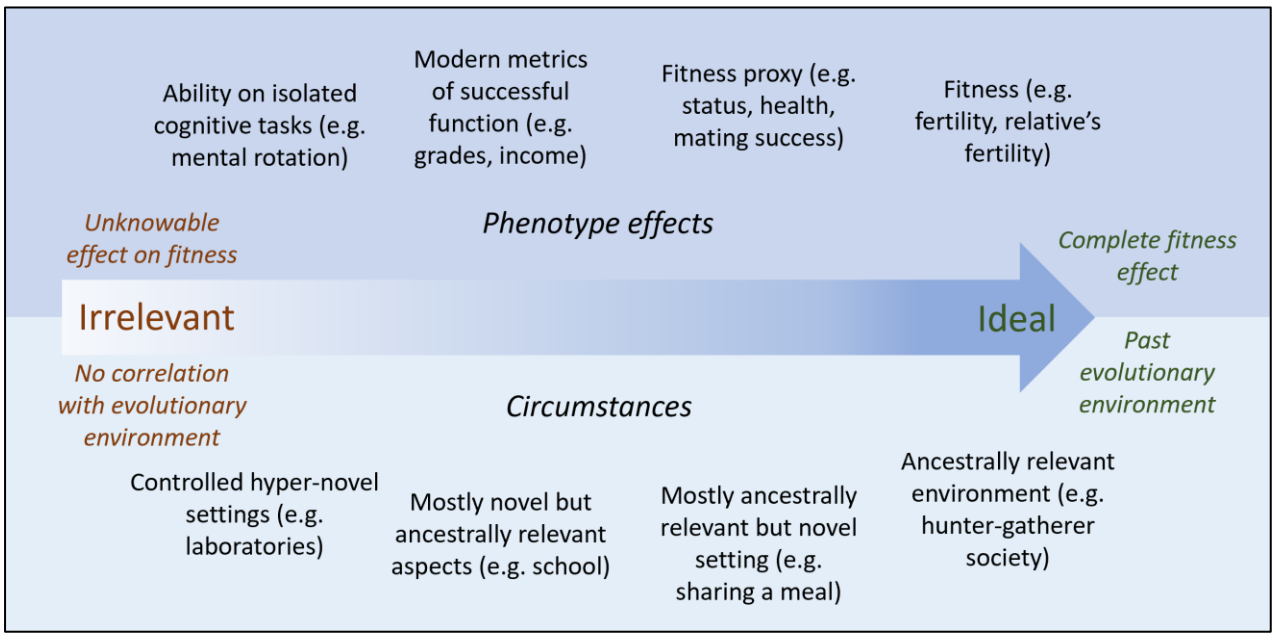

Figure 4: The varying relevance of direct evidence of functioning traits, also considering the circumstances within which that evidence is gathered.

This basic structure of analysis could be adapted to support or question existing hypotheses in evolutionary psychology or pre-existing hypotheses of mental disorders. The multitude of proximate biological or psychological accounts (e.g. referencing lower oxytocin levels (John & Jaeggi, 2021) or weak central coherence in autism (Frith & Happé, 1994)) would, however, need expansion to explain heritable persistence in the species. Pre-existing adaptive accounts are disadvantaged if they relied purely on diagnostic criteria, lacking Categorisation of subtypes or Integration of correlated phenotypes. In autism's case, various accounts have recognised dysfunctional subtypes and the heritable spectrum (informally applying Categorisation and Integration). However, the wider literature of evidence they draw from relies on varying inclusion criteria, limiting interpretability of the analysis. Future research should correct this.

6.1/ Depict: Autism

Hypotheses presented to explain autism are numerous. Any of them should ideally explain all relevant available evidence, regardless of their disciplinary origin. Classic hypotheses (e.g. of refrigerator mothers; Kanner 1968) or recent neurobiological hypotheses (e.g. Markram and Markram 2010) may equally be Depicted and Evaluated if expanded to address the evolutionary paradox of autism's persistence.

Here we Depict three prominent accounts (Baron-Cohen, 2008, 2020; Crespi, 2016; Del Giudice, 2018a) referencing evolutionary function in their explanation of autism. They agree on symptoms of social disability as costly by-products, but specifics regarding functional processes and evolutionary forces differ.

Baron-Cohen (2008, 2020) frames autistic traits as an exaggeration of 'if-then' systematic thinking, which exists on an empathising-systemising dimension. Autistics are hypoempathisers, lacking various empathic abilities (Baron-Cohen, 1995) including complex emotion recognition (Baron-Cohen, Wheelwright, Hill, et al., 2001), recognition of faux pas (Baron-Cohen, O'Riordan, et al., 1999) and spontaneous ascription of internal states (Castelli et al., 2000), but hypersystemisers, with associated advantages in folk physics (Baron-Cohen, Wheelwright, Spong, et al., 2001) and attention to detail (O'Riordan et al., 2001), sometimes allowing then to achieve high levels of domain expertise in fields like mathematics, physics or computer science (Baron-Cohen, Wheelwright, et al., 1999). Obsessions cluster in systemising domains (Baron-Cohen & Wheelwright, 1999). Regarding the broader autism phenotype he points to evidence of relatives of autistics being engineers, mathematicians and scientists (Baron-Cohen, 1998; Baron-Cohen et al., 1997), strong systemisers (Baron-Cohen & Hammer, 1997), and that scientists and mathematicians score higher on the Autism Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, et al., 2001). Baron-Cohen's general hypothesis is that high systemisers filled an ancestral social niche as tool-makers, inventors and experts in areas of their obsession. He accounts for

disabling cases meeting autism diagnostic criteria (without Categorisable pathology) as resulting from systemising overexpression and a lack of empathising, perhaps due to assortative mating.

Del Giudice (2018) explicitly addresses autistic heterogeneity and the necessity of differentiating cases caused by mutations and associated with intellectual disability from cases caused by common alleles, also citing evidence that those common alleles are associated with higher IQ and educational attainment in the general population (Clarke et al., 2016; Hagenaars et al., 2016). He references Baron-Cohen's systemising account of autistic strengths and talents, adding that autistic symptoms relate to a logical, deliberative style of reasoning (Brosnan et al., 2014), and embellishes on the account by arguing for autism as a slow life history skilled/provisioning reproductive strategy, particularly for males. Elsewhere he suggests systemising abilities would be sexually selected as attractive traits in males (Del Giudice et al., 2010), and that autistic-like traits delay reproduction and increase potential investment in family, noting that autistic-like traits correlate with restricted sociosexuality, preserved or heightened interests in romantic relationships, and increased investment in long-term partners. Maladaptive cases again arise from overexpression, exacerbated by assortative mating, mutations and environmental insults. Baron-Cohen and Del Giudice therefore agree on autistic traits as cognitive specialisations (Hunt and Jaeggi 2022) but differ on specifics of function.

Crespi (2016) provides a 'high intelligence imbalance' hypothesis, where strong recent selection for intelligence in humans leads to occasional dysregulation and autism as a by-product. Specifically, he relates autism to exaggerated 'perceptual' intelligence from Johnson and Bouchard Jr.'s (2005) Verbal-Perceptual-Rotational model of intelligence. He cites evidence of autistic reductions in verbal skills but increase in focus of attention (Ploog, 2010; Sabatos-DeVito et al., 2016) enhanced perceptual and spatial abilities (Muth et al., 2014) and superior ability in non-rotational aspects of the mental rotation task (Zapf et al., 2015). He further notes that autism genetically correlates with IQ (Clarke et al., 2016), childhood IQ, college attendance and years of education (Bulik-Sullivan et al., 2015), cognitive function in childhood and educational attainment (W. D. Hill et al., 2016), and verbal-numerical reasoning and educational level reached (Hagenaars et al., 2016). Performance on the Iowa Gambling task of decision-making is enhanced (South et al., 2014). Sensory discrimination and acuity are enhanced across auditory (Stanutz et al., 2014) visual (Muth et al., 2014) and tactile (Nakano et al., 2012) domains. He notes that family members of autistics share cognitive strength profiles (Gizzonio et al., 2014; Noland et al., 2010). These strengths are paradoxical in the face of apparently lower autistic IQ, on average, on most standardised tests. Referencing various neurological similarities in brain size and growth, connectivity and neuronal function between high intelligence and autism, he hypothesises that instances of autism are costly by-products of general selection for

intelligence in humans. Thus, despite noting the same cognitive strengths as relevant to adaptive processes causing autism, the hypothesis doesn't claim that autistic or broad autism phenotype individuals are filling a particular social niche.

Author	Direct Evidence	Adaptive explanation
Baron-Cohen	Lack social abilities but excel in systemising abilities; family members often successful in systemising careers	Systemising ability filled an inventing/expert social niche; systemising overexpressed without balancing empathising abilities in individuals with autism
Del Giudice	Systemising abilities; rational thinking; tendency towards committed, long-lasting relationships	Slow life-history skilled/provisioning strategy, with systemising strengths and altered relationship patterns as valuable for reproductive success and inclusive fitness
Crespi	Enhanced abilities in specific areas of intelligence; particular benefits of correlated traits across the population	Strong recent positive selection on human intelligence, cases of autism as a by-product via overexpression and vulnerability

Table 3: A summary of competing Depictions of autism direct evidence and adaptive explanations

7/ Evaluate

Description, Categorisation and Integration hone a target of adaptive explanation; Depictions attempt to explain it, and need Evaluation. The goal of Evaluation is deciding the optimum coherent explanation given the available evidence. A systematic standardised method is critical here, to avoid just-so storytelling or cherry-picking, and allow fair comparison of hypotheses.

Evaluation proceeds in three steps (Figure 5). The first two are standard existing scientific practise not requiring extended exposition. Firstly, asking about *reliability*. Do Depictions reference reliable direct evidence (e.g. reproducible, replicable, methodologically sufficient, evolutionarily relevant) and reliable explanatory theory (e.g. doesn't rely on naïve group selection (Okasha, 2001))? Secondly, asking whether competing Depictions' direct evidence is *compatible*. Does any referenced reliable evidence falsify a competing hypothesis (e.g. one hypothesis claims jealousy enhances affectionate attachment but another cites evidence that jealousy causes fear and reduces affection)? Once viable Depictions remain, the third, more novel stage of Evaluation progresses. Do the Depicted hypotheses sufficiently explain the circumstantial evidence? Existing evolutionary hypotheses have often cited areas of circumstantial evidence (e.g. male preponderance and low prevalence to justify sociopathy as a cheating strategy (Mealey, 1995) or cross-culturally shared sex

differences to justify jealousy as a mate guarding and retention strategy (Buunk et al., 1996)) but the DCIDE method's Evaluation assesses such evidence systematically, utilising standardised implications from each area for explanation via functional or by-product explanations (Table 4).

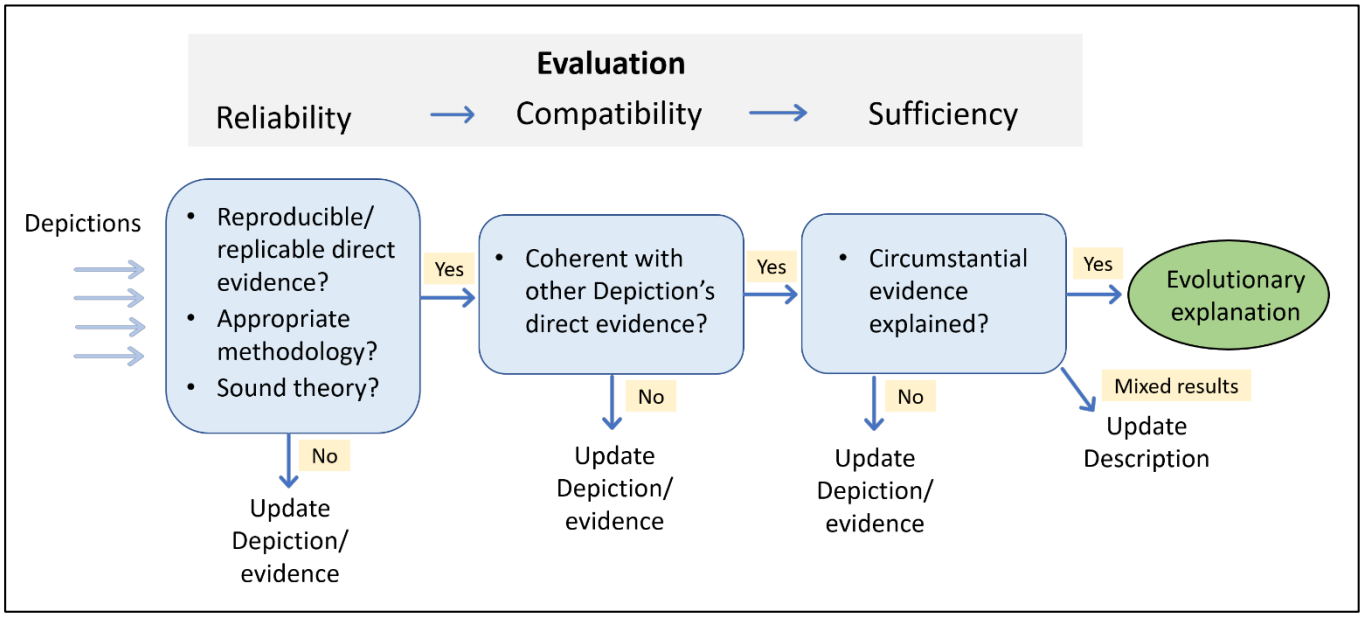

Figure 5: Three steps for Evaluation. Possible outcomes of each step, represented by arrows, can occur simultaneously (e.g. a Depiction can mostly explain the circumstantial evidence, and to some extent be explained, but also show mixed results requiring updating of Description)

Area of Evidence	***Functional explanation***	***By-product explanation***	***Example considerations***
Age of Effect	Adaptation for that stage of life history	Appears at ages when non-functional	Pre-reproductive age; adolescence and early reproduction; prime reproductive age; grandparental alloparenting
Duration	Adaptation functions for this duration	Has non-functional effects	Lifelong traits; occurrence over long (e.g. life stages), medium (e.g. seasons) or short term (e.g. immediate events)
Sex Differences	Sex specific function	Selected for in other sex; consequences of other sex effects	Arising in sexual-selection specific circumstances (e.g. sexual competition); infidelity; sexual jealousy; sexual attraction; childcare
Environmental plasticity	Functional reaction; adaptive	Non-adaptive consequences of plasticity;	Presentation cross culturally. Effects of ancestrally relevant factors (e.g. diet; illness; stress; social status;

	developmental plasticity; degree of plastic strategy	mismatch	relationship status; interpersonal conflict; entrapped; danger; opportunity)
Prevalence of genetic propensity	Functional strategy optimised for a particular prevalence within a group	Appears in cases and groups where non-functional	0.5-1% one individual per extended social group; 3-5% one individual per band; 10-20% one individual per family/friend group, 50% one in a pair, 100% everyone. (Intermediary figures unclear)

Table 4: Examples of circumstantial evidence and the general Evaluative implications for each area of evidence to be explained as either functional or by-product.

Circumstantial evidence can subsume all non-direct evidence associated with Described and Integrated traits, but ideally they should have standardisable interpretations (e.g. Tables 1,2 and 4). The critical question is whether Depictions sufficiently explain that circumstantial evidence: do they explain the age of effect, duration, etc. Visibility variables alone are considered as relevant circumstantial evidence in this paper, but biomedical variables could also be utilised (e.g. genetic selection histories implying environments of adaptiveness; observed pleiotropic effects via mechanisms such as cytokines implying trade-offs). Complexity in psychiatry of co-opted pathways and genes, multi-effect systems and complicated network interactions mean that present usefulness of biomedical evidence in Evaluating autism's adaptive explanation is limited, but relevant brief mention of biomedical variables are made in conclusion.

Visibility variables are circumstantial facts applying to every phenotypic characteristic. Function depends upon visibility: a function can only occur to the extent that it is visible (e.g. at a particular age, for a particular duration). By-products are visible, but that visibility needs to be excused as non-functional. Advantageously, visibility variables have fairly standard implications (e.g. traits appearing for ten minutes versus several years, or in early versus late reproductive age, are eligible for specific functional or by-product explanations – Table 4). This requires framing in terms relevant to human evolution rather than modern descriptive norms: for example, age of effect should be considered in terms of different human life stages instead of decade (Haig, 2010); prevalence in terms of absolute numbers per human ancestral group size instead of percentages (Hunt & Jaeggi 2022). Depictions can be compared in their sufficiency of explaining the specific visibility of Described and Integrated traits; the correct hypothesis should explain them coherently. Insufficiency here should appear either as lacking explanation or just-so storytelling, excessively wrangling the

explanation to fit the evidence without prior theoretical grounds. The correct hypothesis should find each circumstantial fact justifying rather than challenging.

Three main possibilities arise during Evaluation (Figure 2). Described and Integrated traits may be considered sufficiently explained; further evidence or development of hypotheses required; or the results mixed (likely if attempting to explain multiple adaptations simultaneously) and different Description required. These are not mutually exclusive. Part of the explanation may be satisfactory, other questions left unanswered, and some subset of cases not fit that explanation, requiring re-Description.

In this paper, Evaluation has important limitations. Neither empirical research nor theoretical models have been formulated specifically for a DCIDE analysis (e.g. autism research has largely concentrated on pre-reproductive age; Integrated traits are rarely considered). Depiction's accounts of evidence are necessarily assumed if unmentioned by the authors. The necessity of brevity restricts comprehensive analysis. Despite these limitations, exemplifying Evaluation is essential, and assessing Depictions of autism still leads to useful conclusions.

7.1/ Evaluate Depiction reliability and compatibility: Autism

Initially, Depiction reliability and compatibility must be assessed. It should be noted that most cited studies come from before the replication crisis (Shrout & Rodgers, 2018) and improvement in practises (e.g. open data and pre-registration), so Evaluation may change substantially if evidence is revised.

In terms of reliability, direct evidence of autistic strengths is plagued by complexity and heterogeneity between individuals, but holds up to meta-analysis (Muth et al., 2014). Reliance upon novel experimental tests implies questionable evolutionary-relevance (Figure 4), but relationships between successful technical careers and systemising (Baron-Cohen, Wheelwright, Skinner, et al., 2001; Baron-Cohen & Hammer, 1997) imply general aptitudes in valued areas with possible ancestral correlates. Expanding research to non-industrialised societies would dramatically improve confidence here. The only direct evidence of entirely questionable reliability is Del Giudice's (2010) evidence on autistic relationship and parenting strategies: a single study correlating autistic-like traits with a relationship questionnaire in 199 students, but the broad autism phenotype has also been associated with avoidant and less secure romantic attachment (Lamport & Turner, 2014). Due to this, and wider recognised difficulties with applying fast or slow life history strategies to human behaviour (Frankenhuis & Nettle, 2020), this element of Del Giudice's Depiction is too weakly supported to warrant full Evaluation at this time. With this aspect removed, his account more directly aligns with Baron-Cohen's, so for simplicity they can be considered equivalent. Evaluation can proceed comparing the systemising social niche specialisation with the high intelligence by-product hypotheses.

Both hypotheses rely on plausible evolutionary theory. Del Giudice and Crespi's publications specify evolutionary dynamics via verbal models; Baron-Cohen's are more implicit, not mentioning social niche specialisation or social selection explicitly, but providing the same basic argument (that high systemisers were valuable to the social group for their specific abilities). In terms of immediately accounting for one another's direct evidence, the overlap between hypotheses relieves potential conflict between them – their primary discrepancy is in the evolutionary model.

If direct evidence of fitness effects in evolutionarily-relevant environments existed, the key distinguishing finding would be of benefits or costs associated with increasing autistic traits; the by-product model predicting steadily increasing fitness costs as autistic traits increase, the social niche specialisation model predicting fitness benefits for some range of autistic traits in certain environments. Such evidence is not readily available, but Depictions can still be Evaluated for their sufficiency explaining the circumstantial evidence.

7.2/ Evaluate: Age of effect

The relevance of age of effect requires referencing average challenges facing different ages of humans over evolutionary history. Functions of adaptations should appear at appropriate times. For example, traits appearing early in reproductive age could function in courtship and initial acquisition of mates; later ages carry trade-offs between supporting existing offspring and investing in new offspring (Walker et al., 2007). If there is selection for some function at a particular age in life (e.g. high sociosexuality in early adulthood) by-products might appear throughout the life course, despite not being fitness-enhancing. By-product accounts could refer to costs of plasticity (e.g. if it is more costly than beneficial to be able to facultatively turn off high sociosexuality when fitness-enhancing benefits disappear), physical constraints (e.g. if high sociosexuality couldn't evolve restricted to early reproductive age) or trade-offs optimised for reproductive success early in life (e.g. if high sociosexuality causes health problems in old age). Depictions referencing such by-products would benefit if research reveals constrained physiological mechanisms.

7.2.1/ Evaluating Age of effect: Autism

Autism is often considered to have lifelong effects, but its specific manifestation often varies substantially by age. In infancy and childhood, social and behavioural difficulties may be quite severe, but during adolescence and adulthood many autistics, especially those of higher cognitive ability, reduce in symptomology (Waizbard-Bartov & Miller, 2023). Individuals scoring high in autistic traits often show reduced social difficulties over time (Riglin et al., 2021). This trajectory may relate to apparent advantages in autistic and high systemising individuals of accurately predicting social psychological phenomena (Gollwitzer et al., 2019), implying they recruit deliberation and reasoning abilities in social situations

instead of automatically empathising. This change in characteristics over time is important for adaptive accounts, because behaviours occurring during the period of pre-reproductive age only need to prepare an individual for reproduction, whilst behaviours at reproductive age more directly affect fitness.

Depictions by Del Giudice and Baron-Cohen relate to systemising and specialist skills serving useful social functions – the benefits primarily arising in adulthood, when their abilities are appreciated, leading to reproductive success. Baron Cohen proposes that hypersystemising in childhood makes exceptional ability more likely (Ch.2, Baron-Cohen 2020). This is coherent with slow development of systemising abilities and costs which alleviate with age. Crespi might account for this developmental trajectory as a by-product of intelligence imbalance slowly compensating over time. Still, harmful by-products are more likely to appear late in life, where they are less visible to selection, so some constraint explanation of why costly early-life autistic traits persist is required.

7.3/ Evaluate: Duration

Duration is most functionally informative if aligned with immediate environmental stimuli or life circumstances. Durations of adaptations should relate to their adaptive function: adrenaline's effects are short-lasting, love is longer-lasting, capacity for language longer, because of their respective functions. By-products could theoretically occur for any duration, but the specific duration should still be justified.

Described and Integrated characteristics often have different durations – for example, schizophrenia may consist of delusional thoughts and hallucinations for long periods, with brief episodes of intense psychosis requiring hospitalisation or increased medication; aggression may involve multi-year phases of heightened reactivity with occasional outbursts of actual violence. Simplistically, visibility to selection implies longer effects are more likely to be adaptations, contingent on phenotype specifics and strength of selection. Two days of severe anxiety may have a stronger effect on fitness than ten years of moderate anxiety if the two days involve life-saving behaviour.

7.3.1/ Evaluating Duration: Autism

Autism is characterised by general differences in cognition which seem long-lasting, especially in restrictive and repetitive behavioural domains, with more regular improvement in social functioning over time (Waizbard-Bartov & Miller, 2023). Characteristic behaviours also involve bursts of extreme behaviour (e.g. meltdowns, sensory overload) and periods of obsession with particular systems, subjects or activities which may last from days to years.

Systemising associates with differences in occupations, interests and hobbies across the lifespan (Svedholm-Häkkinen & Lindeman, 2016) which fits Baron-Cohen and Del Giudice's Depictions of high systemising filling a social niche. If periods of obsession lead to long-

lasting expertise, this would also support the hypothesis. Various shorter-lasting behaviours such as repetitive play, craving movement (e.g. spinning, jumping) and stereotypies (e.g. hand flapping) reduce with age (Mayes & Calhoun, 2011), and along with meltdowns and sensory overloads could be acceptable by-products. To Crespi, the duration of autism and Integrated traits is due to the developmental imbalance of intelligence which lasts for life. Bursts or phases of behaviour are by-products of this imbalance, presumably related to environmental circumstances. The long duration and thus high visibility are surprising for costly by-products.

7.4/ Evaluate: Sex Differences

Sex differences may relate to sex-specific functions of the characteristics in question or come downstream of other sex differences, environmental or biological. Functional accounts should explain differential presentation (either in prevalence or manifestation) as solving an adaptive challenge experienced uniquely or disproportionately by one sex. Differential challenges between sexes are a common topic of evolutionary psychology, involving intersexual or intrasexual competition, infidelity, courtship, mate guarding, attractiveness and more (Stewart-Williams & Thomas, 2013).

By-product approaches to explaining sex differences might reference adaptive sex differences with costs (potentially in either sex) or differential vulnerability due to other sex effects. As noted in Integration, biological constraints mean by-products in the non- or less-benefitting sex are expected. Biomedical variables may allow validation of such hypotheses (e.g. finding testosterone is critical to trait manifestation implies positive selection in males with by-product effects in females).

7.4.1/ Evaluating Sex Differences: Autism

As noted in Integration, autism sex ratios are male-biased at approximately 3:1, with female autistics more successfully camouflaging and requiring a higher polygenic score to meet diagnostic criteria (Warrier et al., 2022). 44% of general population males also show either a systemising or high systemising bias in the UK Brain Type Study (D. M. Greenberg et al., 2018) in comparison to 27% of females. Baron Cohen's (2008) account particularly concentrates on this difference, postulating that higher empathising in females relates to evolutionary pressures related to childcare, whilst systemising may help in male-biased activities of hunting, tracking and toolmaking and -use. High female empathising may also mean that being shifted towards a high systemising phenotype makes diagnosable autism more likely in males than females, because female empathising mitigates social disability (Baron Cohen, 2020). Camouflaged female cases are thus explained as a combined reduced benefit to females of systemising and increased benefit of empathising. The social niche specialisation hypothesis thus explains the general spectrum of sex differences as due to

systemising traits and niches as particularly fitness-enhancing for males. This explanation of the Integrated spectrum is more complete than Crespi's, which briefly mentions sex differences in autism rates as possibly related to reduced verbal and higher rotational intelligence in males.

7.5/ Evaluate: Environmental plasticity

Environmental effects can precipitate, prevent or alter presentation of traits – they affect phenotype manifestation within a reaction norm. Functional explanations of environmental effects should explain why this reactivity was fitness-enhancing. This can be obvious in short-lasting reactive states such as emotions. Suites of feelings, physiological arousal and behaviour arise and dissipate in particular circumstances of adaptive challenge – for example, anger arising to rectify social undervaluation (Sell et al., 2017) or fear for protection (Stankowich & Blumstein, 2005). For long-lasting traits, predictive adaptive responses (Gluckman et al., 2005) utilise environmental cues to provide information about future environments, allowing functional developmental trajectories via adaptive developmental plasticity (West-Eberhard, 2003). This may affect whether a trait appears at all (e.g. early life experiences causing phobias) or presentation of a trait (e.g. language capacity developing different languages and accents). Assessing the range of presentation of the trait in question allows observation of its fixed and flexible aspects, which inform functional hypotheses regarding the fixed and flexible aspects of the adaptive challenge. Cross-cultural research can be particularly important here.

By-product explanations need to account for specific effects as non-functional. These are expected – external influences have stochastic effects, and functional plastic responses need only be beneficial on average. Arousal of short-lived reactions can be inappropriate, and in longer-lasting plastic responses, developmental trajectories may end up non-functional or harmful in the future environment (so called 'developmental mismatch' (Gluckman et al., 2019)).

7.5.1/ Evaluating Environmental plasticity: Autism

Environmental factors precipitating cases of autism which require adaptive explanations are limited – advancing parental age is the clearest (Wu et al., 2017). Whether advancing parental age also increases systemising and BAP traits is unclear. If so, investigating functional explanations for why systemising traits are more beneficial for children born to older parents is prudent for the social niche Depiction (e.g. perhaps older siblings support less social younger siblings, mitigating harms). If not, undiscovered dysfunction is more likely, and both Depictions would need to explain this effect as a vulnerability trade-off. This awaits further research for clarity.

Autism, BAP and systemising presentation do depend on environment, however. Although autistic and systemising cognitive differences are fairly stable over the life course, the behavioural manifestation of systemising is environmentally dependent, concentrating on local lawful, repeatable, predictable systems following input-output-operation dynamics (Baron Cohen, 2008). Systemisers are 'pattern seekers' (Baron Cohen, 2020), but precise systemising interests may range between mechanical systems, weather, music, animals, and more. This fits social niche specialisation hypotheses, which necessitate flexibly adapting to local ecologies. Although research on autistic presentation in evolutionarily-relevant societies is lacking, Spikins, Wright and Hodgson (2016) note an ethnographic example of an unusually antisocial reindeer herder, highly valued for memorising parentage, medical history and character of individuals within a herd of 2600 reindeer. Similar modern success-stories are often in technical careers. Where such tendencies lead to socially valued expertise, this could be functional. If the BAP and systemising were costly by-products away from an optimum cognitive type, reliable tendencies to attach to potentially valuable local systems and gain above-average ability would be surprising. By-product explanations such as Crespi's must explain these tendencies as compensation strategies, not actually fitness-enhancing by being valued above neurotypical cognition.

7.6/ Evaluate: Prevalence of genetic propensity

Genotypes interact with environments in developing phenotypes, and the range of resulting phenotype expressions – the reaction norm – is the genetic effect which selection acts upon (Martin et al., 2023) and which must relate to functional explanations. Depictions should explain why heritable reaction norms don't exist at lower or higher frequencies.

Species-wide traits (e.g. eyes) or states (e.g. hunger) are eligible for functional explanations relating their species-wide positive effect on fitness. Species-wide by-products may be trade-offs of such adaptations, unavoidable biological constraints, or potentially non-adaptive yet drifted to fixation. Heritable traits below species-wide prevalence are more complicated. They may have fixated within sub-groups (e.g. of geographic location or sex), thus requiring population-specific functional explanations. They may result from 'reactive heritability' (Tooby & Cosmides, 1990), species-wide strategies arising due to other heritable phenotypes (e.g. extraversion arising in attractive people (Lukaszewski & Roney, 2011)). Alternatively, they may represent adaptive individual differences in functional strategy or by-products of such strategies (Figure 1). Functional genetic individual differences are mostly theoretically predicted and empirically observed in social animals with predictable social environments, where costs or constraints prevent plasticity, and social niches exist for alternative strategies (Hunt & Jaeggi, 2022). These likely require referencing negative frequency-dependent selection, social niche specialisation and social selection, with alleles under balancing selection.

When Evaluating hypotheses' sufficiency at explaining specific prevalence of genetic propensity, ancestral social group size is the relevant framing. Hunter-gatherer social structures are composed primarily of residential units ('bands') of individuals (mean size of 28) interacting frequently with other bands in metagroups ('tribes') with a geometric mean of 165 individuals (95% confidence limit: 152-181 (Hamilton et al., 2007)). Gender ratios are usually equal (though can fluctuate stochastically (Kramer et al., 2017)), with population composition about half adult, half children/adolescents (Kelly, 2013). Assuming these broadly represent the average group sizes and demographics of relevance to human evolution, we can infer prevalence per social group throughout human evolutionary history, with standardised functional implications – traits appearing, on average, in five individuals in every band are eligible for different explanations than traits appearing once per band. In general, higher prevalence indicates higher visibility to selection and thus higher likelihood to be an adaptation, but rare strategies may be adaptive at a minimum level of once per aggregated bands (approximately 1%) or once per band (approximately 4%) where specific social niches exist (Hunt & Jaeggi, 2022; Martin et al., 2023). Mealey (1995) argued, for instance, that sociopathy is an adaptive cheating strategy, optimal in around 1% of the population to avoid cheater-detection.

7.6.1/ Evaluating prevalence of genetic propensity: Autism

Autistic traits are distributed continuously throughout the population. Assuming high heritability implies similar cross-cultural rates of relevant heritable cognitive differences (although precise manifestation may differ), trait distributions should be Evaluated referencing hunter-gatherer social group size. The most extreme 0.6% are ineligible for functional explanation (see §2.3.4), but if identifying, for example, 3.14% of individuals (4.64% of boys) as autistic (Li et al., 2022), approximately one adult in every two or three bands would meet diagnostic criteria. Stricter criteria including 1% of individuals implicate one individual per tribe. At 5-9% of the population (Morrison et al., 2018; Sasson et al., 2013), one BAP adult per band is expected.

Del Giudice and Baron Cohen aren't precise about the adaptive range of autistic traits, but such frequencies are coherently explained if one BAP adult finds a social niche within most hunter-gatherer bands, with the possibility that manifestations meeting autism diagnostic criteria find niches in the wider social group. It's also possible band-level niches are advantageous enough to justify costly extremes manifesting more rarely in the wider social group – or indeed, if assortative mating is important, perhaps such phenotypes were essentially non-existent in the past, due to restricted mate choice. Success at different levels of group size could vary between environments and generations – the social niche hypothesis merely implies balancing benefits and costs which prevent persistent increase or decrease in frequency, aligning with various versions of functional explanations of the

observed prevalence. Crespi's model emphasises that autism-causing alleles increase intelligence and have been positively selected across the general population, but autistic trait distribution represents prevalence of costly by-products, persisting due to inefficient negative selection. This makes autism and BAP trait presence in every hunter-gatherer group surprising – the proposed dynamic is that the non-autistic intelligence increase is significant enough that it justifies a constant costly by-product – the persistent churn of selection for intelligence and against autistic individuals just happens to remain stable enough to maintain a rate of almost exactly one BAP adult per band.

7.7/ Concluding Evaluation

Evaluating the high intelligence by-product hypothesis against the social niche specialisation hypothesis with respect to the visibility variables, a general pattern emerges: for the by-product hypothesis, the circumstantial evidence is theoretically surprising and needs to be excused as a failure of natural selection, but for the social niche hypothesis, the evidence is generally expected. Note that if any single area of evidence was substantially different, the social niche hypothesis could be excluded: if prevalence too low, or onset too late, or duration too short, or phenotype not appropriately environmentally reactive. Instead, the evidence fits the Depiction, so we can Evaluate the systemising social niche specialisation hypothesis favourably.

Biomedical variables have been excluded from this Evaluation, which could unfairly disadvantage the by-product model if specific evidence was revealed, particularly directional positive selection for autism alleles associated with intelligence and neurological similarities between high intelligence and autism implying exaggerated features. Regarding selection direction, some research has found positive selection on autism risk variants associated with intelligence (Polimanti & Gelernter, 2017), but only 14 SNPs out of 446 showed such signals (Prakash & Banerjee, 2021). There are also inherent limitations to estimating allele frequency change over relevant time-scales due to small sample sizes of ancient human genomes. Crespi (2016) also cites various neuroscientific findings implying traits associated with intelligence are over-developed or over-expressed in autism, particularly concentrating on brain size and growth, brain connectivity, and neuronal function, which would cohere with the by-product hypothesis. However, sufficient tests comparing autistics and controls whilst looking at variation in intelligence are lacking. The strongest findings relate to brain growth, but one large project found only 15% of autistic boys and 6% of girls show precocious brain growth (Amaral et al., 2017), without excluding Categorisable non-adaptive cases. A very large (n=7005) study also found no significant relationship between autistic traits and subcortical region size (Sharp et al., 2023). Future findings might identify the exaggerating brain differences Crespi proposes. This wouldn't necessarily exclude the social niche specialisation hypothesis (they could be the mechanism of adaptation) but would

provide a mechanistic explanation of the high intelligence by-product, making the claims that selection has not or could not reduce autistic traits whilst maintaining high intelligence in the population more plausible.

Given this DCIDE review, the best evolutionary explanation for autistic traits, including diagnosable autism but excluding non-adaptive cases, is that advantages in systemising and related cognitive enhancements lead to individuals consistently filling valued roles throughout societies over human evolutionary history, with social disadvantages that came as trade-offs. The reason the human population differs cognitively upon the autism spectrum is because the benefits of autistic thinking are balanced by costs: only so many spaces exist for these types of minds before strengths are outweighed by weaknesses, so the individual difference persists.

Given the sufficiency of this explanation in accounting for the available evidence, it would be surprising if a competing Depiction usurped it, but certainly not impossible. Most plausible would be some alternative account of the precise role or niche that autistic or BAP individuals fill, perhaps concentrating on some non-systemising trait such as attention to detail or sensory sensitivities. The circumstantial evidence (especially prevalence) would be hard to explain with completely different evolutionary models to social niche specialisation, but the precise phenotypic traits of relevance, and the dynamics which occur regarding the positive and negative selection effects may differ. For example, although existing accounts concentrate on male preponderance and male niches, it's possible that autistic cognition is more positively selected in females (e.g. for food processing or tool-use skills) with males bearing costly by-products due to background higher systemising cognition. Direct evidence of social niche roles to decide between different versions of the hypothesis ideally requires more research in the least industrialised and small-scale societies (Syme & Hagen, 2019). What is likely is that continuing technological and scientific advances lead to identification of more cases Categorisable as disruption – but given the existing evidence, it is fairly implausible that this will extend to all autistic and BAP individuals.

Despite the general sufficiency of the systemising social niche specialisation hypothesis in explaining the direct and circumstantial evidence, some complexities should be noted. Firstly, not all 'autism spectrum disorder' or BAP individuals show the systemising behavioural phenotype core to the Depiction. This should warrant Evaluation as 'mixed results' – the original Description may be too broad, capturing multiple adaptations or by-products of adaptations, and benefit from tighter Description. Altering Description specifiers to differentiate high and low systemisers may result in a different adaptive explanation for low-systemising autistics. Evidence of distinct heritability between these groups would justify this separation. If so, new research utilising different 'autism spectrum disorder' inclusion criteria would be appropriate.

Furthermore, Evaluation doesn't narrow down the adaptive range of systemising-BAP-autism Integrated traits. The systemising social niche specialisation hypothesis fits the evidence, but exact functional explanations are not constrained to 'the BAP is adaptive, autism is not' or vice versa. This is acceptable – indeed, precise function may vary between generations and environments – only average inclusive fitness matters. Relatedly, although the high intelligence by-product explanation alone struggles to sufficiently explain the circumstantial evidence, positive selection on intelligence across the population may simultaneously be a factor increasing autism-predisposing alleles, accompanying balancing selection for systemising social niche specialisation. Functional accounts are often not mutually exclusive. Of the hundreds of autism-predisposing alleles in a population at any one time, their positive effects on reproductive success – their functions – could be primarily seen in high systemisers who find socially valuable roles, and partially in increasing intelligence in the general population. Also note that reaching such general adaptive explanations does not convey anything like complete explanations of individual phenotypes: every individual has a unique genetic and epigenetic background, inevitably carrying some mildly deleterious rare and *de novo* alleles, experiencing a unique environment and with ontogenetic pathways developing stochastically and uniquely. This caveat applies to any phenotype, from eyes to aggression. Adaptive explanations are necessarily non-individualised, applying to population-shared traits over generations (Hunt, 2023). Finally, it is important to highlight that although society and medicine may be informed by science, evolutionary explanations have no necessary ethical connotations regarding treatment and accommodations.

8/ Further Applications

The DCIDE method is a formal framework for formulating and assessing adaptive hypotheses, particularly suitable for psychiatric traits. Every account that doesn't start by recognising psychiatric conditions' heterogenous causes is instantly dismissible. Once non-adaptive cases are considered explained, again any account that doesn't recognise the relationship between diagnosable traits and their more evolutionarily-visible correlates will fail to capture the critical phenomena. Then, given various hypotheses attempting to explain those phenomena, without using a systematic method of analysing the available evidence and comparing hypotheses with standardised inferences, progress towards the true explanation is difficult. Although autism serves as a useful exemplification, the DCIDE method could be widely applied. Indeed, beyond its use in whittling down the many existing evolutionary hypotheses for psychopathological traits (see del Giudice (2018) for a comprehensive review), its role in strengthening existing work in evolutionary psychology may also be valuable.

One obvious implication is recognising the value of circumstantial evidence and visibility variables, expanding the scope of useful experiments and data collection beyond trying to precisely evoke psychological function, and encouraging collecting simpler behavioural or demographic data to support hypotheses. This is yet another reason to encourage research beyond traditional convenience samples of students (Henrich et al., 2010), especially cross-culturally and across age groups. Although such evidence is sometimes collected and referenced, this has occurred without a framework of standardised analytical principles, and might have been accused of cherry-picking. We should expect such evidence to align with existing leading hypotheses from evolutionary psychology. To briefly DCIDE review a classic target of evolutionary psychology, jealousy may be Described as thoughts or feelings of insecurity, fear, and concern over a romantic relationship. Presumably few cases will Categorise as non-adaptive. Jealousy is certainly visible to selection, so Integration of associated traits is not required to encompass a likely target phenotype of selection. A leading Depiction (Buss & Haselton, 2005) is that jealousy is activated by threats to a valuable relationship, functioning to protect it from partial or total loss, with many specific features of jealous emotions and behaviours noted as direct evidence of the functional trait. In Evaluation, this hypothesis amply explains the age of effect (from sexual maturity), duration (relating to perceived threat), sex differences (more concerned with sexual infidelity in men and emotional infidelity in women), environmental effects (relating to threat likelihood) and disposition prevalence (the capacity for jealousy seems essentially universal, showing fairly low (29%) heritability (Kupfer et al., 2022), which is plausibly explained as reactive heritability).

This could be applied across evolutionary psychology with useful effects in strengthening accounts. Take another example, of shame Depicted as a defence against being devalued by others (Sznycer et al., 2016). Current research (e.g. Leroux, Hétu and Sznycer 2023) concentrates on direct evidence of the shame system functioning by deterring from shameful actions, attempting to prevent shameful information spreading, and minimizing devaluating effects when they occur. Conducting a DCIDE method review will presumably make little difference in Description, Categorisation, Integration or Depiction, but in Evaluation forces the hypothesis to account for the visibility variables. Note that, no matter the direct evidence offered, if the jealousy or shame Depictions can’t account for the circumstantial evidence of observed age of effect, duration, sex differences, environmental effects and prevalence (and, if biomedical evidence is considered, that too) hypotheses will be Evaluated as insufficient. Importantly, this protects against accusations of cherry-picking and just-so storytelling, because there is a standard evidence base with standardised evolutionary implications to be addressed. Sznycer and colleagues mention a competing explanation of shame as pathological, a psychological maladjustment that leads to a crippling of adaptive self-functions (Tangney et al., 1992). They use direct evidence of

shame's specific features in fulfilling its proposed function to contradict this – a strategy which often (however unfairly) attracts criticism of just-so storytelling, because the hypothesised function of shame follows from unique *a prioi* considerations of 'what evolution would select for' (e.g. 'shame should deter the individual from making choices where the prospective costs of devaluation exceed the benefits') rather than standardised principles regarding pre-determined areas of evidence. Both the cherry-picking and just-so storytelling critiques could be avoided if the same hypothesis was also supported by a DCIDE method review. In Evaluation, the devaluation-defence hypothesis must explain the specific visibility variables, with standardised implications, and if it better explains those variables to the pathological explanation (as is likely), it should be preferred.

As for psychiatric conditions, competing pathological hypotheses must explain the same evidence as evolutionary hypotheses. In fact, when forced to try and explain not only the direct evidence of the trait's suggested function, but also the circumstantial evidence implying visibility to evolutionary selection, as well as the lack of biomedical evidence of Categorisable non-adaptive causes, simple pathological hypotheses denying any role for adaptive explanations may often begin to sound like the much more speculative option – perhaps such speculations should be labelled 'just-no' storytelling.

Bibliography

Abed, R., & St John-Smith, P. (Eds.). (2022). *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge University Press. https://doi.org/DOI: 10.1017/9781009030564

Adrion, J. R., Cole, C. B., Dukler, N., Galloway, J. G., Gladstein, A. L., Gower, G., Kyriazis, C. C., Ragsdale, A. P., Tsambos, G., Baumdicker, F., Carlson, J., Cartwright, R. A., Durvasula, A., Gronau, I., Kim, B. Y., McKenzie, P., Messer, P. W., Noskova, E., Vecchyo, D. O. Del, … Kern, A. D. (2020). A community-maintained standard library of population genetic models. *ELife*, *9*, 1–39. https://doi.org/10.7554/ELIFE.54967

Alawieh, A., Zaraket, F. A., Li, J.-L., Mondello, S., Nokkari, A., Razafsha, M., Fadlallah, B., Boustany, R.-M., & Kobeissy, F. H. (2012). Systems biology, bioinformatics, and biomarkers in neuropsychiatry. *Frontiers in Neuroscience*, *6*, 187. https://doi.org/10.3389/fnins.2012.00187

Amaral, D. G., Li, D., Libero, L., Solomon, M., Van de Water, J., Mastergeorge, A., Naigles, L., Rogers, S., & Wu Nordahl, C. (2017). In pursuit of neurophenotypes: The consequences of having autism and a big brain. *Autism Research*, *10*(5), 711–722. https://doi.org/10.1002/AUR.1755

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.

Asperger, H. (1991). 'Autistic psychopathy' in childhood. In U. Frith (Ed.), *Autism and Asperger syndrome* (pp. 37–92). Cambridge University Press. https://doi.org/10.1017/CBO9780511526770.002

Austad, S. N., & Hoffman, J. M. (2018). Is antagonistic pleiotropy ubiquitous in aging biology? *Evolution, Medicine, and Public Health*, *2018*(1), 287–294. https://doi.org/10.1093/emph/eoy033

Baron-Cohen, S. (1998). Does autism occur more often in families of physicists, engineers, and mathematicians?

Autism, *2*(3), 296–301.

Baron-Cohen, S. (2002). Is Asperger Syndrome Necessarily Viewed as a Disability? *Focus on Autism and Other Developmental Disabilities*, *17*(3), 186–191. https://doi.org/10.1177/10883576020170030801

Baron-Cohen, S. (2008). Autism, hypersystemizing , and truth. *Quarterly Journal of Experimental Psychology*, *61*(June), 64–75.

Baron-Cohen, S. (2019). *The Concept of Neurodiversity Is Dividing the Autism Community*. Scientific American Blog Network. https://blogs.scientificamerican.com/observations/the-concept-of-neurodiversity-is-dividing-the-autism-community/

Baron-Cohen, S. (2020). *The Pattern Seekers: How Autism Drives Human Invention*. Basic Books.

Baron-Cohen, S., & Hammer, J. (1997). Parents of Children with Asperger Syndrome: What is the Cognitive Phenotype? *Journal of Cognitive Neuroscience*, *9*(4), 548–554. https://doi.org/10.1162/JOCN.1997.9.4.548

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, *29*(5), 407–418. https://doi.org/10.1023/A:1023035012436

Baron-Cohen, S., & Wheelwright, S. (1999). 'Obsessions'' in children with autism or Asperger syndrome.' *British Journal of Psychiatry*, *175*(5), 484–490. https://doi.org/10.1192/bjp.175.5.484

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent ? A test with children with Asperger Syndrome. *Learning*, *5*(1), 47–78.

Baron-Cohen, S., Wheelwright, S., Stone, V., & Rutherford, M. (1999). A mathematician, a physicist and a computer scientist with Asperger syndrome: Performance on folk psychology and folk physics tests. *Neurocase*, *5*(6), 475–483. https://doi.org/10.1080/13554799908402743

Baron-Cohen, S., Wheelwright, S., Stott, C., Bolton, P., & Goodyer, I. (1997). Is there a link between engineering and autism? *Autism*, *1*(1), 101–109. https://doi.org/10.1177/1362361397011010

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* MIT Press/Bradford Books.

Baselmans, B. M. L., Yengo, L., van Rheenen, W., & Wray, N. R. (2021). Risk in Relatives, Heritability, SNP-Based Heritability, and Genetic Correlations in Psychiatric Disorders: A Review. *Biological Psychiatry*, *89*(1), 11–19. https://doi.org/10.1016/J.BIOPSYCH.2020.05.034

Beauchaine, T. P., Neuhaus, E., Zalewski, M., Crowell, S. E., & Potapova, N. (2011). The effects of allostatic load on neural systems subserving motivation, mood regulation, and social affiliation. *Development and Psychopathology*, *23*(4), 975–999. https://doi.org/10.1017/S0954579411000459

Belmont, J. W., Boudreau, A., Leal, S. M., Hardenbol, P., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., … Stewart, J. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299–1320. https://doi.org/10.1038/nature04226

Belsky, J. (2012). The Development of Human Reproductive Strategies. *Current Directions in Psychological*

Science, *21*(5), 310–316. https://doi.org/10.1177/0963721412453588

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience*, *16*(9), 551–563. https://doi.org/10.1038/nrn3992

Brosnan, M., Chapman, E., & Ashwin, C. (2014). Adolescents with Autism Spectrum Disorder Show a Circumspect Reasoning Bias Rather than 'Jumping-to-Conclusions.' *Journal of Autism and Developmental Disorders*, *44*(3), 513–520. https://doi.org/10.1007/s10803-013-1897-5

Brown, J. S., & Athena Aktipis, C. (2015). Inclusive fitness effects can select for cancer suppression into old age. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1673). https://doi.org/10.1098/RSTB.2015.0160

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics 2015 47:11*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bury, S. M., Jellett, R., Spoor, J. R., & Hedley, D. (2023). "It Defines Who I Am" or "It's Something I Have": What Language Do [Autistic] Australian Adults [on the Autism Spectrum] Prefer? *Journal of Autism and Developmental Disorders*, *53*(2), 677–687. https://doi.org/10.1007/S10803-020-04425-3

Buss, D. M., & Haselton, M. (2005). The evolution of jealousy. *Trends in Cognitive Sciences*, *9*(11), 506–507. https://doi.org/10.1016/j.tics.2005.09.006

Buunk, B. P., Angleitner, A., Oubaid, V., & Buss, D. M. (1996). Sex Differences in Jealousy in Evolutionary and Cultural Perspective: Tests From the Netherlands, Germany, and the United States. *Psychological Science*, *7*(6), 359–363. https://doi.org/10.1111/j.1467-9280.1996.tb00389.x

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*(3), 314–325. https://doi.org/10.1006/NIMG.2000.0612

Chaudhary, N., & Salali, G. D. (2022). Hunter-Gatherers, Mismatch and Mental Disorder. In P. St John-Smith & R. Abed (Eds.), *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health* (pp. 64–83). Cambridge University Press. https://doi.org/DOI: 10.1017/9781009030564.007

Christensen, J., Grnøborg, T. K., Srøensen, M. J., Schendel, D., Parner, E. T., Pedersen, L. H., & Vestergaard, M. (2013). Prenatal Valproate Exposure and Risk of Autism Spectrum Disorders and Childhood Autism. *JAMA*, *309*(16), 1696. https://doi.org/10.1001/JAMA.2013.2270

Clarke, T. K., Lupton, M. K., Fernandez-Pujals, A. M., Starr, J., Davies, G., Cox, S., Pattie, A., Liewald, D. C., Hall, L. S., Macintyre, D. J., Smith, B. H., Hocking, L. J., Padmanabhan, S., Thomson, P. A., Hayward, C., Hansell, N. K., Montgomery, G. W., Medland, S. E., Martin, N. G., … Mcintosh, A. M. (2016). Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular Psychiatry*, *21*(3), 419–425. https://doi.org/10.1038/MP.2015.12

Cosmides, L., & Tooby, J. (1994). Beyond intuition and instinct blindness: toward an evolutionarily rigorous cognitive science. *Cognition*, *50*(1–3), 41–77. https://doi.org/10.1016/0010-0277(94)90020-5

Cosmides, L., & Tooby, J. (2013). Evolutionary Psychology: New Perspectives on Cognition and Motivation. *Https://Doi.Org/10.1146/Annurev.Psych.121208.131628*, *64*, 201–229. https://doi.org/10.1146/ANNUREV.PSYCH.121208.131628

Crespi, B. (2016). Autism as a disorder of high intelligence. *Frontiers in Neuroscience*, *10*(JUN), 300. https://doi.org/10.3389/fnins.2016.00300

Croen, L. A., Najjar, D. V., Fireman, B., & Grether, J. K. (2007). Maternal and Paternal Age and Risk of Autism Spectrum Disorders. *Archives of Pediatrics & Adolescent Medicine*, *161*(4), 334.

https://doi.org/10.1001/archpedi.161.4.334

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Darwin-Online.Org.Uk. http://darwin-online.org.uk/Variorum/1866/1866-576-c-1859.html

Davison, R. J., & Gurven, M. D. (2021). Human uniqueness? Life history diversity among small-scale societies and chimpanzees. *PLOS ONE*, *16*(2), e0239170. https://doi.org/10.1371/JOURNAL.PONE.0239170

De Martino, B., Harrison, N. A., Knafo, S., Bird, G., & Dolan, R. J. (2008). Explaining enhanced logical consistency during decision making in autism. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(42), 10746–10750. https://doi.org/10.1523/JNEUROSCI.2895-08.2008

Del Giudice, M. (2018a). *Evolutionary Psychopathology: A Unified Approach*. Oxford University Press. https://doi.org/10.1093/med-psych/9780190246846.001.0001

Del Giudice, M. (2018b). The Autism Spectrum. In *Evolutionary Psychopathology* (Vol. 4, Issue 3, pp. 243–260). https://doi.org/10.1093/med-psych/9780190246846.003.0010

Del Giudice, M., Angeleri, R., Brizio, A., & Elena, M. R. (2010). The evolution of autistic-like and schizotypal traits: A sexual selection hypothesis. *Frontiers in Psychology*, *1*(AUG), 1–18. https://doi.org/10.3389/fpsyg.2010.00041

DeLong, G. R., & Dwyer, J. T. (1988). Correlation of family history with specific autistic subgroups: Asperger's syndrome and bipolar affective disease. *Journal of Autism and Developmental Disorders*, *18*(4), 593–600. http://www.ncbi.nlm.nih.gov/pubmed/3215885

Dupré, J. (2003). *Human Nature and the Limits of Science*. Clarendon Press.

Durisko, Z., Mulsant, B. H., McKenzie, K., & Andrews, P. W. (2016). Using Evolutionary Theory to Guide Mental Health Research. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *61*(3), 159. https://doi.org/10.1177/0706743716632517

Fletcher, A., & O'Brien, N. (2008). Disability Rights Commission: From Civil Rights to Social Rights. *Journal of Law and Society*, *35*(4), 520–550. https://doi.org/10.1111/J.1467-6478.2008.00449.X

Fombonne, E., Macfarlane, · Heather, & Salem, A. C. (2021). Epidemiological surveys of ASD: advances and remaining challenges. *Journal of Autism and Developmental Disorders*, *51*. https://doi.org/10.1007/s10803-021-05005-9

Frankenhuis, W. E., & Nettle, D. (2020). Current debates in human life history research. *Evolution and Human Behavior*, *41*(6), 469–473. https://doi.org/10.1016/J.EVOLHUMBEHAV.2020.09.005

Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind." *Cognition*, *50*(1–3), 115–132.

Garcia, A. R., Finch, C., Gatz, M., Kraft, T., Rodriguez, D. E., Cummings, D., Charifson, M., Buetow, K., Beheim, B. A., Allayee, H., Thomas, G. S., Stieglitz, J., Gurven, M. D., Kaplan, H., & Trumble, B. C. (2021). Apoe4 is associated with elevated blood lipids and lower levels of innate immune biomarkers in a tropical amerindian subsistence population. *ELife*, *10*. https://doi.org/10.7554/ELIFE.68231

Gardener, H., Spiegelman, D., & Buka, S. L. (2011). Perinatal and Neonatal Risk Factors for Autism: A Comprehensive Meta-analysis. *Pediatrics*, *128*(2), 344. https://doi.org/10.1542/PEDS.2010-1036

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., Ripke, S., Sandin, S., Sklar, P., Svantesson, O., Reichenberg, A., Hultman, C. M., Devlin, B., Roeder, K., & Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

Geschwind, D. H., & Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science (New York, N.Y.)*, *349*(6255), 1489–1494. https://doi.org/10.1126/science.aaa8954

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. *Current Opinion in Neurobiology*, *17*(1), 103–111. https://doi.org/10.1016/j.conb.2007.01.009

Geschwind, D. H., & State, M. W. (2015). Gene hunting in autism spectrum disorder: On the path to precision medicine. In *The Lancet Neurology* (Vol. 14, Issue 11, pp. 1109–1120). Lancet Publishing Group. https://doi.org/10.1016/S1474-4422(15)00044-7

Gizzonio, V., Avanzini, P., Fabbri-Destro, M., Campi, C., & Rizzolatti, G. (2014). Cognitive abilities in siblings of children with autism spectrum disorders. *Experimental Brain Research*, *232*(7), 2381–2390. https://doi.org/10.1007/S00221-014-3935-8

Glinton, K. E., & Elsea, S. H. (2019). Untargeted Metabolomics for Autism Spectrum Disorders: Current Status and Future Directions. In *Frontiers in Psychiatry* (Vol. 10). Frontiers Media S.A. https://doi.org/10.3389/fpsyt.2019.00647

Gluckman, P. D., Hanson, M. A., & Low, F. M. (2019). Evolutionary and developmental mismatches are consequences of adaptive developmental plasticity in humans and have implications for later disease risk. *Philosophical Transactions of the Royal Society B*, *374*(1770). https://doi.org/10.1098/RSTB.2018.0109

Gluckman, P. D., Hanson, M. A., & Spencer, H. G. (2005). Predictive adaptive responses and human evolution. *Trends in Ecology and Evolution*, *20*(10), 527–533. https://doi.org/10.1016/j.tree.2005.08.001

Gollwitzer, A., Martel, C., McPartland, J. C., & Bargh, J. A. (2019). Autism spectrum traits predict higher social psychological skill. *Proceedings of the National Academy of Sciences*, *116*(39), 19245–19247. https://doi.org/10.1073/PNAS.1911460116

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London - Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Gratten, J., Wray, N. R., Peyrot, W. J., McGrath, J. J., Visscher, P. M., & Goddard, M. E. (2016). Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. *Nature Genetics*, *48*(7), 718–724. https://doi.org/10.1038/ng.3577

Greenberg, D. M., Warrier, V., Allison, C., & Baron-Cohen, S. (2018). Testing the Empathizing–Systemizing theory of sex differences and the Extreme Male Brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences*, *115*(48), 12152–12157. https://doi.org/10.1073/pnas.1811032115

Greenberg, G. (2013, May 16). *The Rats of N.I.M.H. | The New Yorker*. The New Yorker. https://www.newyorker.com/tech/annals-of-technology/the-rats-of-n-i-m-h

Guloksuz, S., & Van Os, J. (2018). The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. In *Psychological Medicine* (Vol. 48, Issue 2, pp. 229–244). Cambridge University Press. https://doi.org/10.1017/S0033291717001775

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Cullen, B., Malik, R., Worrall, B. B., Sudlow, C. L. M., Wardlaw, J. M., Gallacher, J., Pell, J., McIntosh, A. M., Smith, D. J., Gale, C. R., & Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular Psychiatry 2016 21:11*, *21*(11), 1624–1632. https://doi.org/10.1038/mp.2015.225

Haig, D. (2010). Transfers and transitions: parent-offspring conflict, genomic imprinting, and the evolution of human life history. *Proceedings of the National Academy of Sciences of the United States of America*, *107 Suppl*, 1731–1735. https://doi.org/10.1073/pnas.0904111106

Hamilton, M. J., Milne, B. T., Walker, R. S., Burger, O., & Brown, J. H. (2007). The complex structure of hunter–gatherer social networks. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1622), 2195–2203. https://doi.org/10.1098/rspb.2007.0564

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*(10), 1218–1220. https://doi.org/10.1038/nn1770

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61–83; discussion 83-135. https://doi.org/10.1017/S0140525X0999152X

Hill, K., & Hurtado, A. M. (1991). The evolution of premature reproductive senescence and menopause in human females - An evaluation of the "grandmother hypothesis." *Human Nature*, *2*(4), 313–350. https://doi.org/10.1007/BF02692196/METRICS

Hill, W. D., Arslan, R. C., Xia, C., Luciano, M., Amador, C., Navarro, P., Hayward, C., Nagy, R., Porteous, D. J., McIntosh, A. M., Deary, I. J., Haley, C. S., & Penke, L. (2018). Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *Molecular Psychiatry*, *23*(12), 2347–2362. https://doi.org/10.1038/s41380-017-0005-1

Hill, W. D., Davies, G., Liewald, D. C., McIntosh, A. M., & Deary, I. J. (2016). Age-Dependent Pleiotropy Between General Cognitive Function and Major Psychiatric Disorders. *Biological Psychiatry*, *80*(4), 266–273. https://doi.org/10.1016/J.BIOPSYCH.2015.08.033

Hull, L., Petrides, K. V., & Mandy, W. (2020). The Female Autism Phenotype and Camouflaging: a Narrative Review. *Review Journal of Autism and Developmental Disorders*, *7*(4), 306–317. https://doi.org/10.1007/s40489-020-00197-9

Hultman, C. M., Sandin, S., Levine, S. Z., Lichtenstein, P., & Reichenberg, A. (2011). Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. *Molecular Psychiatry*, *16*(12), 1203–1212. https://doi.org/10.1038/mp.2010.121

Hunt, A. D. (2023). Making Wakefield Workable: The Fitness and Function Framework for Taxonomising Evolutionary Dysfunction. *PsyArxiv*. https://doi.org/10.31234/OSF.IO/QBGKE

Hunt, A. D., Abed, R., & St John-Smith, P. (2022). The Biopsychosocial Model Advanced by Evolutionary Theory. In R. Abed & P. St John-Smith (Eds.), *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge University Press.

Hunt, A. D., & Jaeggi, A. V. (2022). Specialised minds: extending adaptive explanations of personality to the evolution of psychopathology. *Evolutionary Human Sciences*, *4*, e26. https://doi.org/10.1017/EHS.2022.23

Hunt, A. D., St-John Smith, P., & Abed, R. (2023). Evobiopsychosocial medicine. *Evolution, Medicine, and Public Health*, *11*(1), 67–77. https://doi.org/10.1093/EMPH/EOAC041

Huxley, J., Mayr, E., Osmond, H., & Hoffer, A. (1964). Schizophrenia as a Genetic Morphism. In *Nature* (Vol. 204, pp. 220–221). http://www.ncbi.nlm.nih.gov/pubmed/14212411

Idring, S., Lundberg, M., Sturm, H., Dalman, C., Gumpert, C., Rai, D., Lee, B. K., & Magnusson, C. (2015). Changes in Prevalence of Autism Spectrum Disorders in 2001–2011: Findings from the Stockholm Youth Cohort. *Journal of Autism and Developmental Disorders*, *45*(6), 1766–1773. https://doi.org/10.1007/s10803-014-2336-y

Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y., Yamrom, B., & Wigler, M. (2015). Low load for disruptive mutations in autism genes and their biased transmission. *Proceedings of the National Academy of Sciences*, *112*(41), E5600–E5607. https://doi.org/10.1073/pnas.1516376112

Ip, J. P. K., Mellios, N., & Sur, M. (2018). Rett syndrome: insights into genetic, molecular and circuit mechanisms. *Nature Reviews Neuroscience 2018 19:6*, *19*(6), 368–382. https://doi.org/10.1038/s41583-018-0006-3

Järvinen, A., Korenberg, J. R., & Bellugi, U. (2013). The social phenotype of Williams syndrome. *Current Opinion in Neurobiology*, *23*(3), 414–422. https://doi.org/10.1016/J.CONB.2012.12.006

John, S., & Jaeggi, A. V. (2021). Oxytocin levels tend to be lower in autistic children: A meta-analysis of 31

studies. *Autism : The International Journal of Research and Practice*, *25*(8), 2152–2161. https://doi.org/10.1177/13623613211034375

Johnson, W., & Bouchard Jr., T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. In *Intelligence* (Vol. 33, Issue 4, pp. 393–416). Elsevier Science. https://doi.org/10.1016/j.intell.2004.12.002

Kanner, L. (1968). Autistic disturbances of affective contact. *Acta Paedopsychiatrica*, *35*(4), 100–136. https://pubmed.ncbi.nlm.nih.gov/4880460/

Kanner, Leo. (1971). Follow-up study of eleven autistic children originally reported in 1943. *Journal of Autism and Childhood Schizophrenia*, *1*(2), 119–145. https://doi.org/10.1007/BF01537953

Kapp, S. K., Pripas-Kapit, S., Dekker, M., Murray, D., Tisoncik, L. A., Baggs, M., Seidel, K., Buckle, K. L., Evans, M., Raymaker, D. M., DaVanport, S., Greenburg, C., Des Rochas Rosa, S., Ne'eman, A., Neumeier, S. M., Brown, L. X. Z., Arnold, L., Elder Robison, J., Onaiwu, M. G., … Russell, G. (2020). *Autistic Community and the Neurodiversity Movement - Stories from the Frontline* (S. K. Kapp (Ed.)). Palgrave MacMillan. https://link.springer.com/content/pdf/10.1007%2F978-981-13-8437-0.pdf

Keller, M. C. (2018). Evolutionary Perspectives on Genetic and Environmental Risk Factors for Psychiatric Disorders. *Annual Review of Clinical Psychology*, *14*(1), 471–493. https://doi.org/10.1146/annurev-clinpsy-050817-084854

Keller, M. C., & Miller, G. (2006). Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behavioral and Brain Sciences*, *29*(4), 385–404. https://doi.org/10.1017/S0140525X06009095

Kelly, R. L. (2013). The lifeways of hunter-gatherers: The foraging spectrum. In *The Lifeways of Hunter-Gatherers: The Foraging Spectrum*. https://doi.org/10.1017/CBO9781139176132

Klein, D. F. (1999). Harmful dysfunction, disorder, disease, illness, and evolution. *Journal of Abnormal Psychology*, *108*(3), 421–429. https://doi.org/10.1037/0021-843X.108.3.421

Knapton, S. (2019). *Autism now so over-diagnosed it will be a meaningless condition within 10 years, warn experts*. The Daily Telegraph. https://www.telegraph.co.uk/science/2019/08/21/autism-now-over-diagnosed-will-meaningless-condition-within/

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kramer, K. L., Schacht, R., & Bell, A. (2017). Adult sex ratios and partner scarcity among hunter–gatherers: implications for dispersal patterns and the evolution of human sociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1729). https://doi.org/10.1098/RSTB.2016.0316

Kupfer, T. R., Sidari, M. J., Zietsch, B. P., Jern, P., Tybur, J. M., & Wesseldijk, L. W. (2022). Why are some people more jealous than others? Genetic and environmental factors. *Evolution and Human Behavior*, *43*(1), 26–33. https://doi.org/10.1016/J.EVOLHUMBEHAV.2021.08.002

Lai, M.-C., Lombardo, M. V., Chakrabarti, B., & Baron-Cohen, S. (2013). Subgrouping the Autism "Spectrum": Reflections on DSM-5. *PLoS Biology*, *11*(4), e1001544. https://doi.org/10.1371/journal.pbio.1001544

Lamport, D., & Turner, L. A. (2014). Romantic Attachment, Empathy, and the Broader Autism Phenotype among College Students. *Http://Dx.Doi.Org/10.1080/00221325.2013.856838*, *175*(3), 202–213. https://doi.org/10.1080/00221325.2013.856838

Lande, R. (1980). Sexual Dimorphism, Sexual Selection, and Adaptation in Polygenic Characters. *Evolution*, *34*(2),

292. https://doi.org/10.2307/2407393

Lea, A. J., Clark, A. G., Dahl, A. W., Devinsky, O., Garcia, A. R., Golden, C. D., Kamau, J., Kraft, T. S., Lim, Y. A. L., Martins, D. J., Mogoi, D., Pajukanta, P., Perry, G. H., Pontzer, H., Trumble, B. C., Urlacher, S. S., Venkataraman, V. V., Wallace, I. J., Gurven, M., … Ayroles, J. F. (2023). Applying an evolutionary mismatch framework to understand disease susceptibility. *PLOS Biology*, *21*(9), e3002311. https://doi.org/10.1371/JOURNAL.PBIO.3002311

Leroux, A., Hétu, S., & Sznycer, D. (2023). The Shame System Operates With High Precision. *Https://Doi.Org/10.1177/14747049231203394*, *21*(3). https://doi.org/10.1177/14747049231203394

Lewis, D. M. G., Al-Shawaf, L., Conroy-Beam, D., Asao, K., & Buss, D. M. (2017). Evolutionary psychology: A how-to guide. *American Psychologist*, *72*(4), 353–373. https://doi.org/10.1037/A0040409

Li, Q., Li, Y., Liu, B., Chen, Q., Xing, X., Xu, G., & Yang, W. (2022). Prevalence of Autism Spectrum Disorder Among Children and Adolescents in the United States From 2019 to 2020. *JAMA Pediatrics*, *176*(9), 943–945. https://doi.org/10.1001/JAMAPEDIATRICS.2022.1846

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1), 5. https://doi.org/10.1038/s41572-019-0138-4

Lord, C., Charman, T., Havdahl, A., Carbone, P., Anagnostou, E., Boyd, B., Carr, T., Vries, P. J. de, Dissanayake, C., Divan, G., Freitag, C. M., Gotelli, M. M., Kasari, C., Knapp, M., Mundy, P., Plank, A., Scahill, L., Servili, C., Shattuck, P., … McCauley, J. B. (2021). The Lancet Commission on the future of care and clinical research in autism. *The Lancet*, *0*(0). https://doi.org/10.1016/S0140-6736(21)01541-5

Lukaszewski, A. W., & Roney, J. R. (2011). The origins of extraversion: Joint effects of facultative calibration and genetic polymorphism. *Personality and Social Psychology Bulletin*, *37*(3), 409–421. https://doi.org/10.1177/0146167210397209

Lyall, K., Croen, L., Daniels, J., Fallin, M. D., Ladd-Acosta, C., Lee, B., Park, B., Snyder, N., Schendel, D., Volk, H., Windham, G., & Newschaffer, C. J. (2017). The Changing Epidemiology of Autism Spectrum Disorders. *Ssrn*. https://doi.org/10.1146/annurev-publhealth-031816-044318

Lyst, M. J., & Bird, A. (2015). Rett syndrome: A complex disorder with simple roots. *Nature Reviews Genetics*, *16*(5), 261–274. https://doi.org/10.1038/nrg3897

Majerus, M. E. N. (2008). Industrial Melanism in the Peppered Moth, Biston betularia: An Excellent Teaching Example of Darwinian Evolution in Action. *Evolution: Education and Outreach*, *2*(1), 63–74. https://doi.org/10.1007/S12052-008-0107-Y/FIGURES/5

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753. https://doi.org/10.1038/nature08494

Markram, K., & Markram, H. (2010). The Intense World Theory – A Unifying Theory of the Neurobiology of Autism . In *Frontiers in Human Neuroscience* (Vol. 4). https://www.frontiersin.org/articles/10.3389/fnhum.2010.00224

Martin, J. S., Jaeggi, A. V., & Koski, S. E. (2023). The social evolution of individual differences: Future directions for a comparative science of personality in social behavior. *Neuroscience & Biobehavioral Reviews*, *144*, 104980. https://doi.org/10.1016/J.NEUBIOREV.2022.104980

Mayes, S. D., & Calhoun, S. L. (2011). Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders*, *5*(2), 749–757. https://doi.org/10.1016/J.RASD.2010.09.002

McCrae, R. R., & Costa Jr., P. T. (2008). The five-factor theory of personality. In *Handbook of personality: Theory*

and research, 3rd ed. (pp. 159–181). The Guilford Press.

Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, *18*(03), 523. https://doi.org/10.1017/S0140525X00039595

Meilleur, A. A. S., Jelenic, P., & Mottron, L. (2015). Prevalence of Clinically and Empirically Defined Talents and Strengths in Autism. *Journal of Autism and Developmental Disorders*, *45*(5), 1354–1367. https://doi.org/10.1007/s10803-014-2296-2

Mitchell, K. J. (2015). The Genetic Architecture of Neurodevelopmental Disorders. *The Genetics of Neurodevelopmental Disorders*, 1–28. https://doi.org/10.1002/9781118524947.CH1

Modabbernia, A., Velthorst, E., & Reichenberg, A. (2017). Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. *Molecular Autism*, *8*(1), 1–16. https://doi.org/10.1186/s13229-017-0121-4

Morrison, K. E., Chambers, L. K., Faso, D. J., & Sasson, N. J. (2018). The content and function of interests in the broad autism phenotype. *Research in Autism Spectrum Disorders*, *49*(February), 25–33. https://doi.org/10.1016/j.rasd.2018.02.002

Mottron, L. (2021). A radical change in our autism research strategy is needed: Back to prototypes. *Autism Research*, *14*(10), 2213–2220. https://doi.org/10.1002/AUR.2494

Muhle, R. A., Reed, H. E., Stratigos, K. A., & Veenstra-VanderWeele, J. (2018). The Emerging Clinical Neuroscience of Autism Spectrum Disorder: A Review. *JAMA Psychiatry*, *75*(5), 514–523. https://doi.org/10.1001/JAMAPSYCHIATRY.2017.4685

Muth, A., Hönekopp, J., & Falter, C. M. (2014). Visuo-Spatial Performance in Autism: A Meta-analysis. *Journal of Autism and Developmental Disorders*, *44*(12), 3245–3263. https://doi.org/10.1007/s10803-014-2188-5

Nakano, T., Kato, N., & Kitazawa, S. (2012). Superior haptic-to-visual shape matching in autism spectrum disorders. *Neuropsychologia*, *50*(5), 696–703. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2011.12.024

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

Nesse, R. M. (2023). Evolutionary psychiatry: foundations, progress and challenges. *World Psychiatry*, *22*(2), 177–202. https://doi.org/10.1002/WPS.21072

Nesse, R. M., & Williams, G. C. (George C. (1996). *Why we get sick : the new science of Darwinian medicine*. 290.

Noland, J. S., Steven Reznick, J., Stone, W. L., Walden, T., & Sheridan, E. H. (2010). Better working memory for non-social targets in infant siblings of children with Autism Spectrum Disorder. *Developmental Science*, *13*(1), 244. https://doi.org/10.1111/J.1467-7687.2009.00882.X

O'Riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(3), 719–730. https://doi.org/10.1037//0096-1523.27.3.719

Okasha, S. (2001). Why Won't the Group Selection Controversy Go Away? *The British Journal for the Philosophy of Science*, *52*(1), 25–50. http://www.jstor.org/stable/3541941

Parker, C. C., & Palmer, A. A. (2011). Dark matter: Are mice the solution to missing heritability? *Frontiers in Genetics*, *2*(JUNE), 10810. https://doi.org/10.3389/FGENE.2011.00032/BIBTEX

Ploog, B. O. (2010). Stimulus overselectivity four decades later: a review of the literature and its implications for current research in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*(11), 1332–1349. https://doi.org/10.1007/S10803-010-0990-2

Polimanti, R., & Gelernter, J. (2017). Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLOS Genetics*, *13*(2), e1006618. https://doi.org/10.1371/journal.pgen.1006618

Prakash, A., & Banerjee, M. (2021). Genomic selection signatures in autism spectrum disorder identifies cognitive genomic tradeoff and its relevance in paradoxical phenotypes of deficits versus potentialities. *Scientific Reports 2021 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-89798-w

Protopapas, A., & Parrila, R. (2019). Dyslexia: Still Not a Neurodevelopmental Disorder. *Brain Sciences*, *9*(1). https://doi.org/10.3390/brainsci9010009

Rao, V., Rosenberg, P., Bertrand, M., Salehinia, S., Spiro, J., Vaishnavi, S., Rastogi, P., Noll, K., Schretlen, D. J., Brandt, J., Cornwell, E., Makley, M., & Miles, Q. S. (2009). Aggression after Traumatic Brain Injury: Prevalence & Correlates. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *21*(4), 420. https://doi.org/10.1176/APPI.NEUROPSYCH.21.4.420

Raychaudhuri, S. (2011). Mapping rare and common causal alleles for complex human diseases. *Cell*, *147*(1), 57. https://doi.org/10.1016/J.CELL.2011.09.011

Riglin, L., Wootton, R. E., Thapar, A. K., Livingston, L. A., Langley, K., Collishaw, S., Tagg, J., Smith, G. D., Stergiakouli, E., Tilling, K., & Thapar, A. (2021). Variable Emergence of Autism Spectrum Disorder Symptoms From Childhood to Early Adulthood. *The American Journal of Psychiatry*, *178*(8), 752–760. https://doi.org/10.1176/APPI.AJP.2020.20071119/ASSET/IMAGES/LARGE/APPI.AJP.2020.20071119F3.JPEG

Rimland, B, & Fein, D. (1988). Special talents of autistic savants. In *The exceptional brain: neuropsychology of talent and special abilites* (pp. 472–492). Guilford Press.

Rimland, Bernard. (1978). Savant capabilities of autistic children and their cognitive implications. In *Cognitive defects in the development of mental illness.* (pp. 43–65). Brunner/Mazel.

Robinson, E. B., Samocha, K. E., Kosmicki, J. A., McGrath, L., Neale, B. M., Perlis, R. H., & Daly, M. J. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proceedings of the National Academy of Sciences*, *111*(42), 15161–15165. https://doi.org/10.1073/PNAS.1409204111

Rødgaard, E.-M., Jensen, K., Vergnes, J.-N., Soulières, I., & Mottron, L. (2019). Temporal Changes in Effect Sizes of Studies Comparing Individuals With and Without Autism. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2019.1956

Rogers, A. (2017, May 11). *Star Neuroscientist Tom Insel Leaves the Google-Spawned Verily for ... a Startup? | WIRED*. Wired. https://www.wired.com/2017/05/star-neuroscientist-tom-insel-leaves-google-spawned-verily-startup/?mbid=social_twitter_onsiteshare

Ronemus, M., Iossifov, I., Levy, D., & Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. *Nature Reviews Genetics*, *15*(2), 133–141. https://doi.org/10.1038/nrg3585

Russell, G., Stapley, S., Newlove-Delgado, T., Salmon, A., White, R., Warren, F., Pearson, A., & Ford, T. (2021). Time trends in autism diagnosis over 20 years: a UK population-based cohort study. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/JCPP.13505

Sabatos-DeVito, M., Schipul, S. E., Bulluck, J. C., Belger, A., & Baranek, G. T. (2016). Eye Tracking Reveals Impaired Attentional Disengagement Associated with Sensory Response Patterns in Children with Autism. *Journal of Autism and Developmental Disorders*, *46*(4), 1319–1333. https://doi.org/10.1007/S10803-015-2681-5

Sasson, N. J., Lam, K. S. L., Childress, D., Parlier, M., Daniels, J. L., & Piven, J. (2013). The Broad Autism Phenotype Questionnaire: Prevalence and Diagnostic Classification. *Autism Research*, *6*(2), 134–143. https://doi.org/10.1002/aur.1272

Schraiber, J. G., & Akey, J. M. (2015). Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics 2015 16:12*, *16*(12), 727–740. https://doi.org/10.1038/nrg4005

Seery, M. D., Holman, E. A., & Silver, R. C. (2010). Whatever does not kill us: Cumulative lifetime adversity, vulnerability, and resilience. *Journal of Personality and Social Psychology*, *99*(6), 1025–1041. https://doi.org/10.1037/a0021344

Sell, A., Sznycer, D., Al-Shawaf, L., Lim, J., Krauss, A., Feldman, A., Rascanu, R., Sugiyama, L., Cosmides, L., & Tooby, J. (2017). The grammar of anger: Mapping the computational architecture of a recalibrational emotion. *Cognition*, *168*, 110–128. https://doi.org/10.1016/J.COGNITION.2017.06.002

Sharp, T. H., Elsabbagh, M., Pickles, A., & Bedford, R. (2023). The subcortical correlates of autistic traits in school-age children: a population-based neuroimaging study. *Molecular Autism*, *14*(1), 1–12. https://doi.org/10.1186/S13229-023-00538-5/FIGURES/3

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Https://Doi.Org/10.1146/Annurev-Psych-122216-011845*, *69*, 487–510. https://doi.org/10.1146/ANNUREV-PSYCH-122216-011845

Singer, A. (2022). It's time to embrace 'profound autism.' *Spectrum*. https://doi.org/10.53053/HPJN5392

Skotko, B. G., Levine, S. P., & Goldstein, R. (2011). Self-perceptions from people with Down syndrome. *American Journal of Medical Genetics Part A*, *155*(10), 2360–2369. https://doi.org/10.1002/AJMG.A.34235

Smith, S. E. (2019). Is Evolutionary Psychology Possible? *Biological Theory*, *1*, 3. https://doi.org/10.1007/s13752-019-00336-4

Smoller, J. W., Andreassen, O. A., Edenberg, H. J., Faraone, S. V., Glatt, S. J., & Kendler, K. S. (2018). Psychiatric genetics and the structure of psychopathology. *Molecular Psychiatry 2018 24:3*, *24*(3), 409–420. https://doi.org/10.1038/s41380-017-0010-4

Solmi, M., Song, M., Yon, D. K., Lee, S. W., Fombonne, E., Kim, M. S., Park, S., Lee, M. H., Hwang, J., Keller, R., Koyanagi, A., Jacob, L., Dragioti, E., Smith, L., Correll, C. U., Fusar-Poli, P., Croatto, G., Carvalho, A. F., Oh, J. W., … Cortese, S. (2022). Incidence, prevalence, and global burden of autism spectrum disorder from 1990 to 2019 across 204 countries. *Molecular Psychiatry 2022 27:10*, *27*(10), 4172–4180. https://doi.org/10.1038/s41380-022-01630-7

South, M., Chamberlain, P. D., Wigham, S., Newton, T., Le Couteur, A., McConachie, H., Gray, L., Freeston, M., Parr, J., Kirwan, C. B., & Rodgers, J. (2014). Enhanced decision making and risk avoidance in high-functioning autism spectrum disorder. *Neuropsychology*, *28*(2), 222–228. https://doi.org/10.1037/neu0000016

Spikins, P., Wright, B., & Hodgson, D. (2016). Are there alternative adaptive strategies to human pro-sociality? The role of collaborative morality in the emergence of personality variation and autistic traits. *Time and Mind*, *9*(4), 289–313. https://doi.org/10.1080/1751696X.2016.1244949

Spitzer, R. L. (1999). Harmful dysfunction and the DSM definition of mental disorder. *Journal of Abnormal Psychology*, *108*(3), 430–432. http://www.ncbi.nlm.nih.gov/pubmed/10466265

Stankowich, T., & Blumstein, D. T. (2005). Fear in animals: a meta-analysis and review of risk assessment. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1581), 2627–2634. https://doi.org/10.1098/RSPB.2005.3251

Stanutz, S., Wapnick, J., & Burack, J. A. (2014). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism*, *18*(2), 137–147. https://doi.org/10.1177/1362361312462905

Stein, D. J., Phillips, K. A., Bolton, D., Fulford, K. W. M., Sadler, J. Z., & Kendler, K. S. (2010). What is a mental/psychiatric disorder? from DSM-IV to DSM-V. *Psychological Medicine*, *40*(11), 1759–1765. https://doi.org/10.1017/S0033291709992261

Stewart-Williams, S., & Thomas, A. G. (2013). The Ape That Thought It Was a Peacock: Does Evolutionary Psychology Exaggerate Human Sex Differences? *Psychological Inquiry*, *24*(3), 137–168. https://doi.org/10.1080/1047840X.2013.804899

Svedholm-Häkkinen, A. M., & Lindeman, M. (2016). Testing the Empathizing-Systemizing theory in the general population: Occupations, vocational interests, grades, hobbies, friendship quality, social intelligence, and sex role identity. *Personality and Individual Differences*, *90*, 365–370. https://doi.org/10.1016/J.PAID.2015.11.044

Syme, K. L., & Hagen, E. H. (2019). Mental health is biological health: Why tackling "diseases of the mind" is an imperative for biological anthropology in the 21st century. *American Journal of Physical Anthropology*, ajpa.23965. https://doi.org/10.1002/ajpa.23965

Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., & Halperin, E. (2016). Shame closely tracks the threat of devaluation by others, even across cultures. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2625–2630. https://doi.org/10.1073/PNAS.1514699113/SUPPL_FILE/PNAS.1514699113.SD01.XLSX

Tangney, J. P., Wagner, P., & Gramzow, R. (1992). Proneness to Shame, Proneness to Guilt, and Psychopathology. *Journal of Abnormal Psychology*, *101*(3), 469–478. https://doi.org/10.1037/0021-843X.101.3.469

Taylor, J. L., Debost, J.-C. P. G., Morton, S. U., Wigdor, E. M., Heyne, H. O., Lal, D., Howrigan, D. P., Bloemendal, A., Larsen, J. T., Kosmicki, J. A., Weiner, D. J., Consortium, P. C. G., Homsy, J., Seidman, J. G., Seidman, C. E., Agerbo, E., McGrath, J. J., Mortensen, P. B., Petersen, L., … Robinson, E. B. (2017). Paternal-age-related de novo mutations and risk for five disorders. *BioRxiv*, 192740. https://doi.org/10.1101/192740

Tooby, J., & Cosmides, L. (1990). On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation. *Journal of Personality*, *58*(1), 17–67. https://doi.org/10.1111/j.1467-6494.1990.tb00907.x

Tsuchiya, K. J., Matsumoto, K., Miyachi, T., Tsujii, M., Nakamura, K., Takagai, S., Kawai, M., Yagi, A., Iwaki, K., Suda, S., Sugihara, G., Iwata, Y., Matsuzaki, H., Sekine, Y., Suzuki, K., Sugiyama, T., Mori, N., & Takei, N. (2008). Paternal age at birth and high-functioning autistic-spectrum disorder in offspring. *British Journal of Psychiatry*, *193*(4), 316–321. https://doi.org/10.1192/bjp.bp.107.045120

Uher, R. (2009). The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Molecular Psychiatry*, *14*(12), 1072–1082. https://doi.org/10.1038/mp.2009.85

Verweij, K. J. H., Yang, J., Lahti, J., Veijola, J., Hintsanen, M., Pulkki-Råback, L., Heinonen, K., Pouta, A., Pesonen, A. K., Widen, E., Taanila, A., Isohanni, M., Miettunen, J., Palotie, A., Penke, L., Service, S. K., Heath, A. C., Montgomery, G. W., Raitakari, O., … Zietsch, B. P. (2012). Maintenance of genetic variation in human personality: Testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. *Evolution*, *66*(10), 3238–3251. https://doi.org/10.1111/j.1558-5646.2012.01679.x

Waizbard-Bartov, E., & Miller, M. (2023). Does the severity of autism symptoms change over time? A review of the evidence, impacts, and gaps in current knowledge. *Clinical Psychology Review*, *99*, 102230. https://doi.org/10.1016/J.CPR.2022.102230

Wakefield, J. C. (1992). The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values. *American Psychologist*, *47*(3), 373–388. https://doi.org/10.1037/0003-066X.47.3.373

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

Walker, R. S., Gurven, M., Burger, O., & Hamilton, M. J. (2007). The trade-off between number and size of

offspring in humans and other primates. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1636), 827–834. https://doi.org/10.1098/RSPB.2007.1511

Warrier, V., Zhang, X., Reed, P., Havdahl, A., Moore, T. M., Cliquet, F., Leblond, C. S., Rolland, T., Rosengren, A., Caceres, A. S. J., Hayward, H., Crawley, D., Faulkner, J., Sabet, J., Ellis, C., Oakley, B., Loth, E., Charman, T., Murphy, D., … Baron-Cohen, S. (2022). Genetic correlates of phenotypic heterogeneity in autism. *Nature Genetics 2022 54:9*, *54*(9), 1293–1304. https://doi.org/10.1038/s41588-022-01072-5

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution.* Oxford University Press.

Williams, G. C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution*, *11*(4), 398. https://doi.org/10.2307/2406060

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Williams, T. M., & Carroll, S. B. (2009). Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature Reviews Genetics*, *10*(11), 797–804. https://doi.org/10.1038/nrg2687

Wilson, D. R. (1998). Evolutionary epidemiology and manic depression. *British Journal of Medical Psychology*, *71*(4), 375–395. https://doi.org/10.1111/j.2044-8341.1998.tb00999.x

Wittman, T. N., Robinson, C. D., McGlothlin, J. W., & Cox, R. M. (2021). Hormonal pleiotropy structures genetic covariance. *Evolution Letters*, *5*(4), 397–407. https://doi.org/10.1002/EVL3.240

Wolf, M., Van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, *447*(7144), 581–584. https://doi.org/10.1038/nature05835

Wu, S., Wu, F., Ding, Y., Hou, J., Bi, J., & Zhang, Z. (2017). Advanced parental age and autism risk in children: a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, *135*(1), 29–41. https://doi.org/10.1111/acps.12666

Zachar, P., & Kendler, K. S. (2017). The Philosophy of Nosology. *Annual Review of Clinical Psychology*, *13*(1), 49–71. https://doi.org/10.1146/annurev-clinpsy-032816-045020

Zapf, A. C., Glindemann, L. A., Vogeley, K., & Falter, C. M. (2015). Sex Differences in Mental Rotation and How They Add to the Understanding of Autism. *PLoS ONE*, *10*(4). https://doi.org/10.1371/JOURNAL.PONE.0124628

Zaroff, C. M., & Uhm, S. Y. (2012). Prevalence of autism spectrum disorders and influence of country of measurement and ethnicity. *Social Psychiatry and Psychiatric Epidemiology*, *47*(3), 395–398. https://doi.org/10.1007/s00127-011-0350-3

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*(5), 778–790. https://doi.org/10.1002/AUR.2696

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, in particular 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 as inaccessible to certain individuals. Biological differences thus 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, p49). 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, p4). 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 posed with 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, and having stopped her stimulant medication, 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 tease apart how naturalistic and normative approaches interact in medical conceptualisation and 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 likely 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 mere value judgements, we have a strong case for believing naturalism is core to our conception of health. 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 when both are *necessary,* naturalism's aim at objectivity is counter-acted by the necessity of accounting for values. The advance provided by Wakefield's hybrid view beyond simple normativity is 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." (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" (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. In cases of functional systems currently labelled disorders and treated as medical problems, the HDA 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 both normativity and naturalism falls short of an infallible hybrid concept reflecting medical norms, both in terms of mainstream medical language use 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 implied as less appropriate when concerned suffering related to evolutionarily functional systems (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 (causal explanations, medical research, patient experience and so on) is considered in integration with all other values (religious, traditional, political, humanist; social, familial, individual and so on).*

The HHD formulation is supposed to more accurately capture medical goals and current norms of thinking about treatment of disorder than the HDA. 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 in medical decision-making.

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 remain stalwart without seeming inhumane). This is built into the HHD, meaning it better accords with public and practitioner actions – inevitably so, because it essentially asks 'given this naturalistic fact, do you 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 a condition's disorder status 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, we are unlikely to update our consideration of its disorder status. Similarly, we are unlikely to significantly revise our consideration of whether the anxiety disorder is a true disorder upon receiving a new form of mechanistic information, such as evidence of hormonal or genetic components. 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. We wouldn't update our definition of anxiety disorder as true disorder upon novel mechanistic information, but would if discovering the anxiety was evolutionarily advantageous. 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 often reconsider a condition's status as 'disorder' upon evolutionary but not neuroscientific information, different scientific explanations likely have different effects on attitudes towards medical treatment.

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)) may have a more significant effect on treatment decisions than specific mechanistic explanations – apart from in cases where the mechanistic explanation implies a specific treatment (e.g. depression is caused by a serotonin imbalance, so drugs which act upon the serotonergic system are the ideal treatment). 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). The HHD would account for the effect of chemical imbalance explanations as a form of scientific explanation that was unusually strong in affecting values regarding medical treatment, likely derived from its intuitive simplicity, despite being very poorly evidenced (Moncrieff *et al.* 2022). Evolutionary explanations don't seem to have any such simple direct medical relevance, as they don't suggest a specific ideal intervention (e.g. 'fixing the imbalance').

Certain explanations (e.g. of an 'evolutionary mismatch' (Lea *et al.* 2023) where disorder arises due to differences between contemporary and ancestral environments) might have more precise connotations for medical response (e.g. reversing elements of the environment causing the mismatch; encouraging Mary to spend more time on the farm), but in general, the effects of evolutionary explanations on treatment decisions will depend on their complex interaction with the holistic value systems, as the HHD recognises. Although these are hard to predict and will differ case-by-case, it's notable that 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), and a recent study has found that evolutionary explanations of depression reduce self-stigmatisation (Schroder, Devendorf and Zikmund-Fisher 2023). How evolutionary explanations affect treatment decisions is an interesting avenue of empirical investigation in the future – but the HHD will undoubtedly be better able to account for the complexities of those effects than the HDA.

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 pathologised and treated as a mental illness, justifying the application of electrocution, nauseating drugs and stigmatisation of homosexuals (Taylor 2011). We now recognise this as immoral and harmful. Such environmental relativism is arguably what drives critics such as Boorse (1977) and 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 as a trait needing remedy. 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. Implausibility aside, 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 pathologised for so long was because other systems (e.g. religion) were informing our values, 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, or proving an evolutionary function of homosexuality, 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 (including evolutionists) 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 – it saves us from the naturalistic fallacy. 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' (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. Strong 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.

Naturalistic considerations may or may not be incorporated into deciding medical values in particular cases. Naturalism forms a *subset* of the competing ideas which inform values. When deciding the correct medical response, we could consider etiology, religion, individual suffering and social acceptance, and eventually decide individual suffering justifies medical treatment, even though the condition is evolutionarily 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. Similarly, people may decide on different treatment regimes when faced with different evolutionary explanations of their condition – if those explanations affect their values.

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 provided 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 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.

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.

Changing Perspectives on Autism: overlapping contributions of evolutionary psychiatry and the neurodiversity movement

Adam D. Hunt[1] and Tanya L. Procyshyn[2]

[1]Institute of Evolutionary Medicine, University of Zurich, Switzerland
[2] Autism Research Centre, Department of Psychiatry, University of Cambridge

Abstract

Perspectives on autism and psychiatric conditions are affected by a mix of scientific and social influences. Evolutionary psychiatry and the neurodiversity movement are emerging paradigms that reflect these two influences, with the former grounded in scientific theory and the latter driven by political and social principles. Despite their separate foundations, there is significant overlap between evolutionary psychiatry and neurodiversity that has not been explored. Specifically, both paradigms reframe disorders as natural cognitive differences rather than disease; expand the concept of 'normal' beyond that implied in modern psychiatry; focus on relative strengths; recognise that modern environments disadvantage certain individuals to cause functional impairment; emphasise cognitive variation being socially accommodated and integrated rather than treated or cured; and can help reduce stigmatisation. However, in other ways, they are distinct and sometimes in conflict. EP emphasises scientific explanation, defines 'dysfunction' in objective terms, and differentiates heterogenous cases based on underlying cause (i.e., autism due to *de novo* genetic mutations). The neurodiversity movement emphasises social action, removes barriers to inclusion, promotes inclusive language, and allows unrestricted identification as neurodivergent. By comparing and contrasting these two approaches, we find that evolutionary psychiatry can, to some extent, support the goals of neurodiversity. In particular, evolutionary perspectives could be convincing to groups more responsive to scientific evidence and help achieve a middle ground between neurodiversity advocates and critics of the movement.

1/ Introduction

In recent years, evolutionary psychiatry and neurodiversity have emerged as novel paradigms for reframing cognitive states and traits historically diagnosed as disorders. At first glance, these two movements occupy distinct spheres: evolutionary psychiatry seeks to understand such conditions by examining their roots in human evolution, exploring theories related to potential adaptive functions and mismatches with the modern environment. The neurodiversity social movement advocates for the acceptance and celebration of cognitive differences, emphasising that conditions including, but not limited to, autism and attention-deficit hyperactivity disorder (ADHD) reflect natural variation in human brains and behaviour rather than pathology. Despite their differences, evolutionary psychiatry and neurodiversity

share the overarching goal of broadening perspectives on cognitive differences in a way that leads to reconsideration of conditions typically diagnosed as psychopathology. In this article, we consider the ability of these emerging paradigms to complement each other, to better achieve understanding and support of differently-thinking people, with a particular focus on autism.

2/ Principles of evolutionary psychiatry

Mainstream psychiatric research generally seeks underlying mechanistic causes of diagnosed conditions with intention to treat. Such causes may be biological (e.g., genetic mutations, alterations in brain structure or function, variation in neurochemistry) or environmental (e.g., stress). In contrast, evolutionary psychiatry emphasises explanation at the 'ultimate' level (Mayr, 1961), asking *why* the process of evolution by natural selection has resulted in these 'proximate' causal mechanisms and their particular effects (Hunt et al., 2023; Nesse, 2023). Why do brains, genes, and stress predispose us to psychiatric conditions?

The paradox that psychiatric conditions are common, heritable, and (as perceived by some) harmful is widely recognised (Keller & Miller, 2006): Shouldn't natural selection have eliminated the responsible genetic variants? Why haven't humans evolved optimal mental health? Evolutionists have proposed diverse explanations for the persistence of various psychiatric symptoms and conditions. For example, ADHD may reflect 'evolutionary mismatch': while increased activity levels and the tendency to rapidly switch attention may have been adaptive in the environments of our evolutionary past, these traits may be experienced as challenges in the more sedentary modern world. Moderate anxiety symptoms could have value as crucial, life-saving signals in the face of certain threats, and may be a misinterpretation of adaptation as disease (Nesse, 2019). One outcome of adopting an evolutionary perspective is reconsidering certain presumed cognitive dysfunctions as differences and recognising their potential functional adaptive value (although caution should be taken to avoid excessive evolutionary storytelling (Gould & Lewontin, 1979)).

3/ Evolutionary theories of autism

Autism spectrum conditions (henceforth autism), which are characterised by a combination of social and behavioural symptoms, present a complex case for evolutionary explanations given their high level of phenotypic and aetiological heterogeneity and associated low reproductive success (Ploeger & Galis, 2011). Nevertheless, proposed evolutionary explanations generally recognise the cognitive strengths shown by many autistic individuals, framing social challenges as costs offset by these abilities.

The clinical psychologist Simon Baron-Cohen has a long history of challenging the pathologising of autism (2002), noting that autistic traits could be restated without negative connotations (e.g., 'obsessive behaviour' restated as 'single-minded focus') and calling for autism spectrum disorder to be renamed autism spectrum *condition* (Lai & Baron-Cohen, 2015). His empathising-systemising theory proposes that autism reflects an extreme position along a dimension of people- and system-focused cognition, resulting in strengths in understanding patterns and rule-based systems paired with challenges in social cognition. Support for this theory includes autistic people's strengths in understanding of folk physics

(Baron-Cohen, Wheelwright, Spong, et al., 2001), high attention to detail (O'Riordan et al., 2001), and expertise in science, technology, engineering, and mathematics (STEM) (Baron-Cohen et al., 1999). Notably, individuals working STEM-related domains score higher than average on self-report questionnaires assessing non-clinical levels of autism-related traits, implying such traits (which are, to some degree, heritable) also predispose one to technical ability (Baron-Cohen *et al.* 2001a; Greenberg et al. 2018). Baron-Cohen's book *The Pattern Seekers* (2020) develops a more fully fledged evolutionary theory, suggesting that individuals exceptionally skilled in systemising were valuable in ancestral social groups, potentially playing roles as tool-makers, inventors, and knowledge experts (e.g., of properties of plants, animal behaviour).

Subsequent evolutionary theories similarly propose that autism is related to cognitive specialisation (Hunt and Jaeggi 2022). Crespi & Badcock (2008) frame autism and schizophrenia as "diametrical disorders of the social brain", with the former reflecting hyperdevelopment of mechanistic cognition paired with underdevelopment of mentalistic cognition (and vice versa for psychotic spectrum conditions such as schizophrenia). They collate evidence of opposite patterns of phenotypic alterations between these conditions (e.g., in autism: enhanced visuo-spatial skills, increased local processing, reduced imagination; vice versa in psychotic spectrum conditions) alongside differences in neuroanatomy, neurological function, genetic variants, and levels of hormones and growth factors that may influence the differential trajectories of brain development that predispose individuals to autism or psychotic spectrum conditions. Subsequently, Crespi (2016) developed the 'high intelligence imbalance' hypothesis more specific to autism, arguing that strong recent evolutionary pressures for increased intelligence in humans cause vulnerability to autism when certain aspects of intelligence are exaggerated. As empirical support, Crespi points to recent findings of overlap between aspects of brain size, brain growth, genetic variants, and sensory and visuo-spatial abilities associated with autism and high IQ. Moreover, autistic people and people with a high IQ both typically show more focused attention, a more deliberate decision-making style, and greater interest in STEM. Notably, several cognitive strengths associated with autism are shared by the (non-autistic) relatives of autistic people (Gizzonio et al., 2014; Noland et al., 2010), further suggesting a shared genetic basis between autism and cognitive specialisation.

Such evolutionary hypotheses align with current genetic understanding of autism, namely that autism likelihood is influenced by more than a thousand genes with individually small effects. In the general population, polygenic scores for autism (i.e., an individual's collective "sum" of genes linked to autism) are positively correlated with cognitive ability (Clarke et al., 2016) and verbal-numerical reasoning and educational attainment (Hagenaars et al., 2016). In other words, individuals who carry more autism-related genes (yet are not autistic) tend to perform better on cognitive tests and be more successful in formal education. Studies examining specific genetic variants report similar patterns: Olduvai protein domain family gene (formally *DUF1220*) copy number is linked to increased brain size, higher cognitive function (measured as higher IQ and mathematical aptitude), and higher autism likelihood (Sikela & Searles Quick, 2018). Taken together, these lines of evidence support the evolutionary hypothesis that genes

with pleiotropic effects related to brain development, intelligence, and autism were selected for as humankind found itself living in increasingly technologically complex societies.

From an evolutionary perspective, then, predisposition to autism could be explained as adaptation for increased human cognitive specialisation, systemising and intelligence, with the most disabling cases of autism occurring as occasional costly by-products that are not adaptive in themselves. For a parallel proposed shift of perspective on autism and other cognitive differences independent of adaptive explanations, we turn to the contributions of the neurodiversity movement.

4/ The neurodiversity movement

Neurodiversity, introduced in the milieu of the social model of disability (Shakespeare, 2006), now generally refers to the social movement with the goal of reconceptualising common, lifelong mental conditions as differences to be accepted and integrated, rather than defects or diseases to be treated or cured (Kapp et al., 2020). Neurodiversity shares similarities with, and in some cases was explicitly inspired by, historical movements in antipsychiatry (Szasz, 1974) and disability activism, such as 'Mad Pride' (Lewis, 2006). While the neurodiversity movement was initiated by autistic people, it has expanded to include numerous conditions broadly classified as mental disorders (Nelson, 2020). Advocates generally seek recognition and acceptance of neurodiversity — diversity among minds — as a valuable part of natural human cognitive diversity (alike biodiversity), claiming that differences are harmfully pathologised and strengths of 'neurodivergent' individuals are overlooked by the 'neurotypical' majority. Attempts to cure psychiatric conditions are seen as equivalent to curing someone's sexuality or race; just as there is no 'right' gender or race, there is no 'normal' or 'right' type of mind. Neurodiversity thus offers a social paradigm for reconceiving psychiatric traits as differences rather than diseases, promoting the need for acceptance and accommodation instead of cures or treatment. The movement also highlights how modern society is, to some extent, responsible for turning differences into disabilities. For example, the expectation of fluid social interactions in schools or workplaces may disable autistic people, and their disability could be prevented by removing this expectation.

To advocates, the argument for neurodiversity is founded on the explicit goal of social justice and biological explanations are only engaged with in the context of rejecting pathologisation and emphasising that disorder and dysfunction are contingent on contemporary environments (Chapman, 2021). Although biological and mainstream psychiatric explanations of neurodiversity are generally avoided or outright rejected by the movement, there is a history of speculation that neurodivergent conditions may have been integrated or better valued in the evolutionary past. For example, autistic advocate Temple Grandin mused: "Who do you think made the first stone spear? That wasn't the yakkity yaks sitting around the campfire. It was some Asperger sitting in the back of a cave..." (Weiss, 2010).

5/ Areas of alignment between evolutionary psychiatry and neurodiversity

Despite their common reference to strengths and emphasis on modern environments as responsible for creating disability, extended investigation into the areas of overlap and divergence between evolutionary explanations and neurodiversity perspectives is lacking.

While it must be acknowledged that the neurodiversity movement and evolutionary psychiatry occupy distinct spheres — with the former driven by political and social principles and the latter by scientific theory — many of their goals and implications are shared (Figure 1). Specifically, they overlap in their: (i) reframing of traits currently diagnosed as mental disorders as natural cognitive differences rather than disease; (ii) expansion of the concept of 'normal' beyond that implied in mainstream psychiatry; (iii) encouragement of understanding of psychiatric conditions in terms of relative strengths; (iv) recognition that modern environments unfairly disadvantage certain individuals, and so the environment is causing functional impairment; (v) emphasis on cognitive variation being socially accommodated and integrated rather than treated or cured; and (vi) potentially reducing stigmatisation.

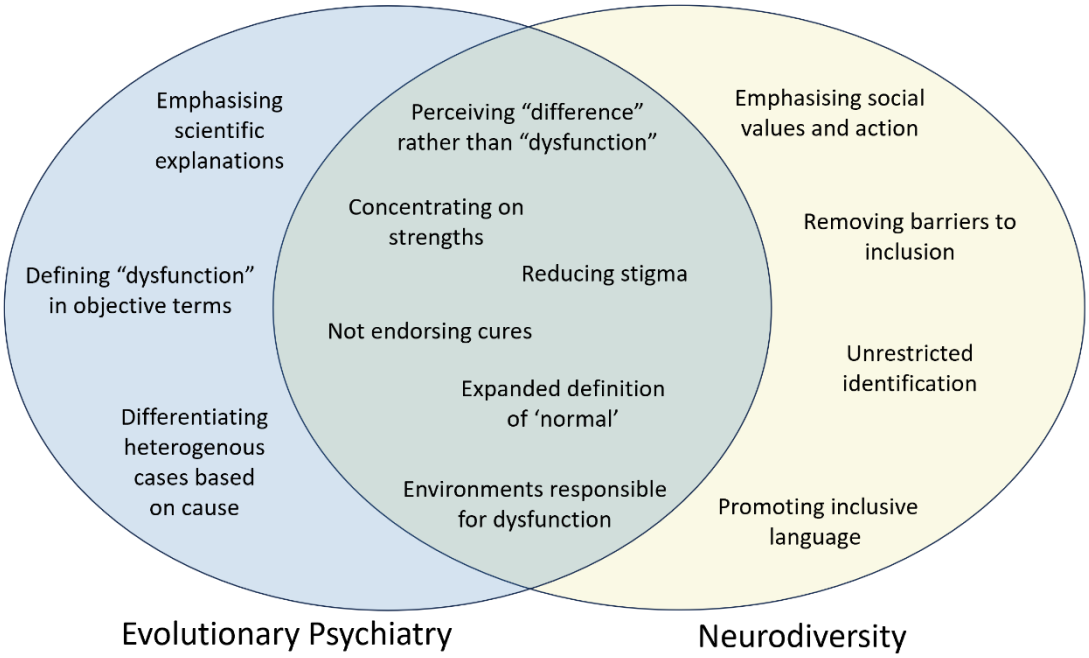

Figure 1: A Venn diagram displaying areas of overlap and difference between evolutionary psychiatry and neurodiversity

Both paradigms are concerned that the pathologising of psychiatric conditions has justified billions of dollars in biomedical research focussed on cures and causes with near zero success in improving outcomes for affected individuals (Insel, 2013). Evolutionary psychiatry directs attention to different areas of research and approaches to inclusion aligned closely with the neurodiversity perspective. Despite the need for separation of scientific fact and social values (to avoid the 'naturalistic fallacy'), evolutionary explanations may play important roles fostering support for neurodiversity. Notably, historic political movements pushing for changes in treatment of minority groups have referenced scientific facts (Prontzos 2019; Brookey 2002). Scientific justification could be useful in convincing demographics unlikely to be swayed by purely social arguments. This includes the scientific research community, who still overwhelmingly research brain and genetic correlates searching for biomarkers, assuming the conditions to be true pathologies in need of cures. Evolutionary accounts instead encourage identifying correlated cognitive benefits of neurodiverse conditions, understanding the specific modern factors that exacerbate harms, and rejecting simple biomedical

treatments if these traits are linked with adaptations. Published findings of strengths associated with various conditions, or genetic evidence that most cases of a particular condition are due to variants common in the general population — rather than "disease alleles" — could support the shared aim of evolutionary psychiatry and the neurodiversity movement in redefining what it means to have a psychiatric diagnosis. Wider recognition of evolutionary perspetives could also increase participation of neurodivergent individuals in scientific research – a concern raised by neurodiversity advocates (Fletcher-Watson et al., 2019) – as study designs go beyond reductionist biomedical models and embrace more holistic approaches to both understand the conditions and improve outcomes.

Social movements endeavouring to alter norms and language relate to what philosophers call 'conceptual engineering': the process by which existing concepts are assessed and improvements are proposed and implemented (Plunkett & Cappelan, 2020). The neurodiversity movement has clearly been engaged in a form of conceptual engineering, introducing the term 'neurodiversity' and attempting to redefine what it means to be labelled as autistic, ADHD, dyslexic, etc. Given that medicalisation of a state carries connotations of negative evaluation – something to be cured or prevented – the need to move away from pathology-oriented language is entirely understandable. While the neurodiversity movement's conceptual engineering efforts are concerned with acceptance and not explanation, evolutionary psychiatry has the power to explain cognitive differences without pathologising. Indeed, there is evidence that explanations of mental health conditions referring to stress, rather than neurological or genetic malfunction, result in lower stigmatisation (Loughman & Haslam, 2018); that evolutionary explanations lead to lower self-stigmatisation in individuals with depression (Schroder et al., 2023); and that cognitive behavioural therapy can be enhanced by integrating evolutionary insights (Abrams, 2020), particularly by destigmatising conditions or encouraging individuals to perceive potential strengths associated with their conditions. As the success of conceptual engineering is related to how well the new concept is adopted (Pinder, 2017), if the communities that neurodiversity advocates communicate with are better convinced by scientific – and specifically evolutionary – arguments, then adopting those hypotheses will help their goals.

In the context of autism, promotion of 'strength-based' approaches are associated with improved life satisfaction (Duan et al., 2014); improved parental perceptions of their autistic children (Steiner, 2011); and improved hospitalisation rates, employment/educational attainment, and intrapersonal outcomes such as self-efficacy and sense of hope (Tse et al., 2016). In education, approaches to utilising strong autistic interests effectively are blossoming (Bianco et al., 2009; Campbell & Tincani, 2011; R. Wood, 2019). Neurodiversity Celebration Week, a yearly occurrence since 2018, specifically reaches out to schools and universities encouraging the recognition of talents and advantages of neurodivergent students (Palumbo, 2023). Neurodiversity has been raised as the next frontier for improving inclusion in the workplace (Salman, 2019), encouraging employers to incorporate neurodivergent employees by framing their strengths as a competitive advantage (Austin & Pisano, 2017; Curry, 2019; V. Wood, 2019). Suggesting that these approaches do work, multinational corporations like SAP are now specifically seeking neurodivergent employees through their Autism at Work program and consulting firms like Auticon have been established to place autistic employees in IT roles.

Although there are not yet large-scale studies on outcomes of such efforts, strengths-based perspectives seem to be increasing educational and employment opportunities for autistic people, potentially with subsequent effects of supporting mental health and general well-being. In allowing spaces to exist where individuals' weaknesses are supported and their strengths incorporated, we may be replicating the ancestral environment that caused that very cognitive diversity to flourish.

6/ Areas of disagreement and on-going debate

Lastly, the areas where the evolutionary and neurodiversity approaches diverge — and even actively conflict — must be noted. Neurodiversity's social model allows extension of the 'neurodivergent' label to any individual with any neurotype, rejecting any sense of 'normal' or 'health' to the traits under its umbrella (Walker, 2013). By contrast, one of evolutionary psychiatry's fundamental scientific appeals is grounding the distinction between health and disorder with reference to evolutionary dysfunction (Hunt et al., 2022). This has been primarily forwarded by Jerome Wakefield's (1992, 2015) 'harmful dysfunction' definition of disorder, which claims that disorder exists wherever there is concurrent harm, defined by social and individual values, and dysfunction, defined by interruption of evolutionarily selected effects. This definition specifically aims to ground definitions of disorder in objective scientific facts, ensuring that diagnoses of health and disease are not merely social.

As a result, evolutionary psychiatry inevitably subtypes the autism spectrum into multiple classes with distinct aetiologies (Del Giudice & Haltigan, 2021; Hunt, 2023). While most cases of autism have no clear pathological cause, roughly 25% of cases (dependent on inclusion criteria) are described as syndromic autism, as they are associated with somatic abnormalities or a clear biomedical cause, such as genetic mutations (e.g., fragile X syndrome, 16p11.2 duplication) or foetal/neonatal damage (e.g., foetal valproate syndrome, obstetric complications) (Fernandez & Scherer, 2017). As these cases of autism are not evolutionarily paradoxical (Keller & Miller, 2006), they do not warrant explanations referencing adaptive traits. The evolutionary approach could thus be perceived as callous and discriminatory by neurodiversity advocates who consider all cases of autism as identical examples of neurodiversity. Similarly, adaptive explanations are not required for cognitive differences arising from other *de novo* genetic causes like Down Syndrome, which often result in intellectual disability. While these conditions are all examples of neurodiversity and can be advocated for under that banner, any scientific approach to understanding cognitive divergence will discriminate between heterogenous cases.

Scientific calls for discerning autistic subtypes are not unique to evolutionary approaches. Even within cases of nonsyndromic autism, high levels of heterogeneity continue to pose challenges to both research and clinical practice. Informed by clinical experience, Mottron and Gagnon (2023) propose more specific diagnostic criteria for what they term "prototypical autism". In this model, there is a developmental trajectory from ages 2 to 5 where most children increasingly process language and information may in a socially-biased manner. In the absence of this bias, children without neurodevelopmental impairment will develop interest in complex information independent of its social content, resulting in the more homogenous phenotype of prototypical autism. In this way, the authors frame prototypical

autism as the non-dominant—but certainly not pathological—outcome of a dynamic biological system, similar to being left-handed. As described above, this rarer developmental trajectory could lead to the higher rates of specialised talents and strengths among "prototypical" autistic people of normal to high intelligence (e.g. Meilleur, Jelenic and Mottron 2015).

To some neurodiversity advocates, suggestions that prototypical autism warrants evolutionary explanations and syndromic cases of autism do not may be unacceptably discriminatory. This is where the major distinction between the evolutionary psychiatry and neurodiversity paradigms emerge: while scientific justification of neurodiversity is constrained to empirical evidence, social norms can be extended indefinitely. As a matter of principle, every cognitive difference can be equally included under the 'neurodiversity' banner – whether by self-identification (Aftab, 2021) or otherwise.

This had led to major criticisms that the neurodiversity movement makes "implausible claims about the distinction between difference and disorder" (Nelson, 2020). Indeed, some individuals are unhappy with their neurodivergent traits and prefer to label them as a disorder, not mere difference. Criticism has also come from parents and family members of severely disabled autistic individuals (e.g. Clements 2019; Escher 2023), as the claim of difference and not disease, strengths over weaknesses, and acceptance rather than cure is contradictory to their experience of loved ones faced with serious disability. Even Judy Singer, who coined the term 'neurodiversity', intended it to be applied to people with normal or high intelligence and thinks excessive inclusivity is unrealistic (Lutz, 2023). Neurodiversity advocates may argue that the core of this distinction should rest in the hands of the affected individuals, particularly whether they perceive their differences as impacting their well-being, which follows the general trend in contemporary psychiatry to diagnose mental states as disorder wherever there is substantial harm. However, this doesn't provide a reliable scientific basis for attributing difference or disorder, leaving the disagreement to be resolved as a purely social matter.

Scientific approaches from evolutionary psychiatry could calm this debate by meeting this disagreement halfway, pushing back on assumptions of simple pathology dominant in contemporary psychiatry while making sense of stark differences by recognising the heterogeneity within wide diagnostic labels like autism spectrum disorder. This is not to say that strength-based approaches, social accommodations, and greater acceptance are not valid for cases of neurodiversity with an identifiable genetic or developmental pathology; rather, evolutionary psychiatry adds nuance that helps resolve the conflation of diverse conditions under singular banners. While the 'neurodiversity' term can encompass all biological causes of differences in minds, acknowledgement of distinct biological causes seems useful for reducing confusion and encouraging productive debate, which is sorely needed (Baron-Cohen, 2019).

Without acknowledgement of heterogeneity of forms of neurodivergence like autism, knowledge of certain causes, and varying levels of disability, critics of the neurodiversity movement may dismiss the *whole* neurodiversity paradigm's validity. As this would unfairly exclude the vast majority of individuals for whom no such cause has been found, evolutionary

explanations might be critical. As both evolutionary psychiatry and neurodiversity continue to grow, it seems prescient to consider how the science may reinforce and support the social goals, as has been recognised as important in the context of challenging sexual prejudice (Rinaldi, 2022) and racism (Donovan et al., 2019).

7/ Conclusion

Evolutionary psychiatry and neurodiversity are novel paradigms developing in parallel with different but complementary aims. While evolutionary psychiatry stipulates scientific theory and neurodiversity stipulates social response, both ultimately advocate for broadening the definition of 'normal', rejecting cures, recognising the role of the environment, and highlighting that differences can be strengths. Notably, scientific perspectives such as evolutionary psychiatry may provide powerful justification of the core ideas of neurodiversity: it could become not just good or socially desirable to agree with the principles of the neurodiversity movement, but an accurate and true understanding of evolutionary processes explaining cognitive differences captured under the umbrella of neurodiversity. Within a world believing in evolutionary explanations for neurodiversity, we would expect shifts in how education, employment, and public spaces include neurodivergent people that are largely in line with what neurodiversity advocates ask for. How much neurodiversity advocates care about integrating scientific explanation into their advocacy is up to them; but of all the psychiatric paradigms, the evolutionary approach may be the one that makes the most sense.

Bibliography

Abrams, M. (2020). *The New CBT: Evolutionary Clinical Psychology*. Cognella Press.

Aftab, A. (2021). The Neurodiversity Paradigm in Psychiatry: Robert Chapman, PhD. *Psychiatric Times*. https://www.psychiatrictimes.com/view/neurodiversity-paradigm-psychiatry

Austin, R. D., & Pisano, G. P. (2017). Neurodiversity as a Competitive Advantage. *Harvard Business Review*. http://www.k12accountability.org/resources/Special-Education/Neurodiversity_Is_a_Competitive_Advantage.pdf

Baron-Cohen, S. (2002). Is Asperger Syndrome Necessarily Viewed as a Disability? *Focus on Autism and Other Developmental Disabilities*, *17*(3), 186–191. https://doi.org/10.1177/10883576020170030801

Baron-Cohen, S. (2019). *The Concept of Neurodiversity Is Dividing the Autism Community*. Scientific American Blog Network. https://blogs.scientificamerican.com/observations/the-concept-of-neurodiversity-is-dividing-the-autism-community/

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent ? A test with children with Asperger Syndrome. *Learning*, *5*(1), 47–78.

Baron-Cohen, S., Wheelwright, S., Stone, V., & Rutherford, M. (1999). A mathematician, a physicist and a computer scientist with Asperger syndrome: Performance on folk psychology and folk physics tests.

Neurocase, *5*(6), 475–483. https://doi.org/10.1080/13554799908402743

Bianco, M., Carothers, D. E., & Smiley, L. R. (2009). Gifted Students With Asperger Syndrome. *Intervention in School and Clinic*, *44*(4), 206–215. https://doi.org/10.1177/1053451208328827

Campbell, A., & Tincani, M. (2011). The Power Card strategy: Strength-based intervention to increase direction following of children with autism spectrum disorder. *Journal of Positive Behavior Interventions*, *13*(4), 240–249. https://doi.org/10.1177/1098300711400608

Chapman, R. (2021). Neurodiversity and the Social Ecology of Mental Functions. *Https://Doi.Org/10.1177/1745691620959833*, *16*(6), 1360–1372. https://doi.org/10.1177/1745691620959833

Clarke, T. K., Lupton, M. K., Fernandez-Pujals, A. M., Starr, J., Davies, G., Cox, S., Pattie, A., Liewald, D. C., Hall, L. S., Macintyre, D. J., Smith, B. H., Hocking, L. J., Padmanabhan, S., Thomson, P. A., Hayward, C., Hansell, N. K., Montgomery, G. W., Medland, S. E., Martin, N. G., … Mcintosh, A. M. (2016). Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular Psychiatry*, *21*(3), 419–425. https://doi.org/10.1038/MP.2015.12

Clements, T. (2019). *What is autism? How the term became too broad to have meaning any more | Tom Clements | Opinion | The Guardian*. The Guardian. https://www.theguardian.com/commentisfree/2019/aug/26/autism-neurodiversity-severe

Crespi, B. (2016). Autism as a disorder of high intelligence. *Frontiers in Neuroscience*, *10*(JUN), 300. https://doi.org/10.3389/fnins.2016.00300

Crespi, B., & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences*, *31*(3), 241–261; discussion 261-320. https://doi.org/10.1017/S0140525X08004214

Curry, S. (2019). *Neurodiversity: A Competitive Advantage in Cybersecurity*. Forbes. https://www.forbes.com/sites/samcurry/2019/05/13/neurodiversity-a-competitive-advantage-in-cybersecurity/

Del Giudice, M., & Haltigan, J. D. (2021). An integrative evolutionary framework for psychopathology. *Development and Psychopathology*. https://doi.org/10.1017/S0954579421000870

Donovan, B. M., Semmens, R., Keck, P., Brimhall, E., Busch, K. C., Weindling, M., Duncan, A., Stuhlsatz, M., Bracey, Z. B., Bloom, M., Kowalski, S., & Salazar, B. (2019). Toward a more humane genetics education: Learning about the social and quantitative complexities of human genetic variation research could reduce racial bias in adolescent and adult populations. *Science Education*, *103*(3), 529–560. https://doi.org/10.1002/SCE.21506

Duan, W., Ho, S. M. Y., Tang, X., Li, T., & Zhang, Y. (2014). Character Strength-Based Intervention to Promote Satisfaction with Life in the Chinese University Context. *Journal of Happiness Studies*, *15*(6), 1347–1361. https://doi.org/10.1007/s10902-013-9479-y

Escher, J. (2023). The Autism Surge: Lies, Conspiracies, and My Own Kids | The Free Press. *The Free Press*. https://www.thefp.com/p/the-autism-surge-lies-conspiracies

Fernandez, B. A., & Scherer, S. W. (2017). Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. *Dialogues in Clinical Neuroscience*, *19*(4), 353. https://doi.org/10.31887/DCNS.2017.19.4/SSCHERER

Fletcher-Watson, S., Adams, J., Brook, K., Charman, T., Crane, L., Cusack, J., Leekam, S., Milton, D., Parr, J. R., & Pellicano, E. (2019). Making the future together: Shaping autism research through meaningful participation. *Autism*, *23*(4), 943–953. https://doi.org/10.1177/1362361318786721

Gizzonio, V., Avanzini, P., Fabbri-Destro, M., Campi, C., & Rizzolatti, G. (2014). Cognitive abilities in siblings of children with autism spectrum disorders. *Experimental Brain Research*, *232*(7), 2381–2390. https://doi.org/10.1007/S00221-014-3935-8

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of

the adaptationist programme. *Proceedings of the Royal Society of London - Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Cullen, B., Malik, R., Worrall, B. B., Sudlow, C. L. M., Wardlaw, J. M., Gallacher, J., Pell, J., McIntosh, A. M., Smith, D. J., Gale, C. R., & Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular Psychiatry 2016 21:11*, *21*(11), 1624–1632. https://doi.org/10.1038/mp.2015.225

Hunt, A. D. (2023). Making Wakefield Workable: The Fitness and Function Framework for Taxonomising Evolutionary Dysfunction. *PsyArxiv*. https://doi.org/10.31234/OSF.IO/QBGKE

Hunt, A. D., Abed, R., & St John-Smith, P. (2022). The Biopsychosocial Model Advanced by Evolutionary Theory. In R. Abed & P. St John-Smith (Eds.), *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge University Press.

Hunt, A. D., & Jaeggi, A. V. (2022). Specialised minds: extending adaptive explanations of personality to the evolution of psychopathology. *Evolutionary Human Sciences*, *4*, e26. https://doi.org/10.1017/EHS.2022.23

Hunt, A. D., St-John Smith, P., & Abed, R. (2023). Evobiopsychosocial medicine. *Evolution, Medicine, and Public Health*, *11*(1), 67–77. https://doi.org/10.1093/EMPH/EOAC041

Insel, T. (2013). Transforming diagnosis. In *NIMH Director's Message*. NIMH (Natl. Instit. Mental Health). https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml

Kapp, S. K., Pripas-Kapit, S., Dekker, M., Murray, D., Tisoncik, L. A., Baggs, M., Seidel, K., Buckle, K. L., Evans, M., Raymaker, D. M., DaVanport, S., Greenburg, C., Des Rochas Rosa, S., Ne'eman, A., Neumeier, S. M., Brown, L. X. Z., Arnold, L., Elder Robison, J., Onaiwu, M. G., … Russell, G. (2020). *Autistic Community and the Neurodiversity Movement - Stories from the Frontline* (S. K. Kapp (ed.)). Palgrave MacMillan. https://link.springer.com/content/pdf/10.1007%2F978-981-13-8437-0.pdf

Keller, M. C., & Miller, G. (2006). Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behavioral and Brain Sciences*, *29*(4), 385–404. https://doi.org/10.1017/S0140525X06009095

Lai, M. C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. In *The Lancet Psychiatry* (Vol. 2, Issue 11, pp. 1013–1027). Elsevier Ltd. https://doi.org/10.1016/S2215-0366(15)00277-1

Lewis, B. (2006). A mad fight: Psychiatry and disability activism. *The Disability Studies Reader*, *3*.

Loughman, A., & Haslam, N. (2018). Neuroscientific explanations and the stigma of mental disorder: a meta-analytic study. *Cognitive Research: Principles and Implications*, *3*(1), 43. https://doi.org/10.1186/s41235-018-0136-1

Lutz, A. (2023). An Interview with Neurodiversity Originator Judy Singer. *Psychology Today*. https://www.psychologytoday.com/intl/blog/inspectrum/202306/an-interview-with-neurodiversity-originator-judy-singer

Mayr, E. (1961). Cause and effect in biology. *Science*, *134*(3489), 1501–1506. https://doi.org/10.1126/science.134.3489.1501

Meilleur, A. A. S., Jelenic, P., & Mottron, L. (2015). Prevalence of Clinically and Empirically Defined Talents and Strengths in Autism. *Journal of Autism and Developmental Disorders*, *45*(5), 1354–1367. https://doi.org/10.1007/s10803-014-2296-2

Mottron, L., & Gagnon, D. (2023). Prototypical autism: New diagnostic criteria and asymmetrical bifurcation model. *Acta Psychologica*, *237*. https://doi.org/10.1016/J.ACTPSY.2023.103938

Nelson, R. H. (2020). A Critique of the Neurodiversity View. *Journal of Applied Philosophy*, japp.12470. https://doi.org/10.1111/japp.12470

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

Land.

Nesse, R. M. (2023). Evolutionary psychiatry: foundations, progress and challenges. *World Psychiatry*, *22*(2), 177–202. https://doi.org/10.1002/WPS.21072

Noland, J. S., Steven Reznick, J., Stone, W. L., Walden, T., & Sheridan, E. H. (2010). Better working memory for non-social targets in infant siblings of children with Autism Spectrum Disorder. *Developmental Science*, *13*(1), 244. https://doi.org/10.1111/J.1467-7687.2009.00882.X

O'Riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(3), 719–730. https://doi.org/10.1037//0096-1523.27.3.719

Palumbo, J. (2023). *What Are The Takeaways Of Neurodiversity Celebration Week*. Forbes. https://www.forbes.com/sites/jenniferpalumbo/2023/03/17/what-are-the-takeaways-of-neurodiversity-celebration-week/

Pinder, M. (2017). Does Experimental Philosophy Have a Role to Play in Carnapian Explication?: X-Phi and Explication. *Ratio*, *30*. https://doi.org/10.1111/rati.12164

Ploeger, A., & Galis, F. (2011). Evolutionary approaches to autism- an overview and integration. *McGill Journal of Medicine : MJM : An International Forum for the Advancement of Medical Sciences by Students*, *13*(2), 38. http://www.ncbi.nlm.nih.gov/pubmed/22363193

Plunkett, D., & Cappelan, H. (2020). "A Guided Tour of Conceptual Engineering and Conceptual Ethics. In A. Burgess, H. Cappelan, & D. Plunkett (Eds.), *Conceptual Engineering and Conceptual Ethics*. Oxford University Press. https://doi.org/10.1093/oso/9780198801856.001.0001

Prontzos, P. G. (2019). *The Concept of Race Is a Lie*. Scientific American Blog Network. https://blogs.scientificamerican.com/observations/the-concept-of-race-is-a-lie/

Rinaldi, A. (2022). I was born this way: New research confirms that a mix of prenatal factors and genetic differences could explain human sexual orientation. *EMBO Reports*, *23*(6). https://doi.org/10.15252/EMBR.202255290

Salman, S. (2019). *Simon Baron-Cohen: 'Neurodiversity is the next frontier. But we're failing autistic people' | Society | The Guardian*. The Guardian. https://www.theguardian.com/society/2019/oct/02/simon-baron-cohen-autism-neurodiversity-brains-money

Schroder, H. S., Devendorf, A., & Zikmund-Fisher, B. J. (2023). Framing depression as a functional signal, not a disease: Rationale and initial randomized controlled trial. *Social Science & Medicine*, *328*, 115995. https://doi.org/10.1016/J.SOCSCIMED.2023.115995

Shakespeare, T. (2006). The social model of disability. *The Disability Studies Reader*, *2*, 197–204.

Sikela, J. M., & Searles Quick, V. B. (2018). Genomic trade-offs: are autism and schizophrenia the steep price of the human brain? *Human Genetics*, *137*(1), 1–13. https://doi.org/10.1007/S00439-017-1865-9

Steiner, A. M. (2011). A strength-based approach to parent education for children with autism. *Journal of Positive Behavior Interventions*, *13*(3), 178–190. https://doi.org/10.1177/1098300710384134

Szasz, T. (1974). *The myth of mental illness: foundations of a theory of personal conduct,*. Harper & Row. https://books.google.co.uk/books?id=x__aAAAAMAAJ&q=ISBN&redir_esc=y

Tse, S., Tsoi, E. W., Hamilton, B., O'Hagan, M., Shepherd, G., Slade, M., Whitley, R., & Petrakis, M. (2016). Uses of strength-based interventions for people with serious mental illness: A critical review. *International Journal of Social Psychiatry*, *62*(3), 281–291. https://doi.org/10.1177/0020764015623970

Wakefield, J. C. (1992). The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values. *American Psychologist*, *47*(3), 373–388. https://doi.org/10.1037/0003-066X.47.3.373

Wakefield, J. C. (2015). Biological Function and Dysfunction: Conceptual Foundations of Evolutionary Psychopathology. In *The Handbook of Evolutionary Psychology* (pp. 1–19). John Wiley & Sons, Inc.

https://doi.org/10.1002/9781119125563.evpsych242

Walker, N. (2013). *THROW AWAY THE MASTER'S TOOLS: LIBERATING OURSELVES FROM THE PATHOLOGY PARADIGM*. Neuroqueer: The Writings of Dr Nick Walker. https://neuroqueer.com/throw-away-the-masters-tools/

Weiss, B. (2010). *Life Among the "Yakkity Yaks."* Wall Street Journal. https://www.wsj.com/articles/SB10001424052748703525704575061123564007514

Wood, R. (2019). Autism, intense interests and support in school: from wasted efforts to shared understandings. *Educational Review*, 1–21. https://doi.org/10.1080/00131911.2019.1566213

Wood, V. (2019). *GCHQ targeting dyslexic and neurodiverse people in recruitment drive, spy chief says*. The Independent. https://www.independent.co.uk/news/uk/home-news/gchq-jobs-recruitment-intelligence-spy-jeremy-fleming-dyslexia-disability-neurodiversity-a9163996.html

Evolving Evolutionary Psychiatry and Explaining Neurodiversity – Conclusion

"In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation."
Charles Darwin (1859 p. 488). The Origin of Species.

1/ Motivations for Evolving Evolutionary Psychiatry

Evolutionary psychiatry offers a tantalising prospect: move beyond seeking elusive simplistic pathological explanations for common and long-unexplained mental disorders, utilising instead the theory explaining all biological variation to further explain the cognitive variation medicalised as psychopathology.

Psychiatry is a discipline with a history dappled by fads and phases; from quirkier approaches referencing perverted unconscious desires, to gruesome movements wantonly lobotomising adults and children (Phillips 2013). In more recent decades, renewed versions of symptom-based psychiatric diagnosis manuals haven't helped disorders be linked with specific pathologies (Shorter 2015), and attempts to achieve such insight by revising nosology persist today (Cuthbert 2022). The 1990s, pronounced the 'decade of the brain' (Jones and Mendell 1999), were expected to reveal neurological correlates behind major mental disorders and offer pathways to prevention and treatment, but came and went without the hoped for success – and after two more decades of research, the replication crisis has undermined much of the effort (Nour, Liu and Dolan 2022). The human genome project was hoped to reveal specific disease genes underlying psychiatric conditions, but pushed the field into even more uncertainty (Cowan, Kopnisky and Hyman 2002) and blurred boundaries between disease states and normal variation (Geschwind and Flint 2015). In a statement noted multiple times throughout this dissertation and worth repeating one final time, Thomas Insel, leader of the National Institute of Mental Health who oversaw around $20 billion of research spending (Rogers 2017) said, 'Whatever we've been doing for five decades, it ain't working ... Maybe we just need to rethink this whole approach ... With

no validated biomarkers and too little in the way of novel medical treatments since 1980 ... it is time to rethink mental disorders' (Greenberg 2013).

Evolutionary psychiatry undoubtedly offers a way to rethink mental disorders – and the steady scientific development of the paradigm and theoretical foundations of the evolutionary sciences it relies upon are unparalleled in psychology – certainly not born of mere fad or phase. However, despite its face appeal, various fundamental problems remain as barriers to its widespread adoption. The primary aim of this dissertation has been to surmount some of them, putting evolutionary psychiatry on firmer footing for future development.

Chapter 1 introduced the dissertation by providing an overview of the theoretical motivation for explaining personality and psychopathological traits with unified evolutionary theory. Treating psychopathology as distinct from normal psychology derives from the social impetus to treat harm, but has mostly not led to discovery of genuine pathophysiological cause. Both personality and psychopathological traits are moderately or highly heritable; early onset and fairly stable; show non-shared environmental effects, and are best understood as continuous traits associated with diffuse, complex brain differences. There is no good scientific justification for explaining these 'normal' and 'abnormal' individual differences with separate theoretical frameworks. Both are somewhat evolutionarily paradoxical, and both warrant explanation at the ultimate, evolutionary level.

Chapter 1 noted that the social dynamics of human groups offer an ideal environment for evolutionary dynamics leading to adaptive individual differences in cognition. The evolution of specialised minds, to cooperate (or cheat) within closely connected human groups should be expected, and is a plausible explanation for a range of our stable cognitive differences, many of which may fall under the banner of 'neurodiversity' (Chapter 3 considers the evidence for this type of explanation of autism in detail). Deeper consideration of the social dynamics and available niches for different cognitive styles in ancestral human groups is critical, and Chapter 1 ends with a particular novel theoretical contribution, noting that we can use such information to address a longstanding question in philosophy of medicine (Rogers and Walker 2017) regarding the difficulty of distinguishing disorder upon continuous trait spectrums: a minimum adaptive prevalence of one specialised individual per extended hunter-gatherer social group is plausible, so a non-arbitrary line can be drawn at around 0.6% of the population, below which adaptive individual differences are highly unlikely to exist. Although this calculation is highly simplified and makes various assumptions (e.g. it doesn't account for environmental effects), future research could expand upon it, and also upon the overarching insight that we can make more use of anthropological data of likely ancestral human social dynamics to make evolutionary inferences – considered in more detail in Chapter 3.

2/ Formulating Solid Foundations

Once the theoretical motivation to pursue evolutionary psychiatry is established, several intricate conceptual and scientific problems remain. Most essentially, some kind of clear scientific paradigm for testing hypotheses in evolutionary psychiatry is required. This is no

simple task: the problem of testing evolutionary hypotheses, which are fundamentally historical, has been long recognised and led Karl Popper to claim, at one point, that such hypotheses are untestable and thus not truly scientific (Sonleitner 1986). The challenge of formulating and assessing scientific hypotheses in evolutionary psychiatry is magnified over normal evolutionary biology and psychology because its hypotheses don't merely suggest a function for a currently useful trait ('eyes are for seeing'). Evolutionary psychiatry considers borderlines between function and dysfunction, seeking the original function of systems which are now failing, or are maladaptive in industrialised environments, or are harmful for health but helpful for fitness, or any of many other complex explanations (Nesse 2005; Matthewson and Griffiths 2017; Al-Shawaf *et al.* 2020; Abed and St John-Smith 2022).

Such complexities are particularly frustrating because evolutionary theory offers a simple, unique benefit which should warrant its place at the foundation of both medicine and psychiatry: providing a truly objective standard for health and disorder. This has been best forwarded by Jerome Wakefield, in his 'harmful dysfunction' analysis (HDA) of disorder (Wakefield 1992, 1997, 2015), which is the single contribution to the field of evolutionary psychiatry that this dissertation has most built upon. Wakefield proposed that disorders must be both harmful, in a value-defined sense, and dysfunctional, in an evolutionary sense. The advantage of utilising an evolutionary sense of dysfunction is that it provides, unlike any other scientific approach, a sound frame of reference for what biology 'should' be doing: natural selection has shaped biology, so what biology 'should' have been doing was achieving reproductive success, which explains its current form and functional capacity. No other scientific paradigm offers such objective normativity. Despite this unique advantage, and general consensus that evolutionary theory lends the most reliable objective framing for distinguishing health and disorder (Wright 1976; Millikan 1989; Neander 1991), Wakefield's specific definition of the 'dysfunction' component as 'interruption of selected effects' has been criticised on multiple fronts (Faucher and Forest 2021). Before conducting a rigorous evolutionary analysis allowing insight into the ultimate causes and proper identification of disorder, we require a reliably precise definition of function and dysfunction. Laying out that definitional framework was the motivation for Chapter 2.

The core move that Chapter 2 made to advance Wakefield's definition of dysfunction revolved around the fundamental quantifiable vector of evolution by natural selection: reproductive success, or fitness. The 'fitness and function framework' (FAFF) shifts the definition of dysfunction away from relying solely upon descriptions of hard-to-specify qualitative processes describing selection effects, instead placing the critical distinction upon differences in reproductive success. Describing the function of hands is much harder than describing the fitness effects of hands; the same obviously goes for depression, and all the psychological traits of interest to evolutionary psychiatry. The FAFF to some extent mirrored a leading definition of adaptation and function in the philosophy of biology (Sober, 2000, p.85) but was adapted for the purpose of elucidating different senses of *dys*function and dealing with the complex cases in psychiatry, where mental disorder description criteria may often miss critical relevant phenotypes (rectified scientifically by using the method in Chapter 3). A particular theoretical strength of the FAFF is that, once a trait has its description and environmental framing specified, its fitness can only be positive, neutral, or

negative: this allows a level of precise demarcation of forms of dysfunction impossible if relying on the qualitative descriptions of dysfunction that previous taxonomies of disorder in evolutionary psychiatry and medicine have utilised (Nesse 2005; Matthewson and Griffiths 2017; Al-Shawaf *et al.* 2020; Abed and St John-Smith 2022).

Many psychiatric disorders categorised in the FAFF will not be 'simple disruptions' but the result of adaptations with trade-offs or manifesting harmfully due to environmental changes. Where within these distinct classes of dysfunction we can truly define 'disorder' here is left open – the aim of Chapter 2 was merely to go beyond the overly simplistic function/dysfunction dichotomy and provide the range of possible strictly definable classes of dysfunction which will stand up to philosophical scrutiny and can be reliably scientifically investigated. That task of scientific investigation is picked up in Chapter 3, which is presented as a standalone chapter aimed at evolutionary scientists hoping to engage in rigorous assessment of possible evolutionary explanations of mental disorders.

The just-so storytelling critique has long been levelled at evolutionary psychology (Gould and Lewontin 1979; Dupré 2003; Smith 2019), and the problem of speculative storytelling with insufficient evidence is magnified in complex cases of mental disorder related to functional capacities which are now obscured. However, plentiful evidence exists from across genetics, neuroscience and epidemiology, with years of accumulated research into the traits we have been labelling as psychopathology. Such evidence can be taken as clues of the past evolutionary process, despite its invisibility. This leads to the possibility of developing a method of systematic review with standardised analytical principles aimed at inferring past selection processes. Chapter 3 thus presented the DCIDE method, which firstly Describes a trait, Categorises subtypes as non-adaptive or potentially adaptive, then Integrates evolutionarily visible correlates of potentially adaptive cases – these initial stages hone analyses upon the correct object of adaptive explanation. After this, Depicted explanatory hypotheses are Evaluated for their sufficiency to explain the totality of evidence, including areas of circumstantial evidence often haphazardly referenced by evolutionary scholars. In the DCIDE method these are assessed under standardised, theoretically derived analytical principles, avoiding accusations of cherry-picking or just-so storytelling as much as possible.

This provides an opportunity for an unprecedentedly rigorous assessment of the evolutionary explanations of autism. The autism spectrum has some non-adaptive subtypes Categorised and considered explained (primarily resulting from *de novo* mutations and foetal traumas), and a wider spectrum of subclinical and female cases Integrated as the proper target of adaptive explanation. Depiction and Evaluation lead to the systemising social niche specialisation hypotheses of Baron-Cohen (2020) and Del Giudice (2018) being validated as fairly sufficiently accounting for the circumstantial evidence of autism's age of effect, duration, sex differences, environmental factors and prevalence of genetic propensity, whilst Crespi's (2016) high intelligence by-product model struggles.

The DCIDE method is aimed specifically at analysing psychopathological trait descriptors which initially capture etiologically heterogenous cases and potentially exclude relevant adaptations, but its broad principles, particularly in the stages of Depiction and Evaluation, could also be used in evolutionary psychology more generally to assess hypothesis strength,

preventing cherry-picking and defending such hypotheses against criticisms of just-so storytelling. The major limitation facing the DCIDE method in imminently providing a range of confident explanations in evolutionary psychiatry derives primarily from the empirical evidence required, particularly in cases where previous research has utilised insufficient initial Description criteria (e.g. excluding subclinical cases). Its usefulness as a systematic review is limited without evidence of appropriately Described and Integrated traits – as are all attempts at evolutionary hypotheses. However, the DCIDE method's structure could also be used to direct such evidence collection (for example, by encouraging a study design which recruits many schizophrenic (or autistic, dyslexic, etc.) individuals, excludes cases Categorising as non-adaptive, Integrates the family members of the potentially adaptive cases into the study, and collects 'direct evidence' of their functioning). Chapter 3 thus offers multiple pathways for a rigorous and holistic empirical paradigm pursuing evolutionary psychiatry in the future.

Because Chapters 2 and 3 were presented as standalone papers, an explication of their relationship is appropriate. At the broadest level, Chapter 2 is a definitional framework, Chapter 3 is the scientific method for implementing it, but certain complexities need mentioning. Some direct overlaps exist between the chapters: for example, the class of 'simple disruption' in Chapter 2 identifies the same cases which Categorise as non-adaptive in Chapter 3. The description specifier noted as critical in Chapter 2 is encompassed in the first stage of the DCIDE method. However, there are places where coherence is less clear. Framing specifiers are not mentioned in Chapter 3, despite being proposed in Chapter 2 as necessary. The method in Chapter 3 does not explicitly aim at fitting explanations of mental disorder into the five classes of dysfunction identified in Chapter 2. Nor does Chapter 3 solely concentrate on fitness effects, but provides a framework for inferring past functional processes, despite Chapter 2 emphasising the difficulty of specifying such processes.

The justification for these differences is primarily in the different audiences and problems the separate chapters aim to address. Chapter 2 aims at philosophical foundations and precise definitional frameworks, primarily the concern of philosophers who seek conceptual clarity. Chapter 3 aims at scientific methods and improved inference from objective evidence, primarily aimed at evolutionary scientists who engage in hypothesising and hypothesis-testing without particular concern over imprecision in definitions. Framing specifiers weren't emphasised in Chapter 3 because most evolutionary scientists broadly specify framing already (including in the accounts of autism Depicted), albeit informally. In the Depiction and Evaluation stages, the DCIDE method emphasises assessment of the evolutionary relevance of direct and circumstantial evidence, forcing consideration of framing, even if not specified explicitly. Regarding the classes of dysfunction identified in Chapter 2 but unmentioned in Chapter 3, although more philosophically rigorous and scientifically tractable, they are also less qualitatively informative. Stories of ancestral function are generally more appealing to evolutionary scientists (and others) than broad fitness-based classification. Chapter 2's strong arguments that these carry inevitable imprecision don't mean such stories *shouldn't* be sought and tested – functional processes are still necessary to explain fitness effects. Chapter 3 supports the entire scientific investigation as much as possible.

Despite their differences, born of addressing different audiences, Chapters 2 and 3 are entirely coherent, and indeed reinforcing. Evolutionary scientists applying imprecise definitions of function and dysfunction were open to the criticisms Chapter 2 addresses (even if these criticisms were rarely raised or noticed), so Chapter 2 prevents this criticism being raised of the scientific method proposed in Chapter 3. Chapter 3 supports Chapter 2 because philosophers might accept the FAFF's analytical consistency but complain that evolutionary history is still scientifically obscure – indeed, this was one of the major complaints against Wakefield's dysfunction component (Bergner 1997; Kennair 2010). This would imply the FAFF is impotent. The DCIDE method addresses this complaint.

Moreover, researchers who take Chapter 2's contributions on board can improve their chances of formulating strong Depictions for assessment in the DCIDE method. Explanatory hypotheses framed within the FAFF benefit from its specificity and tractability if they simultaneously suggest hypotheses of fitness-based classification and functional processes. For example, one could Depict individuals with a broad autism phenotype as exhibiting a 'trade-off' (their autistic traits involve positive and negative fitness effects, but are overall positive or neutral) and autistic individuals meeting strict diagnostic criteria as cases of 'distributed disruption' (their traits are individually costly to fitness but the predisposing genetic variants have positive effects in other individuals in the population); social niche specialisation for systemising could be further suggested as the primary functional processes explaining these fitness effects. In Evaluation, the evidence should more uncontroversially support placements into the fitness-related classes of Chapter 2's taxonomy than validate the functional effects (e.g. the high prevalence of subclinical autistic phenotypes and fairly low prevalence of diagnosable autism makes distributed disruption more likely for diagnosable cases, no matter the functional explanation). Chapter 2 provides a framework to mitigate problems with lack of precision in specific functional explanations by being more precise about the class of trait we are looking at – Chapter 3 then provides an improved method for inferring both fitness effects and functional processes.

3/ How to Help

Presuming the establishment of sound philosophical foundations and a tractable scientific method for providing evolutionary explanations, a critical question remains for evolutionary psychiatry: 'how does this help?'. Chapters 4 and 5 consider this.

What is ultimately sought by impartial observers is specific, empirical evidence of how evolutionary psychiatry leads to improved outcomes: to reduction in harm, through either treatment or prevention. That support is the essential aim of psychiatry and clinical psychology. Every major previous paradigm in psychiatry claimed to have some specific answer, often wrapped up in its theoretical framing, of pathways through which it was poised to provide help: by relieving suppressed unconscious desires; by teaching different thoughts and behaviours; by releasing the trauma; by fixing the biological system which has gone awry. Although these paradigms have generally failed to deliver on their promises, they were almost always explicit about possible avenues of help. In contrast, evolutionary psychiatry has developed for decades as an explanatory paradigm, generally without pretense to offer new treatment regimes. Psychiatry and clinical psychology are directly

associated with psychopharmacology and therapy as modes of intervention, but most work in evolutionary psychiatry hasn't suggested specific advances in either of these treatment modalities (with limited exceptions speculating on improvements to cognitive behavioural therapy (Abrams 2020) or directing pharmacological research (Hunt, St-John Smith and Abed 2023)). The 'how does this help' question has thus gone fairly unanswered – with implications that the actual answer is 'it doesn't'.

This may be an underestimation of potential impact. Empirical work on this question should be an aim for researchers in coming years, but Chapters 4 and 5 provide a perspective on where evolutionary psychiatry seems best poised to affect humans experiencing mental health problems. They concentrate on factors beyond the modalities of therapy or pharmaceuticals.

Chapter 4 builds off of Wakefield's hybrid model for defining disorder to consider the general question of how scientific knowledge interacts with values in medical decision making. In making the HDA's dual components of harm and dysfunction equally necessary, it becomes inevitably decoupled from the goals of medicine. We treat fertility with contraception despite fertility being a critical biological function. We often choose to leave benign cancers under observation, and refrain from treatments if they may cause more harm than the disease. We don't require discovery of pathology in psychiatry before deciding the mental state should be alleviated. A criticism of Wakefield's HDA, and evolutionary psychiatry more generally, arises here – if it ultimately doesn't matter to medicine whether the target of treatment is evolutionarily dysfunctional, why bother with an evolutionary analysis?

Chapter 4 points out that a structural alteration to the hybrid HDA model, as 'hierarchical harmful dysfunction', better accounts for current and past medical normativity, and reframes the relevance of evolutionary explanations. The basic alteration is to say that harm and dysfunction are not equally *necessary*, but exist in a hierarchy: medicine seeks to heal harm as an overriding goal, but those values regarding harmfulness are informed by scientific – and evolutionary – considerations, in combination with all other considerations. In some cases, proving a trait is evolutionarily functional or dysfunctional will alter our values regarding the treatment of that trait. In others it will not. The chapter then argues that, to the HDA's credit, the facts about whether a trait is evolutionarily functional or dysfunctional holds particular sway over our values regarding a condition. Coming to believe in evolutionary explanations for autism, or ADHD, or depression, may significantly change our treatment of those conditions. Although not answering the 'how does this help' question with novel therapies, Chapter 4 provides an iteration of Wakefield's model that is better able to account for observed medical normativity whilst illuminating a pathway for evolutionary psychiatry to affect treatment and outcomes – through altering ideas and values.

Chapter 5 concludes the dissertation by considering an area where the effect of evolutionary explanations seems poised to align with a movement seeking to alter social values regarding mental disorders. Despite being founded upon entirely different aims, of scientific understanding on the one hand and social justice on the other, there are multiple overlaps between evolutionary psychiatry and the 'neurodiversity' movement (Kapp *et al.* 2020). The vector of improvement the neurodiversity movement calls for is not (generally) increased or

improved therapy or medication, but is at the societal level, mediated via changing ideas and values surrounding mental disorder. Neurodiversity advocates call for destigmatisation, integration, accommodation and greater recognition of the value of neurodivergent minds. Thus, both evolutionary psychiatry and the neurodiversity movement aim almost solely at reconceptualisation of mental disorders, differentially emphasising scientific and social motivation.

Despite differing in fundamental motivation, Chapter 5 notes various areas of overlap between evolutionary psychiatry and neurodiversity. Both paradigms reframe disorders as natural cognitive differences rather than disease. They both expand the concept of 'normal' to mental states and traits which are currently diagnosed as disorders. They focus on relative strengths, rather than emphasising weaknesses alone. Both recognise that modern environments disadvantage certain individuals, which can then be a primary cause of functional impairment. Both emphasise social accommodation and integration rather than treatments or cures. Both have been suggested to potentially help reduce stigmatisation. These overlaps are substantial enough that Chapter 5 suggests the neurodiversity movement could adopt theories and findings from evolutionary psychiatry (for example, those discussed in Chapters 1 and 3) as justification for the movement, increasing its efficacy in convincing parties who are more responsive to scientific rather than social reasons (e.g. in employment; in medical research).

In some ways, the scientific explanations offered by evolutionary psychiatry contradict the principles of the neurodiversity movement – for example, neurodiversity can be endlessly inclusive (e.g. to Downs Syndrome), but evolutionary approaches necessarily differentiate different types of dysfunction, as emphasised in Chapter 2, including some which are excluded from any kind of adaptive explanation, as Categorised in Chapter 3. Not every diagnosable case of mental disorder has the same evolutionary explanation, but all can be included as examples of neurodiversity by the social movement. Rather than interpreting this as an irreconcilable incompatibility, Chapter 5 suggest that this fundamental divergence offers an avenue to remedying a heated ongoing debate over the neurodiversity concept, born from the criticism that it ignores seriously disabled individuals (Clements 2019; Nelson 2020; Escher 2023). Neurodiversity advocates can advocate on behalf of individuals with traits falling into any of the five forms of dysfunction identified in Chapter 2, with any kind of adaptive or entirely non-adaptive causes. By recognising differences in causation and phenotype presentation, such advocacy is provided with necessary nuance, preventing accusations of ignoring these differences.

The most obvious path for evolutionary psychiatry to answer the question of 'how does this help' may take inspiration from the neurodiversity movement. Core to the neurodiversity paradigm is the idea that by reconceptualising 'disorders' as 'differences', or recognising neurodivergent strengths, *we will improve the lives of neurodivergent individuals*. It is in this realm of changing ideas around mental health and disorder, too, that evolutionary psychiatry may imminently improve lives. Comparisons of the stigmatisation effects of biogenetic (e.g. neurochemical, genetic) and stress explanations of mental disorder generally find that biogenetic explanations are somewhat more stigmatising (Loughman and Haslam 2018).

One study has found that providing evolutionary explanations of depression to individuals who had personally experienced depression reduced self-stigmatisation and increased willingness to reveal a depression diagnosis in comparison to explanations-as-usual (Schroder, Devendorf and Zikmund-Fisher 2023). This was the first example of a form of empirical research which could be much further developed to answer the 'how does this help' question. For example, if educating teachers on plausible evolutionary explanations of ADHD, autism and dyslexia, will they treat their neurodivergent students differently? If parents, when receiving a diagnosis for their child, are provided information regarding evolutionary explanations, would that affect their optimism about their child's future and encourage different styles of parenting? If employers were aware of evolutionary explanations for mental health conditions, including not only neurodivergent traits but states such as depression and anxiety, would they alter employment practices and workplace dynamics to better accommodate their workforce and mitigate problems, maximising both productivity and happiness? Most of all, how will evolutionary explanations affect diagnosed and self-diagnosed individual's perception of themselves and behaviour? The possibilities for improvement in outcomes here have great *prima facie* appeal.

Better understanding could improve how the diagnosed think about themselves, and how everyone else thinks about them, opening up possibilities for flourishing lives which are now obfuscated and misdirected by labels carrying connotations of lesser, worse, impaired, diseased. The longstanding criticisms of psychiatry for over-medicalisation and overly simplistic pathologising of mental states and traits have likely been scientifically justified in many cases – most common mental disorders are unlikely to be simple disruptions. Evolutionary psychiatry provides a more nuanced scientific perspective for understanding why dysfunctions in cognition are regular features of human minds.

Bibliography

Abed R, St John-Smith P. Introduction to Evolutionary Psychiatry. In: St John-Smith P, Abed R (eds.). *Evolutionary Psychiatry: Current Perspectives on Evolution and Mental Health*. Cambridge: Cambridge University Press, 2022, 1–18.

Abrams M. *The New CBT: Evolutionary Clinical Psychology*. San Diego, CA: Cognella Press., 2020.

Al-Shawaf L, Lewis DMG, Barbaro N *et al.* The Products of Evolution: Conceptual Distinctions, Evidentiary Criteria, and Empirical Examples. *SAGE Handb Evol Psychol* 2020:70–95.

Baron-Cohen S. *The Pattern Seekers: How Autism Drives Human Invention*. Basic Books, 2020.

Bergner RM. What is psychopathology? And so what? *Clin Psychol Sci Pract* 1997;**4**:235–48.

Clements T. What is autism? How the term became too broad to have meaning any more | Tom Clements | Opinion | The Guardian. *Guard* 2019.

Cowan WM, Kopnisky KL, Hyman SE. The Human Genome Project and Its Impact on Psychiatry. *Annu Rev Neurosci* 2002;**25**:1–50.

Crespi B. Autism as a disorder of high intelligence. *Front Neurosci* 2016;**10**:300.

Cuthbert BN. Research Domain Criteria (RDoC): Progress and Potential. *Curr Dir Psychol Sci* 2022;**31**:107–14.

Darwin C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. *darwin-online.org.uk* 1859.

Dupré J. *Human Nature and the Limits of Science*. Oxford: Clarendon Press, 2003.

Escher J. The Autism Surge: Lies, Conspiracies, and My Own Kids | The Free Press. *Free Press* 2023.

Faucher L, Forest D eds. *Defining Mental Disorder: Jerome Wakefield and His Critics*. The MIT Press, 2021.

Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. *Science* 2015;**349**:1489–94.

Del Giudice M. The Autism Spectrum. *Evolutionary Psychopathology*. Vol 4. 2018, 243–60.

Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc London - Biol Sci* 1979;**205**:581–98.

Greenberg G. The Rats of N.I.M.H. | The New Yorker. *New Yorker* 2013.

Hunt AD, St-John Smith P, Abed R. Evobiopsychosocial medicine. *Evol Med Public Heal* 2023;**11**:67–77.

Jones EG, Mendell LM. Assessing the decade of the brain. *Science* 1999;**284**:739.

Kapp SK, Pripas-Kapit S, Dekker M *et al. Autistic Community and the Neurodiversity Movement - Stories from the Frontline*. Kapp SK (ed.). Palgrave MacMillan, 2020.

Kennair LE. The Problem of Defining Psychopathology and Challenges To Evolutionary Psychology Theory. In: Buss DM, Hawley PH (eds.). *The Evolution of Personality and Individual Differences*. 2010, 451–79.

Loughman A, Haslam N. Neuroscientific explanations and the stigma of mental disorder: a meta-analytic study. *Cogn Res Princ Implic* 2018;**3**:43.

Matthewson J, Griffiths P. Biological Criteria of Disease: Four Ways of Going Wrong. *J Med Philos* 2017;**42**, DOI: 10.1093/jmp/jhx004.

Millikan R. In Defense of Proper Functions. *Philos Sci* 1989;**56**:288–302.

Neander K. Functions as Selected Effects: The Conceptual Analyst's Defense. *Philos Sci* 1991;**58**:168–84.

Nelson RH. A Critique of the Neurodiversity View. *J Appl Philos* 2020:japp.12470.

Nesse RM. Maladaptation and natural selection. *Q Rev Biol* 2005;**80**:62–70.

Nour MM, Liu Y, Dolan RJ. Functional neuroimaging in psychiatry and the case for failing better. *Neuron* 2022;**110**:2524–44.

Phillips MM. Neurologist Walter Freeman's Lobotomy Legacy - WSJ.com. *Wall Str J* 2013.

Rogers A. Star Neuroscientist Tom Insel Leaves the Google-Spawned Verily for ... a Startup? | WIRED. *Wired* 2017.

Rogers WA, Walker MJ. The Line-drawing Problem in Disease Definition. *J Med Philos* 2017;**42**:405–23.

Schroder HS, Devendorf A, Zikmund-Fisher BJ. Framing depression as a functional signal, not a disease: Rationale and initial randomized controlled trial. *Soc Sci Med* 2023;**328**:115995.

Shorter E. The history of nosology and the rise of the Diagnostic and Statistical Manual of Mental Disorders. *Dialogues Clin Neurosci* 2015;**17**:59–67.

Smith SE. Is Evolutionary Psychology Possible? *Biol Theory* 2019;**1**:3.

Sober E. *Philosophy of Biology*. Routledge: Routledge, 2000.

Sonleitner FJ. What Did Karl Popper Really Say About Evolution? *Creat Evol J* 1986;**6**:9–14.

Wakefield JC. The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values. *Am Psychol* 1992;**47**:373–88.

Wakefield JC. When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorder. *Dev Psychopathol* 1997;**9**:269–90.

Wakefield JC. Biological Function and Dysfunction: Conceptual Foundations of Evolutionary Psychopathology. *The Handbook of Evolutionary Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2015, 1–19.

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