BEHAVIORAL AND BRAIN SCIENCES (2006) **29**, 385–452
Printed in the United States of America

Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best?

Matthew C. Keller
Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA 23219.
matthew.c.keller@gmail.com **www.matthewckeller.com**

Geoffrey Miller
Department of Psychology, University of New Mexico, Albuquerque, NM 87131-1161.
gfmiller@unm.edu **www.unm.edu/~psych/faculty/gmiller.html**

Abstract: Given that natural selection is so powerful at optimizing complex adaptations, why does it seem unable to eliminate genes (susceptibility alleles) that predispose to common, harmful, heritable mental disorders, such as schizophrenia or bipolar disorder? We assess three leading explanations for this apparent paradox from evolutionary genetic theory: (1) ancestral neutrality (susceptibility alleles were not harmful among ancestors), (2) balancing selection (susceptibility alleles sometimes increased fitness), and (3) polygenic mutation-selection balance (mental disorders reflect the inevitable mutational load on the thousands of genes underlying human behavior). The first two explanations are commonly assumed in psychiatric genetics and Darwinian psychiatry, while mutation-selection has often been discounted. All three models can explain persistent genetic variance in some traits under some conditions, but the first two have serious problems in explaining human mental disorders. Ancestral neutrality fails to explain low mental disorder frequencies and requires implausibly small selection coefficients against mental disorders given the data on the reproductive costs and impairment of mental disorders. Balancing selection (including spatio-temporal variation in selection, heterozygote advantage, antagonistic pleiotropy, and frequency-dependent selection) tends to favor environmentally contingent adaptations (which would show no heritability) or high-frequency alleles (which psychiatric genetics would have already found). Only polygenic mutation-selection balance seems consistent with the data on mental disorder prevalence rates, fitness costs, the likely rarity of susceptibility alleles, and the increased risks of mental disorders with brain trauma, inbreeding, and paternal age. This evolutionary genetic framework for mental disorders has wide-ranging implications for psychology, psychiatry, behavior genetics, molecular genetics, and evolutionary approaches to studying human behavior.

Keywords: adaptation; behavior genetics; Darwinian psychiatry; evolution; evolutionary genetics; evolutionary psychology; mental disorders; mutation-selection balance; psychiatric genetics; quantitative trait loci (QTL)

1. Introduction

Mental disorders such as schizophrenia, depression, phobias, obsessive-compulsive disorder, and mental retardation are surprisingly prevalent and disabling. In industrialized countries such as the United States, an estimated 4% of people have a severe mental disorder (National Institute of Mental Health 1998), and almost half of people will meet the criteria for some type of less severe mental disorder at some point in their lives (Kessler et al. 2005). The annual economic costs in treatment and lost productivity are in the hundreds of billions of dollars (Rice et al. 1992). The less quantifiable personal costs of mental disorders to sufferers, families, and friends are even more distressing. For example, schizophrenia affects about 1% of people worldwide (Jablensky et al. 1992), typically beginning in early adulthood and often following a chronic lifelong course. People with schizophrenia often imagine hostile, confusing voices; they have trouble thinking clearly, feeling normal emotions, or communicating effectively; and they tend to lose jobs, friendships, and sexual partners. In response, many people with schizophrenia kill themselves, and a much larger proportion dies childless.

This is an evolutionary puzzle, because differences in the risk of developing schizophrenia and other common, debilitating mental disorders are due, in large part, to differences in people's genes. Given that natural selection has built the most exquisitely complex machinery known to humankind – millions of species of organic life-forms – why do so many people suffer from such debilitating and heritable mental disorders? If these mental disorders are as disabling as they appear, natural selection should have eliminated the genetic variants (*susceptibility alleles*) that predispose to them long ago. Does the prevalence of heritable mental disorders therefore imply that mental

 0140-525X/06 $12.50

disorder susceptibility alleles were selectively neutral or perhaps even advantageous in the ancestral past, or has natural selection been unable to remove susceptibility alleles for some hidden reason?

1.1. The goal of this article and who should read it

This article tries to develop an understanding of the evolutionary persistence of susceptibility alleles underlying common, heritable, harmful mental disorders. We compare and contrast the three broadest classes of evolutionary genetic models that explain persistent genetic variation: ancestral neutrality, balancing selection, and polygenic mutation-selection balance. Such models have been tested mostly by evolutionary geneticists on traits such as bristle numbers in fruit flies, survival in nematode worms, and growth rates in baker's yeast. Yet these models make strong, discriminating predictions about the genetics, phenotypic patterns, and fitness payoffs of any trait in any species, and so should be equally relevant to explaining mental disorder susceptibility alleles. However, these three main models of persistent genetic variation have never before been directly compared with regard to their theoretical and empirical adequacy for explaining human mental disorders. That is our first main goal.

Our second main goal is to promote more consilience among evolutionary genetics, human behavioral/psychiatric genetics, and Darwinian psychiatry/evolutionary psychology. Trying to integrate these disparate fields is hard, not just because each field has different goals, terms, assumptions, methods, and journals, but also because each field has various outdated misunderstandings of one another. For example, we will argue that Darwinian psychiatry often relies too heavily on balancing selection, whereas psychiatric genetics often assumes fitness neutrality or ignores evolutionary forces altogether. Although balancing selection and neutral evolution were historically seen as primary causes of genetic variation, they have proven less important than expected in explaining persistent genetic variation in traits related to fitness. Conversely, the third model – polygenic mutation-selection balance – has enjoyed a theoretical and empirical renaissance in evolutionary genetics, but remains obscure and misunderstood in psychiatric genetics and Darwinian psychiatry.

Cross-fertilization between these fields promises not only to shed light on deep quandaries regarding the origins of mental disorders; it also may help resolve some ongoing frustrations within each field by guiding research and theory more effectively. Evolutionarily oriented mental health researchers, such as Darwinian psychiatrists and evolutionary psychologists, often go to torturous lengths to find hidden adaptive benefits that could explain the evolutionary persistence of profoundly harmful mental disorders such as schizophrenia or anorexia, but these accounts are often frustratingly implausible or hard to test. New ideas from evolutionary genetics and data from psychiatric genetics can help this audience better understand which evolutionary genetic models are theoretically credible and empirically relevant to mental disorders.

Many psychiatric and behavioral geneticists try to find the specific susceptibility alleles that underlie common, harmful, heritable mental disorders. They are often frustrated that even the most promising loci explain little overall population risk and rarely replicate across studies or populations. Traditional methods for gene hunting implicitly assume that mental disorder susceptibility alleles will be at relatively high frequencies and common across populations. Such a convenient scenario, we will argue, could arise from ancestral neutrality or balancing selection, but is much less likely to arise from a mutation-selection balance. Evolutionary genetics could help guide more fruitful gene hunting based on more realistic assumptions.

Evolutionary geneticists try to understand the origins and implications of natural genetic variation across traits and species. The beautiful empirical and theoretical work in evolutionary genetics is under-funded and too often thought irrelevant to human welfare. Greater familiarity with evolutionary genetics might help funding agencies appreciate the potential relevance of this work to understanding some of the leading causes of human suffering, and may introduce evolutionary geneticists to rich genetic data sets on complex human traits such as mental disorders that can be used to test evolutionary models.

1.2. What this article owes to Darwinian psychiatry

In developing our ideas, we build upon Darwinian psychiatry as it has developed over the last 20 years (McGuire & Troisi 1998; Nesse & Williams 1994; Stevens & Price 2000a). Our starting point is the Darwinian psychiatric view that dysfunction is difficult to infer without an understanding of function (Troisi & McGuire 2002; Wakefield 1992). Mental disorders, by this viewpoint, reflect a failure of one or more psychological adaptations to perform their proper, naturally selected, prehistoric functions (Troisi & McGuire 2002; Wakefield 1992). The heart is an adaptation designed to pump blood, for example, and its failure causes blood-circulation

MATTHEW C. KELLER is a postdoctoral fellow at the Virginia Institute for Psychiatric and Behavioral Genetics. He received a B.A. from the University of Texas, Austin in 1995 and a Ph.D. from the University of Michigan, Ann Arbor in 2004. He has done postdoctoral work in genetic epidemiology at the Queensland Institute of Medical Research in Brisbane, Australia, and at the Center for Society and Genetics at UCLA. His primary interests are in behavioral/psychiatric genetics, evolutionary psychology, evolutionary theory, personality, emotion, and statistical methods.

GEOFFREY MILLER is an evolutionary psychologist at the University of New Mexico. He received a B.A. from Columbia University in 1987, and a Ph.D. from Stanford University in 1993, then worked in Europe until 2001 (at University of Sussex, University College London, London School of Economics, and the Max Planck Institute for Psychological Research in Munich). His book *The Mating Mind* (2000) has been published in 11 languages. His research concerns human mate choice, fitness indicators, evolutionary behavior genetics, intelligence, personality, psychopathology, and consumer behavior.

problems that are functionally distinguishable from a pancreas's failure to regulate blood sugar or a lung's failure to oxygenate blood. Likewise, there is a clear mental health problem when a brain is unable to feel social emotions or make sense of reality. This perspective has some important corollaries.

First, a better understanding of normal psychological adaptations should help delineate harmful dysfunctions in those adaptations. Research on adaptive function (e.g., evolutionary psychology; Barkow et al. 1992; Buss 1995) and research on maladaptive dysfunction (e.g., Darwinian psychiatry) are mutually illuminating. This is equally true when mental disorder symptoms have only indirect relationships to psychological adaptations. For example, reading disorders cannot result from a dysfunction in a "reading adaptation," because the visual and linguistic adaptations that enable reading evolved long before the invention of writing a few thousand years ago (Wakefield 1999a). Likewise, auditory hallucinations in schizophrenia probably do not result directly from dysfunction in a "hallucination-suppression adaptation," but indirectly, as side-effects of dysfunctions in more plausible mechanisms that, for example, coordinate and store short-term information, or that filter irrelevant stimuli (Cannon & Keller 2005).

Second, many mental disorders are probably extreme points along a continuum of symptom severity that ranges from patently *unaffected* to extreme forms of the disorder. This makes distinctions between "normal" and "abnormal" somewhat arbitrary (Farmer et al. 2002), because psychological adaptations often show continuous degradation of performance. In this dimensional view of mental disorders, schizophrenia is an extreme form of schizotypal and schizoaffective personality disorders, mental retardation is an extreme form of low intelligence, chronic depression is an extreme form of normal depressive reactions, and so forth. Even mental disorders that look like discrete categories at the phenotypic macro-level (mainly eating, dissociative, post-traumatic stress, melancholic depressive, and antisocial disorders; Haslam 2003) may be influenced by the cumulative effect of many minor dysfunctions at the micro-level of genes and brain development (Gottesman & Shields 1967).

Third, some apparently pathological behaviors may not really be disorders at all from an evolutionary perspective because they do not reflect genuine maladaptive dysfunctions. In particular, some clinically defined mental disorders such as certain phobias or depressions may be *reactive defenses* analogous to fever, nausea, and bodily pain, which protect against infections, toxins, and tissue damage, respectively (Gilbert 1998; McGuire & Troisi 1998; Nesse & Williams 1994). Aversive defenses are cues that something in the environment is wrong, not pathologies themselves.

Consider, for example, depression in light of the reactive defense model. In response to major failures or losses, normally expressed depressive symptoms (e.g., pessimism and fatigue) may adaptively withdraw effort from unpropitious situations when the marginal fitness returns are likely to be low, and emotional pain may motivate avoidance of such situations in the future (Keller & Nesse 2005; 2006; Nesse 2000). These normal reactions are illustrated by the regression line in Figure 1; more severe situations provoke more protracted and severe

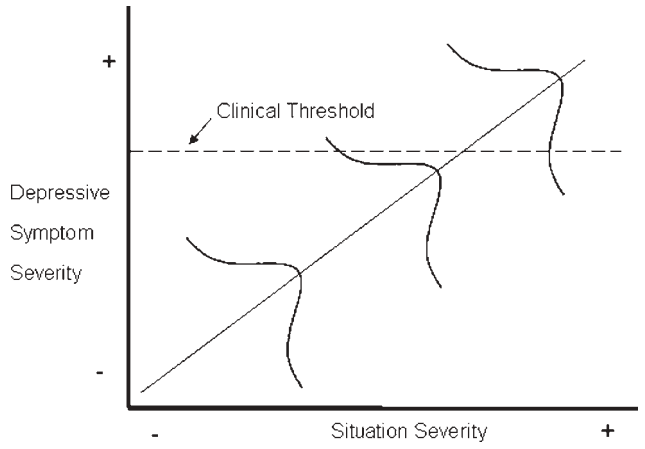

Figure 1. The reactive defense model as applied to depressive reactions (see T.A. text).

reactions. The Gaussian distributions in Figure 1 illustrate interpersonal differences, including genetic differences, which influence symptom severity, given a certain level of situation severity. Since severe situations (e.g., major failures, death of kin) can cause nearly anyone to experience depressive symptoms (Monroe & Simons 1991), some cases of severe and prolonged depressive symptoms (i.e., clinical depression; see above the dashed line in Fig. 1) may simply be normal and adaptive responses to very adverse situations. At the same time, depressive symptoms that are abnormally severe, given the situation (the positive extremes of the Gaussian distributions), may signify malfunctions in the mechanisms responsible for depressive symptoms. Thus, clinical cut-offs based solely on symptom severity and duration, and which do not consider the fitness-relevant precipitating situation, may fail to distinguish truly pathological from non-pathological depressive symptoms (Wakefield 1999a).

These insights are generally appreciated in Darwinian psychiatry and eventually should help build a comprehensive theoretical framework for psychiatry. However, there remains a gaping hole in Darwinian psychiatry's account of mental disorders: there are no good explanations of why human brains seem to malfunction so often, and why these malfunctions are both heritable and disastrous to survival and reproduction. That is, there is still no good answer for why such susceptibility alleles have persisted despite thousands of generations of natural selection for adaptive human behavior.

1.3. What phenomena this article tries to explain

This article tries to develop an understanding of the evolutionary persistence of susceptibility alleles: regions of DNA – broadly defined to include both coding as well as non-coding, regulatory regions (see sect. 6.3) – that differ between individuals in the population and that increase the risk of common mental disorders. In other words, this article is concerned with explaining the genetic rather than the environmental variation associated with mental disorders (with complications such as gene-environment interactions considered later [sect. 4.4]). The reactive defense model offers insight into the environmental triggers for certain disorders – the normal

reactions to environmental stressors and the suites of species-typical, fixed alleles that code for these reactions. However, the reactive defense model is not helpful in explaining genetic variation, because adaptive defenses should be activated by environmental triggers, not heritable risk alleles that differ between individuals.

Although we have continually referred to mental disorders such as schizophrenia, depression, or mental retardation as "common," these disorders are *uncommon* in an absolute sense, generally having lifetime prevalence rates of less than 2%. Rather, they *are* common relative to the thousands of other heritable states that are known to be harmful to fitness, such as achondroplastic dwarfism or Apert's syndrome.

Most rare, harmful, single-gene disorders (*Mendelian disorders*) have frequencies consistent with *mutation-selection balance* – a balance between genetic copying errors that turn normal alleles into harmful mutations, and selection eliminating these mutations (Falconer & Mackay 1996). Mutations arise in parental germ-line cells and are passed on to offspring (and all their cells, including their own germ-line cells) at some low rate (m) per gene, per individual, per generation. Those that affect the phenotype are almost always harmful for the same reason that random changes to a computer's circuitry are almost always harmful: entropy erodes functional complexity (Ridley 2000). Selection removes these mutations at a rate proportional to the fitness cost of the mutation, represented by the selection coefficient (s) against the mutation. If s is reproductively lethal ($s = 1$), the newly arisen mutation exists in only one body before being eliminated from the population, but if s is fairly small, the mutation may pass through and affect many bodies through many generations before being removed by selection. The result of this balance between mutation rate m and selection coefficient s is usually a low equilibrium frequency (p) of mutant alleles that have not yet been removed from the population by selection. Specifically, mutations are expected to have population frequencies of $p = m/s$ if dominant, $p = \sqrt{m/s}$ if recessive, and somewhere in between otherwise (*additive* alleles are exactly midway between). As mutation rate m decreases or selection coefficient s increases, the mutation's frequency p should drop. This process accurately describes, in most cases, why Mendelian disorders are so rare.

The *cumulative* frequency of all Mendelian disorders – around 2% of all births (Sankaranarayanan 2001) – is high only because so many genes are subject to mutation (around 25,000). Heritable harmful disorders that are individually this rare (<1/5,000) pose no evolutionary paradox; no one wrings their hands about trying to find hidden adaptive benefits for such disorders because their frequencies are consistent with a simple balance between mutation and selection. Thus, a more accurate way to classify mental disorders as "common" or "rare" is to assess whether they are much more common than would be expected from a single-gene mutation-selection balance.

Table 1 compares the frequencies of several mental disorders with the frequencies of several Mendelian disorders, all of which are consistent with mutation-selection expectations (except for sickle-cell anemia, discussed in sect. 5.4). Stunningly, common mental disorders tend to be hundreds and even thousands of times more prevalent than expected from a single-gene mutation-selection model. This discrepancy has led many researchers (e.g., D. R. Wilson 1998) to dismiss mutation-selection

Table 1. *Comparisons of frequencies between a small subset of Mendelian disorders and common mental disorders*

Disorder	Genetic basis	Lifetime prevalence per 100,000 in U.S.A.
Mendelian disorders		
Dyskeratosis congenita	Recessive mutations at 3q25	<1
Granulomatous disease, type I	Recessive mutations at 7q11.23	<1
Apert's syndrome	Dominant mutations at 10q26	<1
Juvenile onset Parkinson's	Recessive mutations at 1p & 6q26	<1
Achondroplastic dwarfism	Dominant mutations at 4q	2–3
Sickle-cell anemia	Recessive mutation at 11p15.5	1,000[a]
Common mental disorders		
Autism	Unknown; $h^2 \cong .90$	20–50
Tourette's syndrome	Unknown; $h^2 \cong .90$	50
Anorexia nervosa	Unknown; $h^2 \cong .65$	100
Bipolar disorder	Unknown; $h^2 \cong .60$	800
Schizophrenia	Unknown; $h^2 \cong .80$	1,000
Mild mental retardation[b]	Unknown; $h^2 > .65$	2,000
Obsessive-compulsive disorder	Unknown; $h^2 \cong .45$	2,000
Panic disorders	Unknown; $h^2 \cong .30$	1,700–3,500
Depression	Unknown; $h^2 \cong .45$	5,000–17,000

Note: Data obtained from *Online Mendelian Inheritance of Man* (n.d.) for Mendelian disorders and from the National Institute of Mental Health (1998) for common mental disorders unless otherwise noted. When single or best estimates of heritability or prevalence were unavailable, we used the average of the reported estimates.
[a]Among African Americans.
[b]Heritability and prevalence data derived from Vogel and Motulsky (1997).

balance as a viable explanation for certain mental disorders, and to doubt that mental disorder susceptibility alleles were ancestrally maladaptive. However, such a conclusion is unwarranted. While single-gene mutation-selection models can clearly be eliminated as explanations for the mental disorders listed in Table 1, multiple-gene (polygenic) models (e.g., Shaner et al. 2004) cannot.

This article focuses on the susceptibility alleles of mental disorders that are much more common than would be expected from a single-gene mutation-selection balance; roughly, this corresponds to mental disorders with lifetime prevalence rates above 50 per 100,000 in reproductively aged adults. The best-studied of such disorders are listed in Table 1, but we do not attempt to provide an exhaustive list of precisely what mental disorders this entails, in part because we suspect that the sundry categories of modern mental disorders are not very meaningful biologically (see sects. 6 and 8), but also because our focus is on understanding the persistence of susceptibility alleles in general rather than on understanding mental disorders individually. Nevertheless, the types of common mental disorders that pose the largest paradox are those that are the most harmful (anorexia, bipolar disorder, schizophrenia, mental retardation, and obsessive-compulsive disorder). When we refer to *mental disorders*, these are the types of disorders we have in mind. If we can explain the susceptibility alleles for disorders that are this debilitating, then the same explanations should provide insight into susceptibility alleles for somewhat less debilitating disorders (panic disorders and depression). The following section examines the central paradox of susceptibility alleles in more detail.

2. The paradox of common, harmful, heritable mental disorders

The complexity, optimality, and diversity of life on Earth reveal the awesome power of natural selection. Common, harmful, and heritable mental disorders (as well as other disorders that are not the focus of the current article) seem to be glaring exceptions. They pose an evolutionary paradox because natural selection is expected to make harmful, heritable traits very uncommon very quickly. Over evolutionary time, selection favors higher-fitness alleles; alleles at most genetic loci have gone to *fixation* (virtually 100% prevalence) because they promoted survival and reproduction under ancestral conditions better than other alleles did on average. Such alleles comprise the species-typical human genome; its normal neurodevelopmental product is human nature. Lower-fitness alleles, on the other hand, even those with very minor negative effects, tend to go extinct fairly quickly. Alleles that reach fixation or extinction cause no genetic variation, and so cannot contribute to heritable variation in traits, such as mental disorders. This expectation that selection should minimize genetic variation in fitness-related traits was canonized in evolutionary theory as a major implication of Fisher's *fundamental theorem of natural selection* (Fisher 1930/1999). For decades, biologists expected that the stronger the selection on a trait, the less heritable variation the trait should show, and early empirical data seemed supportive.

Based on such reasoning, evolutionary psychologists have usually argued that genetic variation in human psychological traits is likely to be either adaptively neutral (e.g., Tooby & Cosmides 1990) or adaptively maintained by balancing selection (e.g., Mealey 1995). Both explanations require that the alternative alleles underlying a trait's heritable variation have net fitness effects that are exactly equal to each other, when averaged across evolutionary time and ancestral environments. These explanations seem less relevant to mental disorders, which appear to be the very embodiment of maladaptive traits. Nevertheless, the expectation that selection knows best, and that genetic variation in any common trait cannot be maladaptive, led to something of a cottage industry among Darwinian psychiatrists trying to explain the evolutionary persistence of alleles that increase the risk of such mental disorders as schizophrenia (Horrobin 2002; Huxley et al. 1964; Jarvik & Deckard 1977; Polimeni & Reiss 2002; Stevens & Price 2000a), bipolar disorder (Sherman 2001; D. R. Wilson 1998), depression (D. R. Wilson 2001), and anorexia (Guisinger 2003). In response, clinicians more familiar with psychiatric hospitals, prisons, and detox centers were understandably skeptical that such apparently Panglossian evolutionary ideas could explain real mental illness (e.g., Brüne 2004; McCrone 2003).

Can an evolutionary account of mental disorder susceptibility alleles be reconciled with the clinical view of mental disorders as genuine dysfunctions? Because they reveal interesting misunderstandings of the problem, we begin by considering the most commonly invoked *nonviable* possible evolutionary explanations of mental disorder susceptibility alleles. We next consider the (chiefly theoretical) merits of three explanations – ancestral neutrality, balancing selection, and polygenic mutation-selection balance – that are better grounded in modern evolutionary genetics. We then discuss six pieces of empirical evidence, concerning the relationships between mental disorders and fertility, brain trauma, paternal age, inbreeding, comorbidity, and frequencies and effect sizes of mental disorder susceptibility alleles that help distinguish between these explanations. We conclude with implications for future research.

3. Non-resolutions to the paradox of common, harmful, heritable mental disorders

3.1. Mental disorders are not really heritable

After decades of consistent behavioral genetic research, the hypothesis that genes play no role in mental disorders (e.g., Ross & Pam 1995) is simply no longer tenable. Using different methodologies, behavioral geneticists have consistently found that mental disorder heritability estimates range from about .2 to about .8, meaning that 20% to 80% of the differences between individuals in mental disorder liability are accounted for by differences in alleles between people. Without acknowledging genetic influences on mental disorders, only the most convoluted, post hoc arguments could explain why (a) adopted children are consistently more similar to their biological than to their adoptive parents, (b) siblings and twins reared apart are about as similar as siblings and twins reared together, (c) similarity in extended families

decreases monotonically as a function of genetic similarity, and (d) identical twins are consistently more similar than fraternal twins (Bouchard et al. 1990; Plomin et al. 2001).

Three issues regarding mental disorder heritability estimates do merit clarification, however. First, heritability describes how much genetic or environmental factors play a role in causing *differences* in a trait; it tells us nothing about the causes of *similarities* in a trait. Both environmental and genetic factors are 100% necessary for the species-typical expression of every trait, including every mental disorder. While true, this fact does not provide an answer to why alleles that create differences in mental disorder risk persist. Second, finding positive heritability for a mental disorder does not vindicate the mental disorder as a diagnostic category. To a first approximation, every reliably measured behavioral trait shows positive heritability – even constructs such as television viewing (Plomin et al. 1990) and political attitudes (Eaves et al. 1999). Any arbitrary "disorder" composed of unrelated but heritable symptoms will show credible heritability.

Last, heritability is a statistical construct that averages over a lot of complexity. The causal pathways between genes and the heritable behaviors they influence must be mediated by many factors, both genetic and environmental in nature. If these factors differ across populations, cohorts, or environmental conditions, then heritability estimates – and even the specific genes responsible for the heritability – might also differ across populations, cohorts, or environmental conditions. For example, if body size is associated with successful aggression in one particular society, then genes that normally influence size will also influence aggression in that society (this concept is sometimes called *reactive heritability*; Tooby & Cosmides 1990). Thus, in some cases, contemporary heritability may not accurately reflect ancestral heritability in magnitude or in composition – a point we consider in more depth later (sects. 4.2 and 4.4) when discussing gene-by-environment interactions.

3.2. Mental disorders are not common enough to hurt the species

One might argue that the cumulative frequency of severe mental disorders, around 4%, is not high enough to imperil the survival of the human species. Alternately, one might argue that the genetic variation underlying mental disorders persists because it is the essential raw material for future evolutionary progress (Embry 2002). These points ignore the central lesson of evolutionary genetics: selection acts on competing alleles within a species, without regard to long-term species viability or evolvability (Williams 1966). Natural selection is a purely mechanistic and iterative process whereby alleles from one generation have a non-random probability of being represented in subsequent generations. Natural selection does not – indeed cannot – hedge bets by stockpiling genetic variation in the hope that currently maladaptive alleles might become adaptive in the future.

3.3. Mental disorders are not really harmful to individual fitness

It is sometimes argued that mental disorders were not fitness reducing in ancestral environments because humans reproduced earlier than they do today (e.g., Hardcastle 2004; Weisfeld 2004). However, every mental disorder in Table 1 strikes well before ancestral humans would have finished reproducing. A harmful mental disorder that struck even as late as the forties would have led to a small but evolutionarily significant decrement in number of future offspring (e.g., see the fertility function of hunter-gatherers in Daly & Wilson [1983]), even apart from its negative effect on inclusive fitness through reduced ability to aid relatives (Kaplan et al. 2000). Thus, if mental disorders were debilitating in ancestral conditions, their developmental timing would have harmed fitness given any reasonable model of ancestral life-history profiles.

Another version of the not-really-harmful argument concerns the fitness effects of susceptibility alleles rather than mental disorders per se: mental disorders may be harmful to fitness, but their genetic architecture may be so complex that natural selection has been unable to eliminate the alleles that predispose to them. Used in this sense, "genetic complexity" basically means *nonadditive genetic variation*: variation in fitness effects that depend on particular combinations of alleles, and that selection therefore affects at a much slower rate (Merilä & Sheldon 1999). Such nonadditive effects include *dominance* (interactions between two alleles at the same locus) and *epistasis* (interactions between alleles at different loci). However, for the same reasons that main effects almost always exist in addition to interaction effects in statistical analyses, dominant and epistatic alleles almost always have some average, or additive, phenotypic effects (contributing to *additive genetic variation*) that are more visible to selection (Falconer & Mackay 1996; Mather 1974). Available empirical evidence on mental disorders is consistent with this expectation. Although the vast majority of behavioral genetic studies have used a design (the classical twin design) that cannot simultaneously estimate additive, nonadditive, and shared-environment effects (Eaves et al. 1978; Keller & Coventry 2005), behavioral genetic studies using designs better able to distinguish these (such as the extended twin design; reviewed in Coventry & Keller 2005) have found at least some additive genetic variation for those mental disorders investigated to date: depressive symptoms, panic disorders, and neuroticism (a correlate of many mental disorders). Thus, the harm that mental disorders do is almost certainly visible to natural selection to some degree.

It could also be argued that mental disorders simply have not affected survival and reproduction, and so are not under selection. At least in *modern* environments, however, many mental disorders are associated with markedly lower fertility (summarized in Table 2). These mental disorders seem to undermine fertility not so much through reducing survival, but through reducing attractiveness or ability in the mating arena. Of the studies that examined this issue, reductions in fertility were principally the result of lower marriage rates rather than fewer offspring once married. At this level of socio-sexual competition to attract and retain mates, there may not be so much difference between the fitness effects of mental disorders in prehistoric and contemporary societies (Miller 2000a).

However, modern fertility has an unknown relationship to ancestral fertility (Symons 1989), which is more relevant

Table 2. *Available fertility estimates (1960–2005) of common mental disorders*

Disorder	Fertility[a]	Birth cohorts, location	Sample	Reference
Psychotic disorders				
Schizophrenia	58% ♀	1890–1919, U.S.	4,041 inpatients & outpatients	Erlenmeyer-Kimling et al. 1969
Schizophrenia	36%	1890s–1950s, Germany	306 inpatients	Vogel 1979
Schizophrenia	45% ♀	1890s–1940s, U.K.	1,086 inpatients & outpatients	Slater et al. 1971
Schizophrenia	70% ♀	1911–1940, U.S.A.	4,023 inpatients & outpatients	Erlenmeyer-Kimling et al. 1969
Schizophrenia	40% ♂; 57% ♀	1914–1968, Spain	142 inpatients & outpatients	Fananás & Bertranpetit 1995
Psychosis[b]	29% ♂; 83% ♀	1920s–1970s, Australia	282 primary-care patients	McGrath et al. 1999
Schizophrenia	23% ♂; 51% ♀	1921–1976, Canada	36 primary-care patients	Bassett et al. 1996
Schizophrenia	101% ♀	1932–1951, U.S.A.	223 outpatients	Burr et al. 1979
Schizophrenia	29% ♂; 62% ♀	1930s–1970s, Japan	553 outpatients	Nanko & Moridaira 1993
Schizophrenia	25%	1930s–1970s, Ireland	285 from population register	Kendler et al. 1993
Schizophrenia	27% ♂; 45% ♀	1950s, Finland	11,231 from population register	Haukka et al. 2003
Psychosis[b]	46% ♀	1953–1982, U.K.	4,556 primary-care patients	Howard et al. 2002
Schizophrenia	23% ♂; 12% ♀	20th century, Denmark	27 from adoption database	Rimmer & Jacobsen 1976
Schizophrenia	37%	20th century, Palau	70 unknown	Sullivan & Allen 2004
Mood disorders				
Affective disorder[c]	70%	1890s–1950s, Germany	165 inpatients	Vogel 1979
Bipolar disorder	50% ♂; 62% ♀	1890s–1950s, U.S.A.	134 inpatients	Baron et al. 1982
Bipolar disorder	69% ♀	1890s–1940s, U.K.	2,692 inpatients & outpatients	Slater et al. 1971
Affective disorder[d]	47% ♂; 89% ♀	1920s–1970s, Australia	60 primary-care patients	McGrath et al. 1999
Affective disorder[c]	66% ♀	1953–1982, U.K.	1,705 primary-care patients	Howard et al. 2002
Developmental disorders				
Mental retardation[e]	40% ♂; 72% ♀	1870s–1930s, Minnesota	1,450 descendants of inpatients	Reed 1971
Low intelligence[f]	88%	1870s, Michigan	78 from school register	Bajema 1963
Mental detardation[e]	95%	1870s–1930s, Minnesota	1,300 descendants of inpatients	Waller 1971
Organic disorders[g]	53%	1890s–1950s, Germany	275 inpatients	Vogel 1979
Other disorders				
OCD[h]	47% ♀	1890s–1940s, U.K.	235 inpatients & outpatients	Slater et al. 1971
"Neurosis"[i]	64% ♀	1890s–1940s, U.K.	5,596 inpatients & outpatients	Slater et al. 1971
Mixed[j]	53%	1890s–1950s, Germany	316 inpatients	Vogel 1979

Note: Data include all available studies in 1960–2005 in which overall fertility rates were reported or derivable and in which a suitable comparison group was reported.
[a]Number of offspring as a proportion of number of offspring among general population matched on age, gender, and other pertinent demographic variables.
[b]Schizophrenia, schizoaffective disorder, schizophreniform, delusional disorder, and paranoid psychosis.
[c]Major depression and bipolar disorder.
[d]Bipolar disorder, bipolar disorder with psychosis, mania, mania with psychosis, and depression with psychosis.
[e]IQ < 70.
[f]IQ < 85.
[g]Mental retardation and psychoses caused by trauma.
[h]Obsessive-compulsive disorder.
[i]Usage not described.
[j]Panic disorder, obsessive-compulsive disorder, drug and alcohol dependence, sexual deviance, and personality disorders.

to understanding the evolutionary persistence of susceptibility alleles. Additional and perhaps more persuasive evidence that mental disorders were associated with decreased ancestral fitness is simply based upon the ubiquitous evidence of their deviance and disability in modern societies, irrespective of their effects on fertility (Troisi & McGuire 2002; Wakefield 1992). Any psychiatric book or journal reveals many such examples, which do not need to be enumerated here. If mental disorders existed in ancestral environments in much the same form as they do now, it is reasonable to assume that, *at some level of severity*, they would have resulted in lower ancestral fitness.

Nevertheless, mental disorders may *not* have existed in ancestral environments as they do now. This final version of the not-really-harmful view merits more careful consideration – the idea that, although mental disorders or their susceptibility alleles are harmful under modern conditions, they may not have been harmful under ancestral conditions, when humans lived in small-scale, hunter-gatherer societies. We assess this hypothesis next.

4. Can ancestral neutrality explain common, harmful, heritable mental disorders?

It seems unlikely that mental disorder susceptibility alleles had no effect on ancestral fitness, given that mental disorders are associated with lower fitness (Table 2) and severe impairment in modern environments. Nevertheless, it is possible that mental disorders were associated with more benign symptoms or less ostracism ancestrally so that they were effectively neutral traits. For example, a common speculation is that perhaps prehistoric individuals with schizophrenia were valued shamans, with a special social role as religious visionaries, so perhaps they were not socially and sexually ostracized as in contemporary societies (Polimeni & Reiss 2002; Preti & Miotto 1997). Alternatively, perhaps alleles that increase the risk of mental disorders today had no such effect in ancestral environments. From an extended-phenotype perspective, both cases are examples of gene-by-environment (G–E) interactions, which occur when the effects of alleles differ depending upon the physical or social environment. Is it possible that the fitness effects of mental disorder susceptibility alleles were equal to the fitness effects of non-susceptibility alleles in ancestral environments, enabling them to persist?

4.1. Neutral evolution maintains genetic variation only when combined with recurrent neutral mutation

To assess whether ancestral neutrality is a viable explanation for the persistence of mental disorder susceptibility alleles, we must first understand the conditions under which neutrality maintains genetic variation. The frequencies of neutral alleles are governed by *genetic drift* – random sampling error over evolutionary time. Over the long term, drift leads to genetic uniformity because neutral alleles either fixate or are lost through sampling error. Drift almost never maintains neutral alleles at intermediate frequencies where they could explain heritable variation in mental disorder susceptibility. Drift is stronger in smaller populations, such as ancestral hominid populations, which are more susceptible to sampling error. Without some additional force that either replenishes lost alleles (see the next paragraph) or that counteracts the process of random drift (see the next section), one neutral allele eventually fixates and the alternative alleles go extinct.

Depending on the way that new mutations affect mental disorder risk, recurrent neutral mutations might counteract the loss of genetic variation caused by drift. Mutations can occur anywhere along a locus, the coding region of which is typically about 2,000 base pairs long; like lightning, mutations are very unlikely to hit precisely the same location twice, and thus alleles introduced into the population via recurrent mutation are very unlikely to be the same. If neutral mental disorder susceptibility alleles are specific, in the sense that only one or a few of all the possible mutations at that locus would affect mental disorder risk, while all others would not, then recurrent mutation is too rare an event to replenish lost susceptibility alleles. In this case, random genetic drift would lead to loss or fixation of the mental disorder susceptibility allele. Therefore, models that hypothesize that mental disorders are complex phenotypes, coded by specific alleles that are alternatives to the normal alleles, are not consistent with what is known about the properties of neutral evolution. However, it is probably more biologically plausible that *any* mutation along the locus could increase or decrease mental disorder risk; in this case, random genetic drift plus recurrent mutation could in principle account for substantial genetic variation.

The degree of genetic variation contributed by such a neutral locus, where any mutation affects mental disorder risk, can be quantified. As already noted, only loci that are polymorphic (where more than one allele exists in the population) contribute to genetic variation. Genetic polymorphism can be measured by H, the proportion of heterozygotes at a locus in a population. Kimura (1983) showed that for neutral loci, $H \cong 4N_e\mu/(1 + 4N_e\mu)$, where μ is the probability of a new mutation at the locus per individual per generation, and N_e is the *effective population size* (roughly the harmonic mean of the breeding population size across generations, which tends to be close to the minimum actual population size during genetic bottlenecks). N_e is often estimated to be around 10,000 for humans (Cargill et al. 1999). Assuming μ is around 10^{-6} to 10^{-5} for most loci (Nachman & Crowell 2000), the expected heterozygosity H across neutrally evolving human loci should be around 4% to 29%. Because neutral loci have relatively high average values of H, they can contribute substantially to heritability in human traits and perhaps mental disorders.

To say that neutral evolution *could* maintain the genetic variation underlying mental disorders is very different than saying that such a process is likely. In sections 4.2 and 4.3, we review two reasons that neutral evolution is probably not a general resolution to the paradox of common, heritable, harmful mental disorders, and then we review the types of phenotypes that might be best explained by a weaker version of this process (sect. 4.4).

4.2. Ancestral neutrality must be implausibly precise

For an allele to be truly neutral over the evolutionary long term, the allele must have fitness effects *extremely* close to neutrality within each generation. This statement can be

https://doi.org/10.1017/S0140525X06009095 Published online by Cambridge University Press

quantified simply. For an allele to be neutral (to be governed by genetic drift more than by selection), the selection coefficient s against an allele must be less than $\sim 1/4N_e$. Thus, only if the average fitness of people with an allele is between 99.997% and 100.003% of the fitness of people without this allele (i.e., if $s < 1/40{,}000$) has the frequency of that allele been governed mostly by neutral drift across human evolution. This is an extraordinarily small selection coefficient, equivalent to a difference of just one offspring more or less than average, not in the next generation, but 15 generations into the future, given a roughly constant population size.

Not only must neutral mental disorder susceptibility alleles have been almost exactly neutral in ancestral environments, they must have been *consistently* so. If the alleles were neutral in most but not all environments, or in most but not all cultures, or in most but not all bodies, then averaged across evolutionary time, these alleles would not be neutral. As we have argued, there are strong reasons to believe that the mental disorders listed in Table 2 are fitness-reducing in modern societies. If susceptibility alleles were neutral in ancestral environments but highly dysfunctional today, this implies very large G–E interactions. Yet, very large G–E interactions are implausible, given this consistency requirement that mental disorder susceptibility alleles had to be unfailingly neutral across many different ancestral environments.

Although many evolutionary biologists believe that neutral mutations are the main source of genetic polymorphisms across DNA in general (since most DNA has no phenotypic effect), few now believe that neutral mutations are the main source of phenotypically expressed variation (Ridley 1996). The very fact that neutral alleles have no fitness effects makes them unlikely to affect phenotypic development. By contrast, mental disorder susceptibility alleles do affect the phenotype in modern environments, and it is likely that they would have done so in ancestral environments as well. It is hard to believe that phenotypically expressed alleles associated with conditions that have such harmful effects in modern environments would have been precisely neutral ($s < 1/40{,}000$) across all ancestral environments.

4.3. Ancestral neutrality is hard to reconcile with modern mental disorder prevalence rates

A strictly neutral hypothesis about mental disorder genetic risk factors would dictate that all levels of genetic risk have equal fitness effects. Under such a scenario, any prevalence rate of mental disorders, from 0% to 100%, should be about equally likely. Contrary to this, Table 1 shows that the most harmful mental disorders are consistently rare in an absolute sense, none being more common than about 2%. If neutral evolution were a general answer to the paradox, one would have to explain why the most harmful mental disorders are so consistently rare. The exceptions that prove the general rule are late-onset disorders such dementia (which affects about 30% of people over age 75; Thomas et al. 2001), which are more likely to have been close to selectively neutral under ancestral conditions.

Perhaps the low frequencies of modern mental disorders suggest that they became fitness-reducing only recently and are currently being selected out of human populations (e.g., Burns 2004; T. J. Crow 2000). How plausible is this? As illustrated in Figure 2, alleles with even small fitness effects are quickly driven to extinction. For example, if schizophrenia in Finland has been as disadvantageous over the last 20 generations, as it appears now ($s \sim .50$), and is caused by a single recessive allele with $p = .10$ (explaining the current disease prevalence of 1%), it would follow from standard evolutionary genetics that 42% of Finns were schizophrenic in 1600 – clearly a nonsensical result. Selection on dominant or additive alleles is even faster. Thus, it is not evolutionarily credible to claim that mental disorders are caused by one or even a few genes and have a low but significant prevalence because they became harmful only several thousand years ago.

4.4. Disorders that ancestral (near) neutrality might help explain

We have argued that it is highly unlikely that alleles with substantial fitness-reducing effects today were precisely and consistently neutral across ancestral environments. However, alleles affecting certain disorders might have been much *closer* to being neutral in ancestral environments, and therefore the modern prevalence rates and heritabilities of these disorders may be higher than predicted from modern fitness estimates. This is a plausible hypothesis for heritable disorders that show the hallmarks of G–E interactions: large cross-cultural variation in prevalence rates, increased (or decreased) rates in recent historical time as environments change, and a credible mismatch between ancestral and modern conditions that affects the mental disorder.

Data showing that depression rates vary enormously between cultures, and seem to be rising to very high levels in industrialized nations (Weissman et al. 1996), are consistent with – but by no means prove that – G–E interactions are important in depression. It is also easy to imagine a credible mismatch scenario for depression. For example, social support from kin and friends was probably more available in small-scale ancestral societies

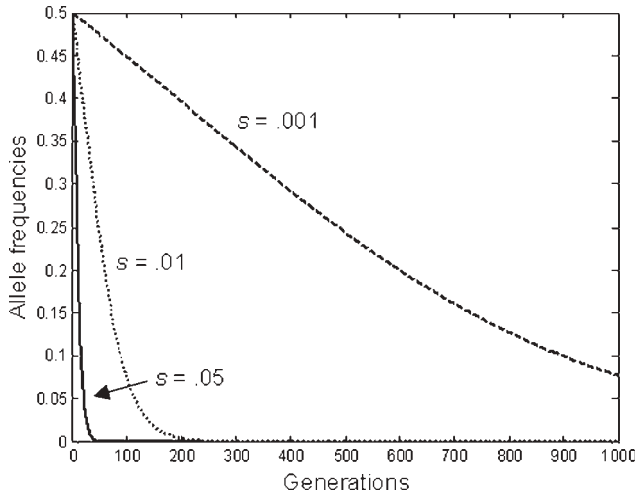

Figure 2. Expected changes (ignoring genetic drift) in allele frequencies across generations, given different levels of selection (s) acting on additive deleterious alleles of minor effect.

https://doi.org/10.1017/S0140525X06009095 Published online by Cambridge University Press

than in modern cities, and such social support may help rescue people from normal periods of transient depression (Kessler 1997). Although heritable shyness may have had little effect on social support in ancestral conditions, "shy alleles" could decrease the social support available when times get tough in modern cities, becoming susceptibility alleles for depression that show strong G–E (or more specifically, in this case, G-culture) interactions.

Other disorders that could plausibly be affected by alleles that were more benign in ancestral environments include: (a) obesity and diabetes, due to unnaturally consistent and appealing food surpluses; (b) asthma, due to unnatural levels and types of antigens and pollutants; and (c) addictions to highly purified, evolutionarily novel drugs, such as heroin or cocaine (Nesse & Berridge 1997). These disorders are heritable within cultures, but their frequencies differ enormously between cultures and environments. They have also probably increased in frequency in cultures most affected within the last 50 to 100 years (Wright & Hastie 2001), as likely environmental risk factors were increasing. It is also reasonable to assume that their environmental risk factors were usually absent in ancestral environments. Finally, in societies with the environmental risk factors, the frequencies of these disorders are not consistently low (e.g., obesity rates are approaching 50% among younger U.S. cohorts).

Nevertheless, it is unlikely that alleles that increase mental disorder risk today were precisely and consistently neutral ancestrally – even those alleles that have become more harmful only recently. Given that natural selection purges even slightly harmful alleles (Fig. 2), the persistence of alleles that were only close to, but not precisely, neutral still requires an explanation. The polygenic mutation-selection paradigm, reviewed in section 6, provides this explanation.

5. Can balancing selection explain common, harmful, heritable mental disorders?

The genetic variation underlying mental disorders, far from being invisible to selection, might have been actually maintained by selection. For example, mental disorders which look harmful and dysfunctional, and which show below-average fitness under some conditions, might show above-average fitness under other conditions. This type of selection, known as *balancing selection*, has been one of the most popular ideas among evolutionary thinkers for resolving the paradox of common, harmful, heritable mental disorders (Allen & Sarich 1988; Barrantes-Vidal 2004; Karlsson 1974; Longley 2001; Mealey 1995; Stevens & Price 2000a), with some researchers even implying that balancing selection is the only possible resolution to the paradox (D. R. Wilson 1998). One purpose of this article is to rebut such claims by showing that there are at least two other potential resolutions to the paradox: neutral evolution and mutation-selection balance.

Balancing selection may be popular among Darwinian psychiatrists in part because it keeps natural selection front and center as the causal force explaining a trait – a comfortable position for adaptationists. Balancing selection might also be appealing for social and moral reasons, because it attributes hidden adaptive benefits to mental disorders in ways that might reduce their social stigma. Morality aside, how feasible is it that balancing selection resolves the paradox?

5.1. Natural selection usually depletes genetic variation

As noted in section 2, selection usually leads to genetic uniformity and therefore depletes heritability. Certain evolutionary models, such as those for phobias and depression (e.g., Keller & Nesse 2005; Watson & Andrews 2002), posit adaptive functions for capacities that are *universal* features of human nature, affected by *universal* suites of genes (i.e., little or no genetic variation), and triggered by adverse situations. These explanations are potentially useful for understanding environmental variation, but do not explain, nor were they intended to explain, the *genetic* variation in phobias and depression.

Other evolutionary models hypothesize that heritable disorders themselves are adaptive without explaining why the disorders have not fixated in the population. Consider three recent hidden-benefit models: Guisinger (2003) viewed symptoms of anorexia as an adaptive response to fleeing famine under ancestral conditions of starvation; Sherman (2001) viewed bipolar disorder as an adaptation to long, severe winters and short summers; and T. J. Crow (2000) viewed schizophrenia as an inevitable risk arising as a side effect of language evolution (see also Burns 2004). None of these offer a compelling explanation for the persistence of heritability in these disorders. If anorexia was simply adaptive under starvation conditions, then the adaptive anorexia alleles would be virtually fixed within those human groups whose ancestors gained such advantages, and the condition should not be heritable within these groups. In truth, however, very few people show these symptoms, and the phenotypic differences between those who do versus those who do not are largely due to genetic differences (Guisinger 2003). Similar arguments can be made for bipolar disorder or schizophrenia. These hidden-benefit models may or may not help explain why humans in general are susceptible to anorexia, bipolar disorder, or schizophrenia, but they do not explain the central paradox addressed in this article: why mental disorder susceptibility alleles have not either fixated, if adaptive, or gone extinct, if maladaptive. This is one of our key points: *Explaining heritable polymorphisms requires special and stringent types of evolutionary explanations that are different from those used to explain species-typical traits*. Most types of selection offer no explanation for mental disorder heritability. Balancing selection can.

5.2. Balancing selection is the only type of selection that actively maintains genetic variation

Balancing selection actively maintains two or more alternative alleles because their net fitness effects balance each other out, being positive in certain genetic or environmental contexts and negative in others. For balancing selection to maintain a stable genetic polymorphism across evolutionary time, (a) the fitness effects of the alternative alleles must be equal across ancestrally relevant genetic and environmental contexts, and (b) some mechanism must assure that these equally fit

alleles are not lost by chance (genetic drift). For the most robust types of balancing selection, if an allele drifts by chance to a lower level, its fitness increases, which then buoys its frequency back up. So long as the equilibrium frequency of one of the alleles is not too low, such a homeostatic mechanism greatly reduces the risk of equally fit alleles being lost by genetic drift.

Before assessing the general utility of balancing selection in explaining mental disorders, we review the explanatory power of four specific forms of it: spatial and temporal variation in selection, heterozygote advantage (also known as heterosis or overdominance), antagonistic pleiotropy, and frequency-dependent selection. Although these are often considered separate evolutionary processes, they have important common features at the evolutionary genetic level that give them similar strengths and weaknesses in explaining mental disorders.

5.3. Temporal or spatial variability in fitness landscapes

Balancing selection can occur when an allele's fitness oscillates over evolutionary time or location. We are aware of no models that try to explain mental disorder heritability by using this mechanism. For this to explain the paradox, a convincing case would need to be made that mental disorders or their susceptibility alleles were advantageous across about half of ancestral populations in different locations or about half of the time, but this seems a priori unlikely, though not disproved, in light of the consistent harmfulness of mental disorders in current environments. A deeper, theoretical problem for this explanation is that no homeostatic mechanism protects alleles against loss through drift; rather, the evolutionary oscillations in an allele's fitness must occur at just the right rate across time or space to keep the allele from fixating or going extinct (Bürger 2000). Such loss of alleles would be especially likely in small prehistoric human populations.

Although this mechanism seems theoretically unlikely to maintain mental disorder susceptibility alleles at *equilibrium*, it is important to remember that we are catching but a snapshot of evolution. It is certainly possible that some susceptibility alleles are at intermediate frequencies because they are sweeping toward fixation or extinction. Such a process may be occurring with a susceptibility allele for heart disease and Alzheimer's disease: APOE*4. APOE*4 is the ancestral allele, being rarest among human groups that have had the longest exposure to agriculture, and is probably headed over the next several thousands of years toward extinction (for two views on why this might be, see Corbo & Scacchi [1999] and Finch & Sapolsky [1999]). Nevertheless, it is unlikely that enough alleles are rising or lowering in frequencies for this to be a general answer to the paradox, given the short time that alleles with fitness effects are at intermediate frequencies (Fig. 2).

5.4. Heterozygote advantage

A genetic polymorphism may be maintained when the heterozygote at some locus has higher fitness than either homozygote (e.g., genotype *Aa* has higher fitness than both *AA* and *aa*). The classic example is sickle-cell anemia. Individuals who are homozygous for the more common allele (*AA*) at the β-hemoglobin locus are susceptible to malaria, whereas those homozygous for the less common allele (*aa*) are more likely to die from sickle-cell anemia. However, heterozygotes (*Aa*) have the best of both worlds: they do not develop anemia, and they are much more likely to survive a malarial infection. In equatorial areas of Africa and Asia where malaria is endemic, heterozygotes have higher fitness than either homozygote. If genotypes rather than genes could be passed to offspring, *Aa* genotypes would have fixated long ago, but this cannot happen. For example, matings between two most-fit heterozygotes nevertheless produce $\frac{1}{4}$ *aa* and $\frac{1}{4}$ *AA* offspring on average. The population frequencies of the two alleles become stable when the average fitness effects of alleles *a* and allele *A* are equal. Here, a homeostatic mechanism keeps alleles from being lost through genetic drift: if the frequency of one allele in the population drifts to a lower level, that allele has an increased chance of finding itself in a heterozygote body, and its average fitness, and hence frequency, increase.

In the case of sickle-cell anemia, Allison (1954) showed that, given the fitness estimates for each genotype at the β-hemoglobin locus, evolutionary genetic theory predicted very well the observed phenotypic frequencies. The sickle-cell story had a large impact on evolutionary biologists in the 1950s, and many suggested that heterozygote advantage might be a general explanation for observed levels of genetic variation in nature (e.g., Lerner 1954). More recently, several evolutionists have theorized that mental disorders such as schizophrenia (Huxley et al. 1964), bipolar disorder (D. R. Wilson 1998), and depression (D. R. Wilson 2001) are maintained by heterozygote advantage.

However, for several reasons, evolutionary biologists have become less enthusiastic about heterozygote advantage as an explanation for persistent heritability in most traits. First, heterozygote advantage appears to be rare in nature: Thirty years of intensive research following the sickle-cell story yielded only six additional examples of polymorphisms maintained in this way (Endler 1986). Second, there are theoretical reasons to doubt that species could sustain widespread maladaptive polymorphisms in this way without going extinct (Crow & Kimura 1970). Third, selection would strongly favor genetic events that overcome the costs of producing homozygotes, such as unequal crossover events that positions both *A* and *a* on the same chromosomal arm, so they can be passed on together without disruption (Ridley 1996), or mutations that reduce the fitness costs of either homozygote. Such genetic events become quite likely across a whole population over evolutionary time, so heterozygote advantage is likely to be an evolutionarily transient stopgap. This is consistent with the fact that the *a* allele at the β-hemoglobin locus evolved fairly recently (Hamblin et al. 2002).

5.5. Antagonistic pleiotropy

Pleiotropy occurs whenever one allele affects more than one trait. Given that traits do not rely on mutually exclusive sets of genes, pleiotropy is ubiquitous. *Antagonistic pleiotropy*, which is also probably ubiquitous, occurs whenever an allele increases the fitness payoffs of one

trait but reduces the fitness payoffs of another trait. For example, an allele might increase fertility but decrease longevity, or increase intelligence but decrease emotional stability.

Generally, this process leads to the fixation of whichever allele has the highest fitness, averaged across the various effects it has on different traits. Even if the net fitness effects of two alternative alleles are precisely equal, which is a priori unlikely, there is no homeostatic mechanism that counteracts the homogenizing effect of genetic drift (Curtisinger et al. 1994; Hedrick 1999; Prout 1999). In fact, this theoretical work suggests that antagonistic pleiotropy is likely to maintain genetic polymorphisms only under a highly restrictive scenario: when individuals with both alleles receive the fitness benefits but not the costs from each allele – a situation called *reversal of dominance*. In this situation, heterozygotes would have the highest fitness, a scenario conceptually equivalent to heterozygote advantage, and which shares the same explanatory weaknesses. The conclusion from theoreticians is that antagonistic pleiotropy cannot maintain genetic variation on its own; it requires a very special type of allelic effect, reversal of dominance, which evolutionary biologists consider unlikely.

Despite these theoretical concerns, antagonistic pleiotropy is probably the most common evolutionary explanation for the persistence of susceptibility alleles. Several researchers have hypothesized that susceptibility alleles underlying bipolar disorder and schizophrenia have two effects: one, to increase creativity, but the second, to increase the risk for the mental disorder (Barrantes-Vidal 2004; Karlsson 1974). These susceptibility alleles are thought to persist because their negative fitness effects from mental disorder risk are precisely offset by their benefits from creativity. The idea that mental disorders are associated with higher creativity is widespread, and supported by some biographical evidence (Jamison 1993) and evidence that relatives of those with mental disorders have higher creativity (reviewed by O'Reilly et al. 2001). However, a literature review of 29 studies found little support for the idea that highly creative people showed an increased rate of mental disorders (Waddell 1998).

5.6. Frequency-dependent selection

Frequency-dependent selection (or more technically, negative frequency-dependent selection) occurs when alleles' fitness effects increase as they become rarer. This process can maintain a stable mix of alleles resulting in persistent trait heritability. Heterozygote advantage can be seen as a special case of this process. Frequency-dependent selection more generally occurs when individuals compete for different resources, such that individuals who are rare relative to their preferred resource are favored (Barton & Keightley 2002).

The classic example of frequency dependence is the evolutionary maintenance of the 50:50 sex ratio (Fisher 1930/1999). If males outnumber females, females necessarily have higher average reproductive success than do males. A mutation increasing the probability of having daughters would be positively selected, and would spread in the population until females began to outnumber males, in which case selection would begin to favor having sons. The evolutionary equilibrium is that both strategies (being male or being female) reach equal frequency, although, in other cases, alternative strategies may have non-equal equilibrium frequencies. Frequency-dependent selection can maintain high levels of heritable genetic variation for as long as the selection pressures remain.

For a few mental disorders such as psychopathy, frequency dependence may be a plausible model. Mealey (1995) argued, forcefully in our opinion, that psychopathy persists at a low base rate as a socially parasitic strategy: it brings high fitness benefits when rare, but becomes less rewarding at higher frequencies because of increased anti-cheater vigilance in the population. Indeed, at the current low base rate (around 1%), male psychopaths seem to have higher-than-average fitness, at least in modern environments – unlike almost all other mental disorders listed in Table 1. In general, frequency-dependent selection can explain polymorphic alleles only when there is a credible explanation of why each allele's fitness increases as its frequency decreases. This is a fairly high standard of evidence. Moreover, there are several problems with balancing selection in general as an explanation for mental disorders, which we explore next.

5.7. General problems with balancing selection explanations for mental disorder susceptibility alleles

Mental disorders are not a random sample of human traits; they are considered "disorders" precisely because they have salient maladaptive outcomes. Rare phenotypes with such severe costs, as opposed to common phenotypes that are not debilitating, are probably the least likely candidates for traits maintained by balancing selection. This is because the devastating negative effects of susceptibility alleles must be balanced by commensurately large, and therefore probably noticeable, positive effects (e.g., sickle-cell anemia being balanced out by malarial resistance). Balancing selection may explain some heritable personality traits such as extroversion and some personality disorders such as psychopathy. Yet it seems a poor candidate as a general explanation of the susceptibility alleles of mental disorders, since their susceptibility alleles would have to show some hidden adaptive benefits that counteract the strongly maladaptive symptoms of these mental disorders.

Another problem for models of both spatio-temporal variation and frequency-dependent selection is that behavioral flexibility, as opposed to fixed, heritable strategies, would probably be favored in the face of differing fitness landscapes (Tooby & Cosmides 1990; although see D. S. Wilson 1994). Fixed, heritable strategies make sense for basic morphological specializations such as growing a male or female body, when it is hard to switch from one to the other after growth. However, the whole point of growing a central nervous system is that different behavioral strategies, which have context-dependent fitness payoffs, can be pursued by the same individual across different situations. Such flexibility circumvents the costs of pursuing fixed strategies when their frequencies are at the wrong level to maximize fitness. Given the extraordinary behavioral flexibility of the human brain, it would be puzzling if such genetically fixed strategies explained mental disorder heritability. Despite these two broad

problems, and the problems specific to the various types of balancing selection, balancing selection cannot be ruled out as a resolution to the paradox on purely theoretical grounds. In section 7, we review several pieces of empirical evidence that support our expectations that balancing selection is not a general explanation for mental disorder susceptibility alleles.

5.8. What balancing selection might explain

Balancing selection might, in theory, maintain mental disorder susceptibility alleles for reasons completely unrelated to mental disorder symptoms. For example, pathogens and parasites are usually poorly adapted to attacking the rarest host genotypes, so rare alleles may help protect the host (Garrigan & Hedrick 2003; Haldane 1949). This anti-pathogen variation could give rise to mental trait variation as a side effect (Tooby & Cosmides 1990), and some researchers have suggested this might explain the high prevalence of schizophrenia susceptibility alleles (J. S. Brown 2003). For this to work, the fitness benefits of improved host defense must outweigh the fitness costs of increased mental disorder risk (Turelli & Barton 2004). Because most loci probably do not affect immunological systems, the vast majority of loci are probably unaffected by parasite-host coevolution. Moreover, selection should have favored minimal overlap between the genes that control anti-pathogen defenses and those that affect other systems, such as the nervous system, although there may be a limit in how far natural selection can go in removing such pleiotropic effects of genes. Although it is certainly possible that some psychological variation is a by-product of frequency-dependent selection for other traits, empirical evidence discussed in section 7 makes it unlikely to be a *general* resolution to the paradox.

We have argued on both theoretical and empirical grounds that antagonistic pleiotropy is unlikely to explain the persistence of mental disorder susceptibility alleles, but a weaker version of it may work better. This version suggests that alleles with conflicting fitness effects on different traits should tend to be *closer* to neutral than alleles without such antagonistic effects, so perhaps they will persist longer at intermediate frequencies and contribute more to heritable variation. If such a near-neutral allele has opposite fitness effects on two traits, those traits should show a negative genetic correlation (Lande 1982). More generally, if antagonistic pleiotropy accounts for substantial genetic variation, most genetic correlations between fitness-related traits should be negative. This logic is compelling, but the evidence among animal traits is not very supportive. A meta-analysis of genetic correlations between fitness-related traits in nonhumans found that 61% were positive (i.e., the higher fitness end of one trait tended to go with higher fitness end of other traits; Roff 1997) – a result less congruent with antagonistic pleiotropy than with polygenic mutation-selection balance models (Charlesworth 1990). Nevertheless, it is plausible that some portion of the genetic variation underlying mental disorders is due to near-neutral alleles that increase mental disorder risk under certain genetic or environmental conditions, but that have some positive benefits in other conditions. Much like the scenario discussed in section 4 for non-pleiotropic near-neutral alleles, this mechanism still begs for an explanation of why near-neutral alleles have not fixated or gone extinct – a topic we turn to next.

6. Can polygenic mutation-selection balance explain common, harmful, heritable mental disorders?

The simplest polygenic mutation-selection model elegantly parallels the single-gene models described earlier: the equilibrium genetic variation (V_G) maintained in a trait affected by many loci is $V_G = V_M/s$, where V_M is the increase in a trait's genetic variation due to new, harmful mutations per generation, and s is the average selection coefficient against these mutations (Barton 1990). It generally takes a while for these harmful mutations to work their way out of the gene pool. For example, a mutation causing a 1% reduction in fitness will persist in the population until it has passed through an average of 100 individuals (García-Dorado et al. 2003). Mutations with the most harmful effects are removed the fastest, so if one is observed, it is probably rare and of recent origin. Mutations with milder effects are removed more slowly, so they tend to be more common (although still very uncommon in an absolute sense) and older, inherited from parents, grandparents, and so forth. Therefore, genetic variation caused by mutation-selection balance is predominantly the result of old mutations that have yet to go extinct, rather than new mutations, a point that is commonly misunderstood. Most mutations are a family legacy, not an individual foible.

Several recent theoretical papers have emphasized the role of mutation in maladaptive human traits (Gangestad & Yeo 1997; Hughes & Burleson 2000) and late-onset diseases (Wright et al. 2003). Such polygenic mutation-selection models suggest that much of the persistent heritability in traits may be due to a large number of harmful alleles that are individually very rare at any given locus in the population, but that are collectively very common across loci. These models recognize that we do not live in the best of all possible worlds. Genetic information is constantly and inevitably eroded by genetic copying mistakes: mutations. Applied to human mental disorders, mutation-selection models suggest that, if a mental disorder appears maladaptive, maybe it really is maladaptive – and always has been.

6.1. Is polygenic mutation-selection a viable explanation for the genetic variation in traits?

For several reasons that now appear misguided, researchers have often doubted that mutation-selection could explain mental disorder susceptibility alleles. First, the results of many animal studies seemed to suggest that just a few loci (around 2–20) account for much of the genetic variation in traits that had been studied (Falconer & Mackay 1996) – too few for mutation-selection balance to play much of a role in mental disorders. However, there are good reasons to think these studies underestimated the number of loci and overestimated their effect sizes (Barton & Keightley 2002). Moreover, the traits analyzed in these studies generally have little relevance to fitness (e.g., number of abdominal bristles in fruit flies), and mutation

https://doi.org/10.1017/S0140525X06009095 Published online by Cambridge University Press

plus drift at few loci can maintain substantial genetic variation in nearly neutral traits. Second, it is estimated that approximately 7 million single-nucleotide polymorphisms (SNPs) have minor allele frequencies greater than 5% (Kruglyak & Nickerson 2001). However, more than 98% of these 7 million SNPs are outside of protein coding regions and are unlikely to affect mental disorder risk (Wright et al. 2003). SNPs that do affect protein production tend to have minor allele frequencies below 5% (Fay et al. 2001), which is consistent with mutation-selection balance.

Third, as discussed earlier (sects. 2, 4.2, and 5.1), there were strong theoretical expectations that maladaptive states should be rare in nature. Fisher's Fundamental Theorem seemed to suggest that additive genetic variation should be lowest in traits under the strongest selection. This prediction seems supported by observations that traits under more intense selection have lower heritability estimates (Roff & Mousseau 1987). However, heritability is but one way to measure additive genetic variation, and alternative measures of additive genetic variation has turned the canonical story about genetic variation in fitness related traits on its head.

Heritability ($h^2 = V_A/V_P$) is the proportion of total phenotypic variation (V_P) due to additive genetic effects (V_A). V_P is influenced by all sources of variation – not just V_A, but also by environmental variation, random noise, and non-additive genetic effects. Low heritability may well be a result of low V_A, but it might also be caused by high V_P (Charlesworth 1987; Price & Schluter 1991). Charlesworth (1984), Houle (1992), and others have argued that the coefficient of additive genetic variation ($CV_A = \sqrt{V_A}/\bar{x}$) is a better way to measure V_A (i.e., to remove *scale dependence* of V_A), because it is standardized by the trait's mean ($\bar{x}$) rather than by V_P and is therefore not confounded by environmental factors, random noise, or nonadditive genetic effects. (Unfortunately, the use of CV_A requires that the trait can be measured on ratio scales, such as number or time, and is therefore unsuitable for measuring genetic variation in most psychological traits.)

In a seminal study, Houle (1992) found that *traits under stronger selection show substantially higher mean-standardized additive genetic variation than do traits under weaker selection*, despite showing lower heritability. For example, fruit-fly wing length (a trait under relatively weak selection) has a heritability of .36, whereas number of offspring (a trait under intense selection) has a heritability of only .06. However, wing length has a CV_A of only 1.6, whereas number of offspring has a CV_A of 11.9. Across many such comparisons, the mean-standardized V_A of traits under the strongest selection is three to ten times *higher* than that for traits under weaker selection, the opposite of what Fisher's Fundamental Theorem would seem to predict. Similar results have been replicated now in many species, including humans (Hughes & Burleson 2000). These results were astonishing at first and created quite a stir among evolutionary geneticists, leading to a paradox that both parallels and informs the paradox of common, harmful, heritable mental disorders.

6.2. The watershed model explains why traits under the strongest selection have the highest genetic variation

Traits under the most intense selection (*fitness-related traits*, such as successful growth or mating) tend to require the adaptive functioning of many subsidiary biological and behavioral processes, and so depend on very many genes (Charlesworth 1987; Houle 1992; Price & Schluter 1991). The most massively polygenic "trait" is, of course, fitness itself – successful survival and reproduction – which requires the functional coordination of every adaptive mechanism in the body. The mutational "target size" of fitness is quite obviously the entire genome with any effect on fitness, which is probably the vast majority of genes with any phenotypic effect.

The biological network of mechanisms that must function together to create adaptive behaviors can be conceptualized by using a watershed analogy. Much like the numerous tributaries of the Amazonian watershed that coalesce and eventually empty into the Atlantic Ocean, there are many "upstream" micro-biological processes (e.g., rates of neuron proliferation, dendritic pruning, glucose metabolism) that flow into (affect) further "downstream" macro-biological processes (e.g., finding food, making friends, securing mates). A mutation at a locus that affects an upstream process disrupts not only that upstream process, but also every trait downstream of that process. A slightly harmful mutation that affects dendritic pruning may not affect glucose metabolism, but will probably undermine downstream processes such as learning ability, attracting mates, and eventually fitness itself. Figure 3 illustrates this watershed analogy.

The watershed analogy suggests that fitness-related traits have high additive genetic variation because they integrate many processes, and so are massively polygenic. Thus, they are vulnerable to harmful mutations at many loci, and have higher additive genetic variation due to new and old mutations. Fitness-related traits have high V_A, *despite* being under intense selection, not because of it. Their high V_A reflects that they tend to be massively

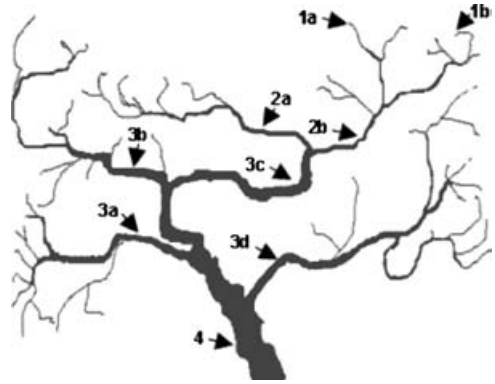

Figure 3. The watershed model of the pathways connecting upstream genes to downstream phenotypes. Mutations at specific loci (1a, 1b) disrupt narrowly defined mechanisms such as transmission of dopamine in the prefrontal cortex (2b). This and other narrowly defined mechanisms contribute noise to more broadly defined mechanisms, such as working memory (3c). Working memory in conjunction with several other mechanisms (3a, 3b, 3d) affects observable phenotypes, such as cognitive ability (4). If enough noise is present in particular upstream processes, specific behavioral syndromes may arise, such as mental disorder symptoms. All tributaries eventually flow into fitness. (Reprinted, with permission, from the *Annual Review of Clinical Psychology, volume 2*. © 2006 by Annual Reviews www.annualreviews.org. [Cannon & Keller 2005])

polygenic. This is not symmetrical: neutral traits are not necessarily influenced by few genes, but fitness-related traits are almost always influenced by many genes. There is now a good deal of support for this model, at least in fruit flies – the animal model of choice for evolutionary geneticists. Among the most compelling pieces of evidence are the high, positive intercorrelations between (a) the estimated number of loci influencing different traits, (b) the estimated trait-level mutation rates, and (c) traits' CV_A (Houle 1998). Charlesworth and Hughes (1999) further estimated that rare, harmful mutations account for 33% to 66% of the additive genetic variation in fitness-related traits in fruit flies.

The watershed model also clarifies why fitness-related traits typically have very high phenotypic variation and therefore moderate to low heritabilities. Downstream traits accumulate any type of noise from upstream traits – not only mutational noise, but also environmental noise (e.g., bad luck with injuries, predators, pathogens, and mates), random noise (e.g., the inherent stochasticity of development), and non-additive genetic effects. Because selection has much less power to reduce the variation in these latter factors compared to additive genetic variation, these factors tend to be proportionately more influential for fitness-related traits, leading to their lower heritabilities (Houle 1992; Merilä & Sheldon 1999).

6.3. The mutational target size of the human brain

The watershed model suggests that fitness-related traits have high genetic variation because they are massively polygenic. How might the watershed model help explain mental disorders? The answer depends upon how many loci influence the mechanisms that, when dysfunctional, cause the behavioral syndromes defined as mental disorders.

Consider the complexity of human brain function in watershed terms. The human brain is the most complex system known to science, with about 100 billion neurons and about a thousand times that many synapses. At least 55% of coding DNA is probably expressed in the human brain (Sandberg et al. 2000). Thus, the brain has an enormous mutational target size – out of the 25,000 protein-coding genes estimated in the human genome, mutations in at least half of them are likely to disrupt brain function, and hence behavior, to some extent (Prokosch et al. 2005).

Yet, there is more to the human genome than protein-coding regions. About as much non-coding DNA as coding DNA is evolutionarily constrained between species, implying that non-coding, regulatory regions are about as important to fitness as protein-coding regions (Keightley & Gaffney 2003). Importantly, non-coding regulatory regions of DNA rarely contribute to Mendelian disorders (McKusick 1998). Thus, harmful mutations in non-coding, regulatory regions seem to have mainly subtle quantitative effects rather than producing dramatic Mendelian catastrophes, and may be especially relevant in explaining the continuously distributed liabilities thought to underlie mental disorders.

How high is the typical human mutation load in brain-expressed loci? Based on conservative estimates, each human carries about 500 to 2,000 slightly harmful older point mutations inherited from ancestors in protein-coding regions (Fay et al. 2001; Sunyaev et al. 2001), plus an average of one or two new fitness-reducing mutations (Eyre-Walker & Keightley 1999). These mutation-load estimates should be at least doubled to account for mutations in non-coding, regulatory DNA, and should be increased slightly to account for mutations involving insertions, deletions, and other changes to chromosomal structure. Given that perhaps half of these mutations affect the brain, we estimate that the average human brain is disrupted by an average of at least 500 genetic mutations.

Apart from a high average mutation load, humans are likely to show high variation in mutational effects. If the numbers of mutations across individuals follows a Poisson distribution, as it would under random mating (S. Gangestad, personal communication, March 3, 2005), the mean and variance in numbers of mutations would be equal, implying a standard deviation of at least 22 mutations ($\sqrt{500}$). However, because humans probably assortatively mate for genetic quality through mutual mate choice (Miller 2000a), the variation in mutation number would be further amplified, so some people should inherit many fewer, and others many more, brain-expressed mutations than average. Moreover, the genetic variation caused by these varying numbers of mutations would be higher still, given that different mutations vary enormously in their effect sizes. The end result will be continuous distributions with respect to almost all psychological dimensions. Individuals with a high load of mutations that affect a particular configuration of upstream cognitive processes would be at higher risk of having mental disorders associated with deficits in downstream behaviors, and would tend to pass this risk on to their offspring. The importance of brain-expressed mutations is consistent with evidence for good genes sexual selection for human mental traits (e.g., Haselton & Miller 2006; Keller, in press; Miller 2000a; 2000b; 2000c; Miller & Todd 1998; Prokosch et al. 2005; Shaner et al. 2004).

6.4. How many loci affect mental disorders?

Before considering this question, it is important to note that the *number of loci affecting a trait* means something different to psychiatric geneticists versus evolutionary geneticists. To psychiatric geneticists, this phrase usually refers only to the loci that currently contribute to the bulk of a trait's genetic variation, which we refer to as the number of *polymorphic loci*. To evolutionary geneticists, however, the "number of loci affecting a trait" usually refers to the much larger number of loci that *could* affect the trait if those loci were polymorphic. It is this latter meaning, which we refer to as the number of *potential loci*, that is relevant to mutation-selection models. Pritchard (2001) estimated that only about 10% of a trait's potential loci will actually be polymorphic at any given time (assuming weak selection), a figure corroborated using a different method by Rudan et al. (2003b).

Recent reviews have invariably concluded that polygenic models (including at least two polymorphic loci) best describe the inheritance of mental disorders such as unipolar depression (Johansson et al. 2001), bipolar disorder (Blackwood et al. 2001), schizophrenia (Sobell et al. 2002), mental retardation (Plomin 1999), and autism (Folstein & Rosen-Sheidley 2001). Beyond this,

however, little is known regarding how many polymorphic loci affect mental disorders, because there has been so little success in actually finding them or modeling their numbers. For example, the data on schizophrenia inheritance are fit equally well by models that predict just a few polymorphic loci (e.g., Risch 1990) and by models that predict an "infinite" number of loci (e.g., Sullivan et al. 2003). The differences in conclusions are largely due to differences in assumptions (additive or epistatic allelic effects; a distinctive syndrome or an extreme of a normally distributed liability) about which no definitive information is available. Nonetheless, it is becoming clear from gene-mapping studies that many loci, at least 5–10 and perhaps many more, must influence the best-studied mental disorders: schizophrenia and bipolar disorder (Kendler & Greenspan, in press).

Rather than further considering assumption-laden models or preliminary empirical results, perhaps it is worthwhile to take a step back and consider carefully what mental disorders, as categories, truly are. Mental disorders are much less objective qualities than age, gender, height, or white blood cell count. Mental disorders are constellations of aberrant behaviors that were lexicalized as unitary disorders by psychiatrists in the nineteenth and early twentieth centuries. There are several possible reasons why mental disorder categories were chosen as they were. First, maybe each mental disorder really has a unitary etiology – a single consistent genetic, neurological, or environmental cause – but few psychiatrists subscribe to such a notion today. Most mental disorders show too much heterogeneity within categories, comorbidity across categories, and continuity with normality, to qualify as discrete, unitary diseases.

Second, as Bleuler (1911) and Jaspers (1923) argued regarding "the schizophrenias," an apparently unitary mental disorder may be a heterogeneous group of dysfunctions in different mechanisms whose final common behavioral pathways lead to similar symptoms. Upstream biological processes that ultimately affect abstract psychological traits are largely hidden from human perception (see 3a and 3b in Fig. 3). They are microscopic neuroanatomical problems hidden within the brain. When these upstream processes dysfunction, humans can usually observe only the downstream behavioral outcomes, and not the specific dysfunctions themselves. Such etiological heterogeneity becomes apparent only in rare cases when dysfunctions in specific upstream mechanisms leave a unique phenotypic signature, in addition to normal symptoms of mental disorders. For example, at least 20 genetic conditions, such as the XXX and XYY karyotypes, congenital adrenal hyperplasia, Wilson's disease, and velocardiofacial syndrome, increase schizophrenic symptoms (Propping 1983). As Vogel and Motulsky (1997) put it,

> Survival of this diagnostic concept was achieved – at least in part – by an interesting strategy: whenever symptoms characteristic of schizophrenia were observed in association with findings that suggested organic disease, the diagnosis of schizophrenia was withheld. . . . [W]hen all [such] patients . . . were excluded, a disease group remained for which specific causative factors could not be found." (p. 700)

Third, and most radically, a mental disorder may be perceived as a coherent category not because it is a "natural kind" with a common etiology at any level, but because it was evolutionarily or culturally adaptive for people to categorize others in particular ways in order to make certain social decisions about them. Thus, insanity may be like ugliness, dishonesty, or aggressiveness – things to avoid and stigmatize in social and sexual interactions – not because they have a unitary etiology, but because they have a common set of fitness costs for observers.

The latter two explanations are not mutually exclusive, of course. We find it likely that apparently unitary mental disorders are partly in the dysfunctions of the sufferer, and partly in the person-perception adaptations of the beholder. Mental disorder categories may reflect a mix of historical convention, diagnostic convenience, innate categorization biases in person perception, and common final pathways of partially overlapping yet distinct dysfunctions. This suggests that *the number of loci affecting a mental disorder depends in large part on the way human minds categorize behavioral symptoms*. The search for endophenotypes (Cannon & Keller 2005; Gottesman & Gould 2003) is critically important because it enables researchers to discern more directly the varied upstream processes whose dysfunctions increase mental disorder risk, while relying less on perceived symptom similarity. The most useful endophenotypes should be those that are further upstream and etiologically less complex. If the past is any guide, the heterogeneity documented in mental disorders so far may be only the tip of the iceberg. Underneath a few simplistic mental disorder categories may lie a vast diversity of potential behavior-impairing mutations across the thousands of genes involved in brain development.

7. Empirical evidence on the three models for common, harmful, heritable mental disorders

We have reviewed several theoretical reasons why polygenic mutation-selection balance may explain the genetic variation underlying mental disorders, much as it explains rare Mendelian disorders. We have also presented some theoretical and empirical reasons to doubt that neutral evolution or balancing selection are good general resolutions to the paradox, although they may play a role under certain specific conditions that we delineated. Fortunately, empirical evidence can help distinguish which of these models goes the farthest in explaining mental disorder susceptibility alleles. We now review six lines of evidence that, taken together, strongly suggest that harmful mutations underlie a substantial portion of the genetic risk in mental disorders.

7.1. Fitness and mental disorders

As noted above (sect. 3.3), mental disorders are associated with lower fertility (due in large part to reduced mating opportunities; see Table 2) and a high level of disability in modern industrialized environments. This is consistent with mutation-selection models, but is less easily reconciled with models of ancestral neutrality and balancing selection. There is one classic example of balancing selection maintaining a highly deleterious condition in humans – sickle-cell anemia – where the strong selection against anemia is balanced by strong selection favoring malarial resistance. To our knowledge, such benefits that balance the harm done by mental disorders have not

been reliably documented for any mental disorder. Indeed, recent evidence on schizophrenia casts doubt that susceptibility alleles for schizophrenia have any hidden benefits, at least in modern environments. If schizophrenia susceptibility alleles are being maintained by either heterozygote advantage or antagonistic pleiotropy, non-affected siblings of schizophrenics should have higher fitness than the general population. However, the best-controlled and largest study of its kind found that 24,000 siblings of 11,000 schizophrenics (sample sizes from all previous studies were fewer than 200 schizophrenics) had the same reproductive success (99.8%) compared to the general population (Haukka et al. 2003). The 2003 study by Haukka and colleagues had plenty of power to detect even minor differences in fitness among relatives of schizophrenics, such as those (around 5%) that might be required if heterozygote advantage maintains the susceptibility allele (Allen & Sarich 1988). Because modern reproductive success may not correlate with ancestral fitness, as we discussed earlier (sects. 3.3 and 4), such evidence does not disprove heterozygote advantage or antagonistic pleiotropy as mechanisms responsible for schizophrenia, but it does weigh against them.

7.2. The effect of trauma on mental disorders

Major genetic abnormalities and environmental insults tend to increase rather than decrease mental disorder risk. For example, chromosomal abnormalities such as trisomy, translocations, and mutations of major effect cause syndromes consistent with autism, mental retardation, schizophrenia, bipolar disorder, and major depression (reviewed in MacIntyre et al. 2003). Traumatic brain injuries increase the risk of mental retardation, schizophrenia, anxiety disorders, and depression (Max et al. 1998; Rao & Lyketsos 2000; Schoenhuber & Gentilini 1988). This type of evidence poses a serious challenge to balancing selection models, particularly those that posit that mental disorders themselves are alternative, complex adaptations maintained by selection. Given that adaptations require the complex coordination of many mechanisms, traumas should disrupt adaptive complexity, not lead to it. Receiving a blow to the head, for example, should not lead to higher intelligence or attractiveness. The direction in which traits move after traumas provide information about the direction of fitness. The mutation-selection model seems most consistent with this evidence: the fact that major phenotypic disruptions (traumas and genetic abnormalities) increase the risk for mental disorders is consistent with the hypothesis that minor phenotypic disruptions (mutations of generally minor effect) do likewise.

7.3. The effect of paternal age on mental disorders

Female humans are born with their full supply of 400+ eggs, and these eggs have gone through only 23 replications, a number that does not change as females age. By contrast, males must continue to produce new sperm throughout life. At age 15, sperm cells have gone through about 35 chromosomal replications, increasing to 380 by age 30, and 840 by age 50 (J. F. Crow 2000). Because each chromosomal replication carries a small chance of a copying error (mutation), the probability of germ-line mutations increases, at a greater than linear rate, with paternal age. Consistent with a mutation-selection model, higher paternal age, but not maternal age, is associated not only with many Mendelian disorders, but also – tellingly – with lower intelligence (Auroux et al. 1989), and an increased risk of mental retardation (Zhang 1992), schizophrenia (Brown et al. 2002; Malaspina et al. 2001; Sipos et al. 2004; although see Pulver et al. 2004), and mental disorders in general (Hare & Moran 1979). Perhaps 15% to 25% of all cases of schizophrenia are a result of this paternal age effect (Malaspina et al. 2001; Sipos et al. 2004), which would be consistent with most other cases being a result of milder, older, more numerous mutations. These paternal age effects are a direct challenge to neutral and balancing selection explanations of mental disorders, but are exactly what would be expected under a mutation-selection model (J. F. Crow 2000).

7.4. The effect of inbreeding on mental disorders

Older harmful mutations tend to be more recessive than new mutations because selection quickly removes mutations with the largest and most dominant harmful effects. *Inbreeding*, or mating between close genetic relatives, reveals the full harmful effects of these old, mostly recessive mutations because offspring of close relatives are homozygous at more loci. Consistent with a mutational role in mental disorder risk, inbreeding in humans has been associated with mental retardation and low intelligence (Vogel & Motulsky 1997), unipolar and bipolar depression (Rudan et al. 2003a), and schizophrenia (Abaskuliev & Skoblo 1975; Bulayeva et al. 2005; Gindilis et al. 1989; Rudan et al. 2003a; although see Chaleby & Tuma 1986; Saugstad & Ödegard 1986). If true, this phenomenon of *inbreeding depression* not only implicates partially recessive harmful mutations in mental disorder risk among *non*-inbred populations; it also shows that selection acted to minimize mental disorder risk in the ancestral past. It is well known in evolutionary genetics that inbreeding depression occurs among traits that have been under directional selection. Ancestral neutrality and balancing selection cannot explain why inbreeding increases mental disorder rates. For example, if schizophrenia risk alleles were maintained by frequency dependence, then inbreeding would be as likely to reduce as to increase schizophrenia risk. Selection only enriches the gene pool with recessive alleles when higher trait values (in this case, higher mental disorder risk) lead to lower fitness.

7.5. Comorbidity between mental disorders

Studies have typically found strong associations between mental disorders; for example, a recent study found that mental disorder comorbidity ranged from 44% to 94%, depending on the mental disorder (Jacobi et al. 2004). This comorbidity appears to be driven in part by pleiotropic genes that simultaneously affect different disorders: there are positive genetic correlations between unipolar depression and generalized anxiety disorder (Kendler et al. 1992), unipolar depression and bipolar disorder (McGuffin et al. 2003), bipolar disorder and schizophrenia (Craddock et al. 2005), autism and unipolar depression (Piven & Palmer 1999), and schizophrenia and several

types of mental retardation (Vogel & Motulsky 1997). Mental disorders are also highly comorbid with many heritable somatic conditions, such as asthma and hypertension (Buist-Bouwman et al. 2005). Comorbidity and positive genetic correlations among mental disorders are nicely explained by mutation-selection models, but would not be expected under ancestral neutrality or balancing selection models. For example, if susceptibility alleles for schizophrenia and bipolar disorder were both ancestrally neutral in their fitness effects, or if their alleles were maintained by balancing selection, there would no particular reason for them to become genetically correlated with each other. On the other hand, if mental disorders are influenced by mutations at hundreds of (potential) loci, which is in the neighborhood of what would be needed for mutation-selection models to explain their prevalence, it would be vanishingly unlikely for each disorder to arise through a mutually exclusive set of genes, given that the human genome includes only about 25,000 protein-coding loci. The genetic risk alleles for mental disorders must overlap quite a lot. This is where the watershed metaphor falls apart: a small mutation (a tributary) can contribute to many different symptoms (rivers); the mapping from genes to mental disorders is many-to-many rather than many-to-one.

7.6. The likely frequencies of mental disorder susceptibility alleles

To guide the search for mental disorder susceptibility alleles, it is crucial to know whether susceptibility alleles are common (one or a few susceptibility alleles per disease locus at high frequencies in the population), or individually rare (one exceedingly predominant non-susceptibility wild-type allele and many different rare susceptibility alleles at each disease locus). Gene mappers differentiate these two possibilities; the first is called the common disease, common variant (CDCV) hypothesis, whereas the latter has been dubbed the common disease, rare variant (CDRV) hypothesis (Wright et al. 2003). To the degree that the CDCV hypothesis reflects the state of the world, current methods of gene mapping should suffice for finding mental disorder susceptibility alleles. On the other hand, the CDRV model suggests that future progress in locating susceptibility alleles will continue to be slow, because the statistical association, between common "marker" alleles and rare susceptibility alleles that gene mapping requires, will be low or nonexistent (Terwilliger & Weiss 1998; Weiss & Clark 2002; Wright & Hastie 2001). Moreover, if susceptibility alleles are rare, they must exist at a large number of loci to explain mental disorder rates and heritabilities, which would further decrease the power of gene-mapping studies. Understandably, the CDRV model has not been well received among psychiatric geneticists. A speaker at a major gene-mapping conference conceded that this CDRV scenario was too depressing to contemplate, and so it was better to proceed as if it were not true (Wright & Hastie 2001).

The three models of selection each leave different signatures in the genome that correspond roughly to the CDCV model or the CDRV model (Bamshad & Wooding 2003; Kreitman 2000). One of the strongest predictions from practically every model of balancing selection is that it will lead to relatively few polymorphic loci, each harboring just a few (usually two) different alleles at fairly high frequencies (minor allele frequencies greater than about 5%, which we call *common alleles*), that account for most of the genetic variation in the trait (Barton & Keightley 2002; Roff 1997). This appears to hold whether the balancing selection is for discrete or continuous trait variation (Mani et al. 1990).

Whereas the prediction that balancing selection leads to common alleles appears robust, the prediction that balancing selection leads to just one or a few loci being polymorphic is more nuanced. The latter prediction applies only to the number of loci that influence traits directly under balancing selection. If a trait is not under balancing selection (i.e., is under directional or stabilizing selection), some of the alleles that influence the trait may nevertheless be pleiotropic and under balancing selection for reasons unrelated to the trait in question (e.g., Turelli & Barton 2004). In this case, there is no limit on the number of loci under balancing selection that might influence the trait. For example, it is possible that schizophrenia risk has always been maladaptive (under directional selection), but that many of the (pleiotropic) loci affecting schizophrenia risk also affect immune functioning and have been under frequency-dependent selection for immunity (see sect. 5.8). Therefore, if mental disorder risk is a pleiotropic side effect of genes that are under balancing selection on other traits, then common alleles – but at an unknown number of loci – should be responsible for most of the mental disorder genetic risk. If mental disorder risk is directly under balancing selection, as many Darwinian psychiatrists have postulated, common alleles at just a few loci should be responsible for most of the genetic risk of mental disorders. Regardless, if balancing selection maintains susceptibility alleles for whatever reason, there should be only a few common susceptibility alleles at each risk locus, and the CDCV model should be true.

Neutral evolution predicts that alleles will be somewhat less common than they would if governed by balancing selection. If neutral susceptibility alleles happened to be common in ancestral human populations, they should still be common today (Reich & Landers 2001). However, as noted earlier (sect. 4.1), genetic drift in small ancestral human populations tends to drive neutral loci to fixation, through random sampling error. Indeed, most neutral loci seem to have one predominant allele and, due to the recent increase in human population size, many individually rare alleles, although some neutral loci also have common alleles (Cargill et al. 1999; Halushka et al. 1999). Thus, neutral evolution should lead to a situation somewhere between the CDCV model and the CDRV model.

Widespread mutation-selection, on the other hand, should lead to a world where the CDRV hypothesis is true. A trait's genetic variation should be a result of mutations at many different loci. The more deleterious and common the trait was ancestrally, the more loci would have to be involved; very serious and common mental disorders may be affected by hundreds or even thousands of potential loci, but only a portion of these should contribute to the bulk of standing genetic variation in any given population at any given time (Pritchard 2001). At each locus, numerous different mutations should exist,

none of which should be at high frequencies (e.g., minor allele frequencies of less than 5%). However, in cases where selection against susceptibility alleles has been minute (e.g., $s < 1/5{,}000$), such as might occur in the case of gene-by-environment interactions (sect. 4.4), some susceptibility alleles could be at high frequencies, despite selection, due to random genetic drift (Pritchard 2001).

The historical success or failure of psychiatric gene hunting helps clarify which of the three evolutionary models – ancestral neutrality, balancing selection, mutation-selection balance – best explains the existence of the bulk of susceptibility alleles. The CDRV model, most consistent with mutation-selection, predicts the least progress in psychiatric gene hunting; whereas the CDCV model, most consistent with balancing selection, predicts the most. Where does the evidence stand? Once again, mutation-selection seems to best fit the evidence. Only a handful of replicable susceptibility alleles for mental disorders have been found despite two decades of intensive research involving thousands of scientists and hundreds of millions of dollars. Acclaimed discoveries of mental disorder susceptibility alleles have typically been followed by repeated failures to replicate (Terwilliger & Weiss 1998; Weiss & Clark 2002). At the same time, the molecular bases for over 1,700 Mendelian phenotypes have been definitively found to date (*Online Mendelian Inheritance of Man*, April 10, 2006), showing that current methods are wildly successful at finding alleles responsible for single-gene, Mendelian disorders.

Even for these susceptibility alleles that have been located, the effect sizes have been very small. One of the more comprehensive recent meta-analyses (Lohmueller et al. 2003) showed that only two of the eight most-studied mental disorder susceptibility alleles (at the DRD3 and HTR2A loci) were reliably associated with a mental disorder (schizophrenia). The meta-analysis estimated the true odds ratio for the larger of the two associations was just 1.12, meaning that given 1,000 people with the DRD3 susceptibility allele and 1,000 people without it, 11 people in the first group and 10 in the second group will probably develop schizophrenia (given its 1% base rate). Several other meta-analyses have also recently concluded that discovered mental disorder susceptibility alleles tend to have small effects (odds ratios less than 1.5; Kendler 2005). The susceptibility alleles underlying most of the genetic risk for mental disorders have not yet been found. If those that have been found represent the "low-hanging fruit" (explaining the most variation in the population), then the remaining susceptibility alleles may be even rarer and harder to detect.

We are not casting doubt on the entire enterprise of gene hunting. Susceptibility alleles explaining the most risk variation in the population, many of which may have been found already, could be common because of balancing selection on separate traits, recent bottlenecks among certain groups, or genetic drift (caused by fitness effects that were closer to neutral ancestrally). If such susceptibility alleles happened to reach frequencies above 5% in ancestral times, their current allelic complexity should still be low, and gene-hunting techniques should be sufficient for finding them (Reich & Lander 2001). Some protective alleles may be sweeping toward fixation caused by recent selection. Some lineage-specific susceptibility alleles may be missed within an analysis or not replicated across analyses because of hidden population substructures that arose across evolutionary history. Technological advancements may eventually enable discovery of even the rarest susceptibility alleles, the base-pair sequences of which would provide important information about the relative importance of ancestral neutrality, balancing selection, and mutation-selection balance (Bamshad & Wooding 2003; Otto 2000). Nevertheless, the slow progress in finding mental disorder susceptibility alleles so far, and the small amount of explained population risk of those that have been found, are generally consistent with the mutation-selection model and the CDRV model. If balancing selection, and to a lesser degree ancestral neutrality, were general explanations for mental disorders, then psychiatric genetics probably would have already found the susceptibility alleles responsible for most of the genetic variation underlying them.

8. Conclusions: Toward a resolution of the paradox of common, harmful, heritable mental disorders

Evolutionary anthropologist Donald Symons observed that "you cannot understand what a person is saying unless you understand who they are arguing with" (Cosmides & Tooby, n.d.). In this article, we are arguing mostly against those evolutionary thinkers who assume that adaptive forces are the only possible explanations for common, heritable polymorphisms such as mental disorders, even when those traits look profoundly harmful to survival and reproduction. We are also arguing against those psychiatric geneticists who disregard evolutionary theory when trying to understand mental disorders or their susceptibility alleles. This article has tried to show how evolutionary genetic theories are important to both fields.

Evolutionary psychologists have struggled to explain genetic variation in the context of species-typical adaptive design – sometimes ignoring it, sometimes citing mismatches between ancestral and current environments, and sometimes trying to find hidden adaptive benefits maintained by balancing selection. These approaches all draw upon the familiar adaptive toolbox, in which the optimizing power of natural selection is assumed. This is a great toolbox to use when trying to reverse engineer universal aspects of human nature such as vision, mate choice, or normal reactions to depression-inducing situations. Indeed, the search for possible adaptive functions of mental disorder symptoms, especially when the capacity to express these symptoms is universal and they are environmentally triggered, is an important counterbalance to the prevailing assumption that subjective distress equals biological disorder. However, a very different set of tools is required to explain persistent genetic variation, especially in traits related to fitness. These tools must be drawn from contemporary evolutionary genetics.

Psychiatric genetics has, with some pride, traditionally been an empirically driven field. This approach is commendable to a degree. However, as Einstein once observed, "It is the theory that decides what we can observe," and evolutionary genetics provides a rigorous

mathematical framework that could better guide psychiatric gene hunting (e.g., Pritchard 2001; Reich & Lander 2001; Rudan et al. 2003a; Wright et al. 2003). For example, mutation-selection models suggest that susceptibility alleles with the largest effect sizes may also be the rarest, the most recent, and the most population specific – an insight with important implications for the methods most likely to locate mental disorder susceptibility alleles (see Wright et al. 2003). Moreover, mutation-selection explanations further justify the search for less polygenic, and more genetically mappable, endophenotypes (Cannon & Keller 2005).

The existence of common, heritable, harmful mental disorders creates an apparent evolutionary paradox, but we think it can be resolved by recognizing the enormous mutational target size of human behaviors. According to this model, behavioral traits are especially susceptible to harmful mutations because they depend on the most complex organ in the human body. The brain is affected by over half of the hundreds of mutations that all humans carry. Some of these mutations have large, distinctive effects, and so are reliably recognized as Mendelian disorders. Tellingly, some of *these* mutations cause syndromes inherited in Mendelian fashion but that are otherwise phenotypically identical to mental disorders (MacIntyre et al. 2003). Mendelian disorders are rare because selection keeps very harmful mutations very rare.

Most other mutations, especially in regulatory regions of the genome, have much milder effects and cause mostly quantitative differences in traits. Individuals with an especially high load in mutations that disrupt a particular configuration of brain systems will tend to act in aberrant, harmful ways that provoke social comment and psychiatric categorization. Lacking a map of the neurogenetic watershed, psychiatrists have struggled to identify criteria that could enable these behavioral syndromes to be meaningfully categorized. Current criteria reflect perceived similarity of symptoms and prognoses, which is potentially influenced not only by actual etiological similarity, but also by the cultural and inherent person-perception biases of those perceiving the sufferer, and the categorization demands of legal, medical, and research systems. Common mental disorders are common because they are defined that way.

It was natural that these mental disorder categories became reified, and that scientists looked for single genes underlying them, which was so successfully accomplished with Mendelian disorders. But common mental disorders are probably fundamentally different than Mendelian disorders – not, as has often been presumed, in that the former were not selected against while the latter were – but rather, in that common mental disorders are influenced by a much larger number of environmental and genetic factors, most of which have only minor influences on overall population risk.

Everyone alive, according to this model, has minor brain abnormalities that cause them to be a little bit mentally retarded, a little bit emotionally unstable, and a little bit schizophrenic. If so, this framework may help explain much more than just mental disorders; it may help explain genetic differences between people in personality, health, athleticism, intelligence, attractiveness, and virtually any other trait related to Darwinian fitness. If scientists so chose, they could define the low-fitness extremes of any of these dimensions as "disorders." The susceptibility alleles contributing to such "disorders" would be the same ones responsible for genetic variation across the whole dimension in the general population. All other things being equal, someone of below-average athleticism harbors an above-average number of athleticism-degrading mutations. Adaptive organic complexity is exquisite as an abstraction, but riddled with errors within any living, breathing individual. We are all very imperfect versions of that Platonic ideal, the species-typical genome. This perspective may help explain evidence of ubiquitous maladaptation; for example, why nearly everyone suffers from some type of heritable physical ailment, or why about half of people will meet DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders, 4th edition*) criteria for a mental disorder at some point in their lives (Kessler et al. 2005).

The theoretical and empirical evidence reviewed in this article is most consistent with a polygenic mutation-selection balance model for explaining common, harmful, heritable mental disorders. Ancestral neutrality and balancing selection almost certainly play roles in maintaining some susceptibility alleles, but, as general explanations, they are difficult to reconcile with empirical evidence that mental disorders are associated with (1) reduced fitness, (2) brain trauma, (3) higher paternal age, (4) inbreeding, (5) genetic comorbidity, and (6) many susceptibility alleles that explain little population risk. So far, the evidence suggests that mutation-selection plays an important role in maintaining susceptibility alleles of mental disorders, whereas the other forces play less certain roles. At the very least, we hope to have demonstrated that there is no necessary paradox in the existence of common, heritable, harmful traits, such as mental disorders, and we hope to have shown the types of empirical evidence that can test different evolutionary theories of susceptibility alleles. It is possible, of course, that new empirical evidence, or new understandings of how genes affect phenotypes, will show that our conclusions were substantively wrong. It is also possible that we have made mistakes in interpretations of data or theory. This is, after all, a persistent danger in multidisciplinary work, but we feel strongly that the difficulties of integrating such disparate fields are far outweighed by the potential advantages. We look forward to a future in which Darwinian psychiatry, psychiatric genetics, and evolutionary genetics become more mutually informative and supportive.

ACKNOWLEDGMENTS

For helpful guidance and generous feedback, thanks to Paul Andrews, Rosalind Arden, Nick Barton, Reinhard Bürger, Tyrone Cannon, Greg Carey, Dan Fessler, A. J. Figueredo, Steven Gangestad, Martie Haselton, Kenneth Kendler, Nick Martin, Emily Messersmith, Michael Neale, Randolph Nesse, Andrew Shaner, Srijan Sri, Kim Tremblay, Jerome Wakefield, X. T. Wang, and Kenneth Weiss. The first author was supported by a National Science Foundation Graduate Research Fellowship, a fellowship from the UCLA Center for Society and Genetics, and a National Research Service Award from the National Institutes of Health, T32 MH-20030 (PI M. C. Neale).