The Evolution of Personality Variation in Humans and Other Animals

Daniel Nettle
University of Newcastle

A comprehensive evolutionary framework for understanding the maintenance of heritable behavioral variation in humans is yet to be developed. Some evolutionary psychologists have argued that heritable variation will not be found in important, fitness-relevant characteristics because of the winnowing effect of natural selection. This article propounds the opposite view. Heritable variation is ubiquitous in all species, and there are a number of frameworks for understanding its persistence. The author argues that each of the Big Five dimensions of human personality can be seen as the result of a trade-off between different fitness costs and benefits. As there is no unconditionally optimal value of these trade-offs, it is to be expected that genetic diversity will be retained in the population.

Keywords: personality, evolutionary psychology, five-factor model

In recent years, there has been an extraordinary growth of interest in giving ultimate, evolutionary explanations for psychological phenomena alongside the proximate, mechanistic explanations that are psychology’s traditional fare (Barkow, Cosmides, & Tooby, 1992; Buss, 1995). The logic of ultimate explanations is that for psychological mechanisms and behavioral tendencies to have become and remain prevalent, they must serve or have served some fitness-enhancing function. The explanatory program of evolutionary psychology has concentrated strongly on human universals, such as jealousy, sexual attraction, and reasoning about social exchange (Buss, 1989; Buss, Larsen, Westen, & Semmelroth, 1992; Cosmides, 1989). The focus has been on the central tendency of the psychology of these domains, rather than the observed variation, and explanation has been in terms of adaptations shared by all individuals.

Indeed, some evolutionary psychologists have implied that one should not expect there to be any important variation in traits that have a history of selection. For example, Tooby and Cosmides (1992) argued that “human genetic variation . . . is overwhelmingly sequestered into functionally superficial biochemical differences, leaving our complex functional design universal and species typical” (p. 25). The reason invoked for this assertion is that natural selection, which is a winnowing procedure, should, if there are no countering forces, eventually remove all but the highest-fitness variant at a particular locus (Fisher, 1930; Tooby & Cosmides, 1990), especially because complex adaptations are built by suites of genes whose overall functioning tends to be disrupted by variation. Because of the winnowing nature of selection, the existence of heritable variation in a trait is argued to be evidence for a trait’s not having been under natural selection: “Heritable variation in a trait generally signals a lack of adaptive significance” (Tooby & Cosmides, 1990, p. 38, italics in original).

Tooby and Cosmides (1992) thus suggested that most of the genetic variation between human individuals is neutral or functionally superficial. They did, however, concede a possible role for “some thin films” of functionally relevant heritable interindividual differences (Tooby & Cosmides, 1992, p. 80). The possible sources of these thin films—frequency-dependent selection and selective responses to local ecological conditions—are discussed in greater detail below. What is relevant for the present purposes is the low priority given to understanding these thin films relative to the task of describing and explaining the universal psychological mechanisms that humans undoubtedly all share.

Personality Variation and Evolutionary Psychology

There has however, been a response from researchers seeking to marry differential and evolutionary psychology in a way that gives greater weight to the study of individual differences (see, e.g., Buss, 1991; Buss & Greiling, 1999; Figueredo et al., 2005; Gangestad & Simpson, 1990; MacDonald, 1995; Nettle, 2005). David Buss made an early contribution to this literature by enumerating possible sources of functionally important interindividual variation (Buss, 1991; see also Buss & Greiling, 1999). Most of these mechanisms that do not rely on heritable variation in psychological mechanisms, for example, enduring situational evocation, or calibration by early life events, or calibration of behavior by the size or state of the individual. However, Buss also discussed the possibility that there are equally adaptive alternative behavioral strategies underlain by genetic polymorphisms, or continua of reactivity of
psychological mechanisms, in which there is no universal optimum, and so genetic variation is maintained.

The idea of continua of reactivity was taken up by Kevin MacDonald (1995). MacDonald proposed that the normal range of observed variation on personality dimensions represents a continuum of viable alternative strategies for maximizing fitness. In this view, average fitness would be about equal across the normal range of any given personality dimension, but individuals of different personality levels might differ in the way that they achieved their fitness—for example, by investing in reproductive rather than parental effort. Implicit in MacDonald’s formulation, but perhaps not examined in enough detail, is the concept of trade-offs. The idea of trade-offs is reviewed in detail below, but the key point is that if two levels of a trait have roughly equal fitness overall and if increasing the trait increases some component of fitness, then it must also decrease other components. Every benefit produced by increasing a trait must also produce a cost. If this is not the case, there is no trade-off, and natural selection is directional toward the higher value of the trait.

The purposes of the present article are several. First, no reasonable biologist or psychologist should disagree with Tooby and Cosmides (1990, 1992) that humans’ psychological mechanisms show evidence of complex design and are largely species-specific. Nor need differential psychologists deny the importance of the branches of psychology that are devoted to the study of species-typical mechanisms. However, I argue that a more up-to-date reading of the very biology from which Tooby and Cosmides draw their inspiration leads to a rather different view of the extent and significance of variation. The films of functionally significant interindividual variation need not be particularly thin. The first purpose of this article, then, is to review interindividual variation in nonhuman species, with particular attention to the way that selection can allow variation to persist even when it is relevant to fitness.

Second, although Buss’s (1991) and MacDonald’s (1995) reviews have been influential in enumerating possible evolutionary mechanisms that lie behind the persistence of personality differences, there has as yet been relatively little work in evolutionary personality psychology that actually tests the predictions of these models empirically (for some exceptions, see Figueredo et al., 2005; Gangestad & Simpson, 1990; Nettle, 2005). The bulk of the work in personality psychology goes on uninspired by considerations of ultimate evolutionary origins. The second purpose of this article, therefore, is to build from MacDonald’s ideas of personality dimensions as alternative viable strategies, outlining a more explicit framework of costs and benefits, and to apply this framework to each of the dimensions of the five-factor model of personality. This approach allows existing studies that were done from a largely inductive, atheoretical perspective to be interpreted more coherently through the long lens of adaptive costs and benefits. In addition, the approach allows the generation of novel predictions and ideas for future research.

Variation in Humans

There is abundant evidence of interindividual biological variation in humans, not just at the phenotypic level but also at the genotypic level. For example, the genes coding for the serotonin and dopamine transmitter-receptor systems are massively and ubiquitously polymorphic in the human population (Cravchik & Goldman, 2000). These systems seem to be importantly involved in health, stress, motivated behavior, and sociality, and the genetic variation in them is associated with variation in behavioral outcomes in these domains (Chen, Burton, Greengberger, & Dmitrieva, 1999; Depue & Collins, 1999; Knutson et al., 1998; Lesch et al., 1996; Reif & Lesch, 2003). It is hard to call such variation functionally superficial.

Moreover, behavior geneticists have unanimously estimated high heritability coefficients for a wide range of important human traits, from personality factors (Bouchard & Loehlin, 2001) to intelligence (Bouchard & McGue, 1981), attitudes (Alford, Funk, & Hibbing, 2005), and vulnerability to psychological illnesses (Cardno et al., 1999; Sullivan, Neale, & Kendler, 2000). Many of these characteristics are demonstrably related to reproductive success. For example, schizophrenia affects life expectancy and reproductive success (Avila, Thaker, & Adami, 2001; Bassett, Bury, Hodgkinson, & Honer, 1996; Brown, Inskipp, & Barracough, 2000), and personality affects mating success (Nettle, 2005), health (Neelamman, Sytenta, & Wadsworth, 2002) and life expectancy (Friedman et al., 1995). High heritability coefficients, especially when the trait is common, mean that the population contains abundant genetic variation. Given the winnowing effect of selection, such variation—which is clearly more than a “thin film”—seems to require explanation. With a view to providing one, I now turn to a review of genetic variation in nonhuman populations.

The Evolution of Variation

Heritable variation is ubiquitous in natural populations. Variation is, after all, the grist for the mill of evolution, and it was the observation of variation that provided Darwin with his foundational insights. For quantitative traits, Lynch and Walsh (1998), in their textbook of evolutionary genetics, stated that “almost every character in every species that has been studied intensely exhibits nonzero heritability” (p. 174). This includes abundant variation in traits demonstrably related to fitness (Houle, 1992). For behavior, heritable diversity has been observed in many species, sometimes taking the form of continuous dimensions akin to personality axes (Gosling, 2001; an example is the propensity to call in male crickets, Gryllus integer; Cade, 1981) and sometimes taking the form of discrete morphs or strategies with a simple genetic basis (an example is independent vs. satellite mating in the ruff, Philomachus pugnax; Lank, Smith, Hanotte, Burke, & Cooke, 1995).

Why does variation persist? The ultimate source of variation is of course mutation. The level of variation found in a population at any point in time reflects the balance between mutation introducing new variants and selection
removing them. For a trait influenced by a single gene, selection does not have to be very strong to keep the level of standing variation close to zero, because mutations arise relatively infrequently. However, the rate of appearance of mutations affecting a trait rises directly with the number of genes involved in building it, and selection does not remove them instantaneously. Thus, for a trait affected by many genes, even if selection is strong, there will be a significant amount of standing genetic variation (Houle, 1998).

Varying the level of investment in a quantitative trait is rarely simply advantageous or simply disadvantageous. There are many components to overall fitness, and increasing investment in one is usually done at the expense of others. For example, growing large, though possibly beneficial in intraspecific competition, raises metabolic costs and also lengthens the time spent growing and thus delays the onset of reproduction. In the pygmy swordtail, *Xiphophorus nigrensis*, there are large males who perform elaborate courtship displays and small males who do not display but seek “sneak” copulations. The difference is underlain by a genetic polymorphism (Zimmerer & Kallman, 1989). Though the large displaying males are generally preferred by females, it takes them 27 weeks to reach sexual maturity, as opposed to 14 weeks for the sneak males (Zimmerer & Kallman, 1989). Thus, continuing growth can be seen as a strategy, one that has both costs and benefits, and the same reasoning can be applied to many aspects of behavior and development (Roff, 1992; Stearns, 1992).

At any point in time and space there will be an optimum value of a trait from the point of view of fitness. However, this optimum may vary rapidly across time and space. This has been best documented in a long series of investigations of Galapagos finches by Peter Grant and colleagues (B. R. Grant & Grant, 1989; P. R. Grant, 1986). For example, a dry year in 1977 drastically reduced food availability for the seed-eating finches (*Geospiza fortis*) on the island of Daphne, favoring larger birds with deeper beaks who could open large hard seeds. Selection was so strong that by 1978, the average body size of a young finch on Daphne was 0.31 standard deviations higher than it had been in 1976, and beak depth had increased by 4% (P. R. Grant, 1986). However, with the return of the rains in 1978, smaller, soft seeds would have become abundant once more and this selection possibly reversed. Galapagos finch populations maintain high levels of genetic diversity as a consequence of this fluctuating selection.

Experimental genetics also confirms this relationship. In an experimental colony of the fruit fly *Drosophila melanogaster*, allowing the food regime to fluctuate over time led to the maintenance of more heritable genetic variation than was maintained by holding the food regime constant (MacKay, 1981).

The complex interaction between spatiotemporal variation in selective optima and trade-offs in different components of fitness can also be applied to behavioral dispositions. To give one example, in the Trinidadian guppy (*Poecilia reticulata*), there is heritable variation in behavioral traits that affect the probability of survival in the presence of a piscivorous fish (Dugatkin, 1992; O’Steen, Cullum, & Bennett, 2002). Guppies that come from populations living upstream of waterfalls, where there are no such predators, are bolder and less likely to survive in the presence of a pike, whereas those downstream show an enhanced ability to do so. These effects are heritable and thus not based on individual experience with predators. When predators are introduced into previously predator-free streams, change in the population distribution of behavior in the presence of predators is rapid. However, when predators are removed, circumspection in the presence of predators is rapidly lost, and within 20 years or so, the population becomes like those with no history of predation at all (O’Steen et al., 2002). This suggests that antipredator vigilance is imposing costs—in terms, for example, of lost foraging or mating time—as well as providing benefits, so selection will move the population distribution around in response to local predation intensity.

There is gene flow between different guppy populations, and so if the population of guppies is considered in toto, there is a normal distribution, underlain by genetic polymorphism, of antipredator vigilance, even though at any specific point in time and space there may be a single optimum level.

Such interactions can be even more complex, as Dingemanse and colleagues have shown in a series of exquisite recent experiments on the great tit, *Parus major* (Dingemanse, Both, Drent, & Timmergen, 2004; Dingemanse, Both, Drent, Van Oers, & Van Noordwijk, 2002; Dingemanse, Both, van Noordwijk, Rutten, & Drent, 2003). Individuals of this species differ on a behavioral dimension called exploration, with high scorers being aggressive and bold in exploring the environment and low scorers exploring less freely and showing low levels of aggression. Exploration is consistent within individuals and substantially heritable (*h*² = 0.3–0.6, which is in the same range as human personality dimensions). Dingemanse et al. (2004) showed that in poor years, when food resources are scarce, there is a linear positive relationship between exploration score and the probability of survival for females. This is because the bolder individuals are more successful at locating and competing for what resources there are. However, in years of abundance, when there is mast seeding of beech trees, there is a strongly negative linear relationship between female survival and exploration score. The authors suggested this may be because high-scoring individuals become involved in dangerous and costly aggressive encounters that, with resources limitless, have no benefit. For males, the patterns are diametrically opposite. Males are generally dominant in this species, and much of their effort is directed to defending territories. In poor years, males lower in boldness do well. This is because the bolder individuals are more successful at locating and defending territories. In poor years, males lower in boldness do well. This is because with higher overall mortality, competition for territories is relaxed, and thus aggressive interactions with other males are less important. However, in good years, more fledglings survive and compete for territories, and so there is increased male–male aggression. In this situation, males with a high exploration score fare better.
The great tit experiments show that levels of exploration have both costs and benefits—costs in terms of getting into aggressive encounters, with all the harm that can bring, and benefits in terms of holding resources or territories when these are limited. The optimal balance between these costs and benefits depends on exact local conditions and the sex of the individual. As these vary, the overall population maintains a normal distribution of exploratory tendencies, with genetic polymorphism underlying it.

In the guppy and great tit examples, there is a single fitness optimum, albeit one that varies in time and space. Fitness functions need not, however, be unimodal. In the coho salmon, *Oncorhynchus kisutch*, there are two male morphs: “hooknose” males, who delay maturity, grow large, and compete with other males to fertilize female egg deposits, and rarer “jacks,” who remain small, hide near nests, and sneak in to fertilize eggs. Gross has shown that fitness for large and small-sized males is roughly equal, whereas intermediate-sized males, too small to fight effectively and too large to hide in refuges and follow a sneak strategy, are at a large disadvantage (Gross, 1985). Thus, selection on size at maturity is disruptive, and heritable variation is maintained.

The extreme case of variable optima is what is known as negative frequency-dependent selection. This describes the situation in which the relative fitness of a trait is high as long as it is rare in the local population but declines as it becomes widespread. It has long been recognized that negative frequency dependency can in theory lead to the maintenance of polymorphism (Maynard-Smith, 1982), though it has been harder to demonstrate empirically and unequivocally that such a mechanism is in operation. One such demonstration comes from the blugill sunfish, *Lepomis macrochirus* (Gross, 1991). Males of this species can be either parental or cuckold, with the former delaying reproduction and building nests and the latter maturing small and early and sneaking into nests built by other males. Gross (1991) showed that the reproductive success of the cuckold males is high when they are rare in a colony but declines steeply as they become more common, whereas intermediate-sized males, too small to fight effectively and too large to hide in refuges and follow a sneak strategy, are at a large disadvantage (Gross, 1985). Thus, selection on size at maturity is disruptive, and heritable variation is maintained.

This brief review of behavioral polymorphism in other species has several implications for the maintenance of personality variation in humans. Here, I wish to focus on two generalizations. First, variation is a normal and ubiquitous result of the fluctuating nature of selection, coupled with the large numbers of genes that can affect behavior. Frequency-dependent selection, oft discussed as a maintainer of variation, is in fact just a subcase of the more general phenomenon of fluctuating selection. Second, behavioral alternatives can be considered as trade-offs, with a particular trait producing not unalloyed advantage but a mixture of costs and benefits such that the optimal value for fitness may depend on very specific local circumstances. With these generalizations in mind, I now turn to the consideration of personality variation in humans.

### Human Personality Traits

The rest of this article follows the structure of the five-factor model of personality (Costa & McCrae, 1985, 1992; Digman, 1990). Though the five broad factors, or domains, are decomposable into finer facets (Costa & McCrae, 1985) and certainly do not capture all the variation in human personality (Paunonen & Jackson, 2000), there is broad consensus that they are useful representations of the major axes of variation in human disposition (Digman, 1990). Following the considerations outlined in the previous section, I briefly examine the nature of each domain and consider the kinds of costs and benefits that increasing the level of the domain might have with respect to biological fitness. The reviews here are speculative, but they are offered in the hope of stimulating empirical work and of drawing psychologists’ attention to the idea that changing the level of a trait is associated with fitness costs as well as fitness benefits.

### Extraversion

A dimension related to positive emotion, exploratory activity, and reward is a feature common to all personality frameworks and theories. Its most common label is *extraversion*, and its proximate basis is thought to involve variation in dopamine-mediated reward circuits in the brain (Depue & Collins, 1999).

I have outlined a trade-offs-based evolutionary model for the maintenance of polymorphism in extraversion (Nettle, 2005). Extraversion is strongly and positively related to number of sexual partners (Heaven, Fitzpatrick, Craig, Kelly, & Sebur, 2000; Nettle, 2005), which, for men in particular, can increase fitness. High scorers are also more likely to engage in extrapair copulations or to terminate a relationship for another. This may lead to their securing mates of higher quality than those secured by individuals who are more constant in their choice of partners. The benefits of extraversion are not limited to mating, as extraverts, or those high on the closely correlated trait of sensation seeking, initiate more social behavior (Buchanan, Johnson, & Goldberg, 2005) and have more social support (Franken, Gibson, & Mohan, 1990) than others. Moreover, they are more physically active and undertake more exploration of their environment (Chen et al., 1999; Kirkaldy, 1982).

However, in pursuing high sexual diversity, and high levels of exploration and activity in general, extraverts also expose themselves to risk. Those who are hospitalized due to accident or illness are higher in extraversion than those who are not (Nettle, 2005), and those who suffer traumatic injury have been found to be high in sensation seeking (Field & O’Keefe, 2004). High extraversion or sensation-seeking scorers also have elevated probabilities of migrating (Chen et al., 1999), becoming involved in criminal or antisocial behavior (Ellis, 1987), and being arrested (Samuels et al., 2004). All of these are sources of risk, risk that in the ancestral environment might have meant social ostracism or death. Moreover, because of their turnover of relationships, extraverts have an elevated probability of
exposing their offspring to stepparenting, which is a known risk factor for child well-being.

One can thus conceive of extraversion as leading to benefits in terms of mating opportunities and exploration of novel aspects of the environment but carrying costs in terms of personal survival and possibly offspring welfare. It is unlikely that there will be a universal optimal position on this trade-off curve. Instead, local conditions, including the density and behavioral strategies of surrounding individuals, could lead to a constant fluctuation in the optimal value, and hence genetic polymorphism would be retained.

Neuroticism

The neuroticism personality axis is associated with variation in the activity levels of negative emotion systems such as fear, sadness, anxiety, and guilt. The negative effects of neuroticism are well-known in the psychological literature. High neuroticism is a strong predictor of psychiatric disorder in general (Claridge & Davis, 2001), particularly depression and anxiety. Neuroticism is also associated with impaired physical health, presumably through chronic activation of stress mechanisms (Neeleman et al., 2002). Neuroticism is a predictor of relationship failure and social isolation (Kelly & Conley, 1987).

A much more challenging issue, then, is finding any compensatory benefit to neuroticism. However, given the normal distribution observed in the human population, and the persistence of lineages demonstrably high in the trait, such a benefit seems likely.

Studies in nonhuman animals, such as puppies (see the Evolution of Variation section), suggest that vigilance and wariness are both highly beneficial in avoiding predation and highly costly because they are quickly lost when predation pressure is absent. In ancestral environments, a level of neuroticism may have been necessary for avoidance of acute dangers. Anxiety, of which neuroticism can be considered a trait measure, enhances detection of threatening stimuli by speeding up the reaction to them, interpreting ambiguous stimuli as negative, and locking attention onto them (Mathews, Mackintosh, & Fulcher, 1997).

Because actual physical threats are generally attenuated in contemporary situations, the safety benefits of neuroticism may be hard to detect empirically. However, certain groups who take extreme risks, such as alpinists (Goma-i-Freixanet, 1991) and Mount Everest climbers (Egan & Stelmack, 2003), have been found to be unusually low in neuroticism. Given the high mortality involved in such endeavors (around 300 people have died in attempting Everest), this finding suggests that neuroticism can be protective.

There may also be other kinds of benefits to neuroticism. Neuroticism is positively correlated with competitiveness (Ross, Stewart, Mugge, & Fultz, 2001). McKenzie has shown that, among university students, academic success is strongly positively correlated with neuroticism among those who are resilient enough to cope with its effects (McKenzie, 1989; McKenzie, Taghani-Knosary, & Tindell, 2000). Thus negative affect can be channeled into striving to better one’s position. However, here neuroticism certainly interacts with other factors. When intelligence or conscientiousness is high, for example, the outcomes of neuroticism may be significantly different than when such factors are low.

Thus it is quite possible that very low neuroticism has fitness disadvantages in terms of lack of striving or hazard avoidance. Although very high neuroticism has evident drawbacks, it may also serve as a motivator to achievement in competitive fields among those equipped to succeed. Thus the optimal value of neuroticism would plausibly depend on precise local conditions and other attributes of the person, leading to the maintenance of polymorphism.

Openness

The trait of openness to experience again seems, at first blush, to be an unalloyed good. Openness is positively related to artistic creativity (McCrae, 1987). According to Miller’s (1999; 2000a) cultural courtship model, creative production in artistic domains serves to attract mates, and there is evidence that women find creativity attractive, especially during the most fertile phase of the menstrual cycle (Haselton & Miller, 2006), and that poets and visual artists have higher numbers of sexual partners than controls (Nettle & Clegg, 2006).

The core of openness seems to be a divergent cognitive style that seeks novelty and complexity and makes associations or mappings between apparently disparate domains (McCrae, 1987). Though such a cognitive style might appear purely beneficial, it is conceptually very similar to components of schizotypy, or proneness to psychosis (Green & Williams, 1999; Woody & Claridge, 1977). Indeed, five-factor Openness correlates positively with the Unusual Experiences scale of the Oxford–Liverpool Inventory for Feelings and Experiences schizotypy inventory (Mason, Claridge, & Jackson, 1995; Rawlings & Freeman, 1997). The Unusual Experiences scale is also correlated with measures of creativity (Nettle, in press-b; Schuldberg, 2000).

Individuals scoring high in Unusual Experiences and on measures of creativity have increased levels of paranormal belief (McCrey & Claridge, 2002; Thalbourne, 2000; Thalbourne & Delin, 1994), and five-factor Openness itself is positively correlated with beliefs in the paranormal (Charlton, 2005). The Unusual Experiences trait is elevated in schizophrenia patients (Nettle, in press), and an extremely similar scale predicted the onset of schizophrenia in a longitudinal study (Chapman, Chapman, Kwapił, Eckblad, & Zinser, 1994). Thus, openness and its covariates are associated with damaging psychotic and delusional phenomena as well as high function. Openness itself has been found to be associated with depression (Nowakowska, Strong, Santos, Wang, & Ketter, 2005), as has a high score on the Unusual Experiences scale (Nettle, in press-b). Thus, the unusual thinking style characteristic of openness can lead to nonveridical ideas about the world, from supernatural or paranormal belief systems to the frank break with reality that is psychosis.

What determines whether the outcome of openness is benign or pathological is not fully understood. It may be a
simple matter of degree, or there may be interactions with developmental events. Poets, for example, differ from schizophrenia patients not in their Unusual Experiences scores, which are in the same range, but in the absence of negative symptoms such as anhedonia and social withdrawal (Nettle, in press-b). The Unusual Experiences trait is positively correlated with mating success in nonclinical populations, at least partly because it leads to creativity (Nettle & Clegg, 2006). However, when it leads to schizophrenia, reproductive success is much reduced (Avila et al., 2001; Bassett et al., 1996). Thus the fitness payoffs to openness appear to be very context or condition dependent, leading to the retention of variation.

Conscientiousness

The remaining two personality domains, conscientiousness and agreeableness, are often thought of as being unalloyed in their benefits, because they are generally negatively related to measures of delinquency and antisocial behavior (e.g. Heaven, 1996). However, it is important not to conflate social desirability with positive effects on fitness. Natural selection favors traits that increase reproductive success, including many cases in which this success comes at the expense of other individuals. It is likely that fitness can be enhanced by a capacity to demand a free ride, break rules, and cheat on others under certain circumstances.

Conscientiousness involves orderliness and self-control in the pursuit of goals. A by-product of conscientiousness is that immediate gratification is often delayed in favor of a longer term plan. This leads, for example, to a positive association of conscientiousness with life expectancy (Friedman et al., 1995), which works through adoption of healthy behaviors and avoidance of unhygienic risks. Very high levels of traits related to conscientiousness—moral principle, perfectionism, and self-control—are found in patients with eating disorders and with obsessive-compulsive personality disorder (Austin & Deary, 2000; Claridge & Davis, 2003).

Though some obsessional individuals can be very high achievers in the modern context, it is not evident that their fitness would always have been maximal in a variable and unpredictable ancestral environment. Their extreme self-control not only may be damaging, as their routines become pathological, but may lead to the missing of spontaneous opportunities to enhance reproductive success. Highly conscientious individuals have fewer short-term mating episodes (Schmidt, 2004) and will forgo opportunities to take an immediate return that may be to their advantage. Adaptations that orient the organism toward working for long-term payoffs will tend to have the effect of reducing the opportunistic taking of immediate ones. This can have fitness costs and benefits, which will vary with local conditions.

Agreeableness

Agreeableness, with its correlates of empathy and trust, is also generally seen as beneficial by personality psychologists, and its absence is associated with antisocial personality disorder (Austin & Deary, 2000). Agreeableness is strongly correlated with Baron-Cohen’s empathizing scale (Nettle, in press-a), which is in turn argued to measure theory of mind abilities and the awareness of others’ mental states (Baron-Cohen & Wheelwright, 2004). Several evolutionary psychologists have argued plausibly that as a highly social species, humans have been under strong selection to attend to and track the mental states of others (Byrne & Whiten, 1988; Dunbar, 1996; Humphrey, 1976). Others have noted that we seem to be unique among mammals in the extent of our cooperation with unrelated conspecifics. Inasmuch as agreeableness facilitates these interactions, it would be highly advantageous. Agreeable individuals have harmonious interpersonal interactions and avoid violence and interpersonal hostility (Caprara, Barbaranelli, & Zimbardo, 1996; Heaven, 1996; Suls, Martin, & David, 1998). They are much valued as friends and coalition partners.

Although this may be true, a vast literature in theoretical biology has been devoted to demonstrating that unconditional trust of others is almost never an adaptive strategy. Across a wide variety of conditions, unconditional trusters are invariably outcompeted by defectors or by those whose trust is conditional or selective (see, e.g., Axelrod & Hamilton, 1981; Maynard-Smith, 1982; Trivers, 1971). Levels of aggression can often be selected for (Maynard-Smith, 1982). Very high agreeableness, if it led to an excessive attention to the needs and interests of others, or excessive trusting, would be detrimental to fitness. Among modern executives, agreeableness is negatively related to achieved remuneration and status (Boudreau, Boswell, & Judge, 2001), and creative accomplishment (as distinct from creative potential) is negatively related to agreeableness (King, Walker, & Broyles, 1996).

Though it is an uncomfortable truth to recognize, it is unlikely that fitness is unconditionally maximized by investing energy in positive attention to others. Instead, though an empathic cognitive style may be useful in the whirl of social life, it may have costs in terms of exploitation or inattention to personal fitness gains. Moreover, sociopaths, who are low in agreeableness, may at least sometimes do very well in terms of fitness, especially when they are rare in a population (Mealey, 1995). The balance of advantages between being agreeable and looking after personal interests will obviously vary enormously according to context. For example, in a small isolated group with a limited number of people to interact with and a need for common actions, high agreeableness may be selected for. Larger, looser social formations, or situations in which the environment allows solitary foraging, may select agreeableness downward.

Conclusions

This article has had several purposes. The first has been to stress that heritable variation is ubiquitous in wild populations and therefore should be expected as the normal outcome of evolutionary processes acting on human behavioral tendencies. Thus, personality variation can be understood in the context of a large literature, both theoretical and empirical, on variation in other species.
Second, I have suggested that a fruitful way of looking at variation is in terms of trade-offs of different fitness benefits and costs (summarized in Table 1 for the Big Five personality factors). Theories based on trade-offs have been very successful in providing an understanding of evolution in other species. Moreover, the idea of trade-offs can be usefully married to the notion of fluctuating selection to explain the persistence of diversity. Such accounts are not speculative. Studies such as those on great tits, guppies, finches, and sunfish (see the section on "Evolution of Variation") have demonstrated how fluctuations in environmental context change the fitness outcomes associated with particular phenotypes, which in turn affects the future shape of the population through natural selection. Thus, researchers examining nonhuman variation have been able to go well beyond post hoc explanations and actually observe evolution in action.

The current trade-off account builds on the ideas of MacDonald (1995), who argued that the observed range of variation represents the range of viable human behavioral strategies and who stressed that there are fitness disadvantages at the extremes. Thus, he stressed stabilizing selection. The present argument is that selection can fluctuate, such that it may sometimes be directional for increasing a trait and sometimes be directional for decreasing it. Among the great tits, for example, selection on exploration is clearly directional in any given year (Dingemanse et al., 2004). The retention of a normal distribution is a consequence of the inconsistency of the direction of selection, not its stabilizing form. That said, I agree with MacDonald that there could be quite general disadvantages at the extremes of some personality dimensions, such as chronic depression with high neuroticism, or obsessive–compulsive personality disorder with high conscientiousness. It is not a necessary feature of the current approach that there always be stabilizing effects.

The other major difference between the current approach and that of MacDonald (1995) is that he did not fully develop the notion of trade-offs across the middle range of a continuum, and in particular, he did not develop empirical predictions for the nature of trade-offs for all the different five-factor dimensions.

It is important to stress that trade-offs and fluctuating selection are not the only possible approaches to the maintenance of heritable variation. Biologists have also observed that there are a number of traits that are unidirectionally correlated with fitness and yet in which substantial heritable variation is maintained (Rowe & Houle, 1996). An example would be physical symmetry. In general, the more symmetrical an individual, the higher its fitness, and yet heritable variation in symmetry persists. The maintenance of variation in such cases appears paradoxical, because directional selection might be expected to home in on perfect symmetry and winnow out all variation.

The solution to the paradox appears to be that such global traits as symmetry are affected by mutations to many, if not most, genes. Most mutations that arise are to some extent deleterious, so deviation from physical symmetry becomes an index of the load of mutations an individual is carrying. Selection, particularly that operating via mate choice, favors symmetry, and thus individual deleterious mutations are winnowed from the population. However, so many genes are involved that there is a constant stream of new mutations maintaining population diversity. Thus, symmetry is a fitness indicator trait in that it is a reliable signal of genetic quality.

Some heritable human traits may be better explained by fitness indicator theory than by trade-off theory. Miller (2000b), for example, has applied such reasoning to intelligence. Intelligence is correlated with physical symmetry, suggesting that it taps overall quality (Prokosch, Yeo, & Miller, 2005). Thus, a fitness indicator approach seems likely to be fruitful in such a case.

For personality, however, I suggest that an evolutionary trade-off account is likely to be useful. This does not mean that all personality differences are to be explained by the same mechanism. There are likely to be developmental calibration effects, too, as indicated by behavior genetics data showing a role for the unique environment and also as suggested by recent studies on early life stress and adult

<table>
<thead>
<tr>
<th>Domain</th>
<th>Benefits</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraversion</td>
<td>Mating success; social allies; exploration of environment</td>
<td>Physical risks; family stability</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>Vigilance to dangers; striving and competitiveness</td>
<td>Stress and depression, with interpersonal and health consequences</td>
</tr>
<tr>
<td>Openness</td>
<td>Creativity, with effect on attractiveness</td>
<td>Unusual beliefs; psychosis</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>Attention to long-term fitness benefits; life expectancy and desirable social qualities</td>
<td>Missing of immediate fitness gains; obsessionality; rigidity</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>Attention to mental states of others; harmonious interpersonal relationships; valued coalitional partner</td>
<td>Subject to social cheating; failure to maximize selfish advantage</td>
</tr>
</tbody>
</table>
behavior (Figueroed et al., 2005). However, for the heritable basis of personality, the combination of trade-off and genetic polymorphism seems a fruitful avenue to pursue. It might be objected that the particular costs and benefits put forward here are speculative and as such amount to just-so stories about how personality variation has arisen.

The former is true; as for the latter, such a charge misunderstands the utility of adaptive explanation in psychology. The evolutionary framework used here is hypothesis generating. That is, an article such as this one, which draws on evolutionary biology, is not an end in itself but rather an engine for generating testable empirical ideas. The particular costs and benefits listed here may not turn out to be the correct ones. However, the framework makes testable predictions that would not have been arrived at inductively. For extraversion, the hypothesis that high scorers will have greater numbers of sexual partners but more serious injuries has already been confirmed (Nettle, 2005). For neuroticism, the current framework makes the prediction that performance on certain types of perceptual monitoring tasks, such as detecting an artificial predator, will actually be improved by neuroticism. Because neuroticism impairs performance on many kinds of tasks, this is a novel prediction.

For openness, the model predicts that high scorers will either be socially successful through creative activity or be socially and culturally marginalized through bizarre beliefs, and the determinants of which outcome prevails may depend on overall condition. This is a hypothesis that certainly merits further investigation (see Nettle & Clegg, 2006). For conscientiousness, the model predicts that high-scoring individuals might perform badly on tasks in which they have to respond spontaneously to changes in the affordances of the local environment, because they will be rigidly attached to previously defined goals. Finally, for agreeableness, the theory predicts that high scorers will avoid being victims of interpersonal conflict but may often emerge as suckers in games such as the public goods game and the iterated prisoner’s dilemma game, which are well studied by psychologists and in which the usual equilibrium is a mixture of cooperation and exploitation.

Thus, the current framework should be seen not as a post hoc explanation of the past but as an engine of predictions about the consequences of dispositional variation in the present. Such consequences are a central explanatory concern of personality psychology, and as such, the evolutionary framework, with its emphasis on costs, benefits, and trade-offs, could be of great utility.

REFERENCES


Genetic diversity among human dopamine and serotonin receptors and transporters. *Archives of General Psychiatry*, 57, 1105–1114.


