How Can Evolution and Neuroscience Help Us Understand Moral Capacities?

Randolph M. Nesse

Trying to understand morality has been a central human preoccupation for as far back as human history extends, and for very good reasons. The core phenomenon is readily observable: we humans judge each other's behaviour as right or wrong, and each other's selves as moral or immoral. If others view you as moral, you will thrive in the bosom of a human group. If, however, others view you as immoral, you are in deep trouble; you may even die young, either at the hands of others, or alone in the bush. These are very good reasons indeed for close attention to morality.

There are, however, two problems. The first is how to distinguish right from wrong, the second is inhibiting temptations to do what is wrong (that temptations to do right are not a problem is most interesting). The first problem poses few concerns for most people – they are confident that they know what is right, based on their intuitive emotional responses. However, different people have different emotional responses, and intuition makes a poor argument. Finding a general principle that explains the individual instances would be incredibly valuable. So, for several thousands of years, philosophers have tried to find general moral principles. They have also argued about where they come from, why they have normative force, and how they are best applied to individual instances (Darwall, Gibbard, & Railton, 1997). Thousands of books chronicle the human quest for moral knowledge.

Now, in a mere eye blink of history, the scene has changed. Completely new kinds of knowledge are being brought to bear. Neuroscience is investigating the brain mechanisms involved in moral decisions, moral actions, and responding to moral and immoral actions by self and others. Evolutionary biology is investigating why those brain mechanisms exist, how they give a selective advantage, and why there is genetic variation that influences moral tendencies. This is an exciting time for those of us curious about morality.

R.M. Nesse (✉)
The University of Michigan, Ann Arbor, MI, USA
e-mail: nesse@umich.edu

Before jumping in with new theory and data, however, it is worth stepping back to see how existing explanations fit what we know already. Otherwise, our efforts to help may be received very much like those of a child who holds up his portable video game in a symphony concert and yells excitedly, "Everyone has to try this!" It is impossible to summarize the accomplishments of moral philosophy over even the Western tradition, but three generalizations may be useful. First, general agreement has been reached on some issues. Second, a remarkable amount of disagreement persists. Third, much of the disagreement comes from confusion about what kind of question is at issue, and much of the agreement is about the need to frame questions carefully.

The most fundamental distinction is between descriptive statements and normative statements. One set of questions is descriptive – what are moral capacities and moral behaviors like, and where do they come from? They are about what is. They are, for the most part, science. The other set of questions is normative. They are about what is right and what is wrong. They are about what we ought to do.

The importance of keeping these questions separate is a major recent (as these things go) advance in moral philosophy, attributed usually to David Hume in the 18th century (Hume, 1985 [1740]). Knowing what the world is like cannot directly tell us what is right. Framed in a more familiar form, you cannot get an Ought from an Is. Normative principles about what we should do cannot be derived from knowledge about what the world is like, not even from complete scientific knowledge about the brain and how its moral capacities were shaped by natural selection. (Subtle arguments apply, of course, but those are for another time).

The attempt to derive normative principles from facts is often called the naturalistic fallacy, but that is not exactly accurate. GE Moore, in his 1903 book *Principia Ethica* (Moore, 1903), defined the naturalistic fallacy as attempts to prove a moral claim based on a definition of "good" based on natural properties such as what is desirable or pleasant. Moore's point is that "good" cannot be defined in such terms, because it is not an object in the natural world. This is obviously closely related to the Is-Ought barrier, so, like others, I will use the phrase "naturalistic fallacy" in the more general sense.

While nearly every article on evolution and morals notes the importance of avoiding the naturalistic fallacy, many then go ahead and nonetheless draw moral guidance from observations about the world. Or, if they don't, readers do, often with blissful naïveté. To observe the phenomenon for yourself, explain the naturalistic fallacy to undergraduate students, then explain how natural selection has shaped male mammals to compete to get as many matings as possible, and open the discussion. The fact of increased competition among males leads some students to conclude that such behavior is natural and therefore right. Others will disagree, and the resulting animated exchanges demonstrate how people draw normative implications from the most abstract principles in biology, all warnings notwithstanding.

The naturalistic fallacy is, for humans, remarkably natural. For instance, when an evolutionary biologist describes forced mating as a potential adaptive strategy, he can warn against the naturalistic fallacy page after page, but few readers will even

notice; most will be outraged because they think he approves of rape. His protests that he said no such thing will be ignored.

While such readers may be illogical, their reactions reflect important human tendencies. They intuitively recognize that describing a human behavior as a "normal" adaptation will lead many to conclude that the behavior is right, or at least not wrong. For instance, some young men, upon learning that natural selection has shaped organisms to maximize reproduction, change their personal sexual behaviour dramatically. I have also observed several people who changed their behavior after learning that relationships are mutually beneficial exchanges in which the maximum payoff goes those who can best deceive others. They changed not only their view of relationships, but their actual relationships; previously secure close partnerships became much more difficult. And then, of course, there is the dire history of people leaping from the fact of natural selection to justifications for eugenics and even genocide. Human views about the origins and functions of the moral capacities have tangible effects on behavior. Caution is warranted.

With this background, we can ask what neuroscience and evolution offer to understanding morality and immorality. The simultaneous focus in this book on psychopathy and normal moral capacities is particularly useful. Medicine has consistently found it difficult to explain abnormal conditions until both the normal mechanisms and their functions are clear. Conversely, studies of pathology often offer the best evidence about the functional significance of a trait. If you want to know what the thyroid gland is for, clues come from observing what happens when the thyroid gland is missing or malfunctioning. If you want to understand the benefits of moral capacities, study individuals who lack them.

Understanding the moral capacities requires two kinds of knowledge, evolutionary and proximate. They address fundamentally different questions (Dewsbury, 1999; Tinbergen, 1963). A proximate explanation is about how a mechanism works. Neuroscience offers proximate explanations at a low level. Psychology offers proximate explanations about morality at a higher level. Evolutionary explanations are different. They address why the mechanisms exist at all, in terms of selection and other evolutionary forces that account for the mechanism being the way it is. This is usually described as "the function" of a trait (although that often turns out to be too simple). Proximate and evolutionary investigations can inform each other, but they are about different questions.

Evolution

An evolutionary explanation of how moral capacities have increased fitness is the essential foundation for understanding morality. For the purposes of this book, the most important conclusion is that this foundation is still under construction. I have written much about this (Nesse, 2006), others have devoted their lives to it (Hammerstein, 2003; Katz, 2000) and this volume contains a comprehensive review (van Veelen in this volume). While much remains to be done, a rough framework is in place.

As most readers will know, naïve group selection seemed sufficient to explain altruism and morality until Williams pointed out its deficiencies (Williams, 1966), and Hamilton (Hamilton, 1964) and Trivers (Trivers, 1971) offered alternative explanations of kin selection and reciprocity, respectively. A vast amount of research since then has framed a general solution (Hammerstein, 2003). In very broad brush strokes, the vast majority of cooperative behavior in animals can be explained by kin selection or mutualisms. Well-documented examples of reciprocity in animals turn out to be rare (Stevens, Cushman, & Hauser, 2005), with the exception of our species. For humans, trading favors is at the center of life. We have emotions shaped to cope with the situations that routinely arise in reciprocity relationships (Nesse, 1990). Extraordinary social institutions enforce agreements, thus allowing vast social complexity.

Despite dozens of issues still on the table, a general explanation for moral capacities that facilitate exchange relationships is within reach. Skill in managing such relationships brings a net gain, and so selection should shape tendencies to do what works. That means paying close attention to who you are dealing with, and it usually means following rules. Explaining altruism beyond reciprocity and kinship (for instance, helping a dying friend), is more difficult. One approach is to explain such altruism away as self-deception or mistakes, another is to attribute it to group selection or cultural influences. I have previously argued that a capacity for commitment (in the economic game-theory sense of the word) can shape capacities for communal relationships that explain some aspects of altruism and our moral capacities. I still think that is important, but it leaves much unexplained.

Finally, I found articles by Mary Jane West-Eberhard (West-Eberhard, 1979; West-Eberhard, 1983) that offered a new perspective. In the late nineteen seventies she discovered that extraordinary social traits can result from the same kind of runaway selection that shape extraordinary sexually dimorphic traits, like peacock tails. The only difference is that the fitness benefits come not from being chosen as a mate, but from being chosen as a social partner. In the course of evolutionary history, once personal relationships began yielding a selective advantage, individuals who chose better partners began to gain an advantage. However, getting the best partners is not merely a matter of choosing, it depends more on being preferred as a partner. Individuals therefore display resources they can offer to their partners, and personal and moral characteristics that make them desirable partners, such as generosity and honesty. My paper, reprinted in this volume, unites this fundamental idea with modern mathematical models and findings from human social science, to argue that runaway selection based on partner choice can explain human moral and other social capacities that are otherwise inexplicable (Nesse, 2007).

One particularly interesting aspect, not developed in the article, is how social selection can explain why we value certain diverse personal characteristics as "virtue." Recent sophisticated social science methods have revisited and confirmed long-recognized virtues such as bravery, creativity, wisdom, persistence, integrity, vitality, love, kindness, social intelligence, fairness, forgiveness, humility, gratitude, hope, and humor (Peterson & Seligman, 2004). How do all of these characteristics come together to be recognized as virtues? I suspect it is because they are the

characteristics we want in our partners. And, because we want to be valued as partners, they are also the traits we seek to display, and to live up to. They are products of social selection. Their unity may arise, not from a grand unifying philosophical principle, but from their origins in social selection for tendencies to choose, and to be, excellent relationship partners.

Evolution and Pathology

The framework of Darwinian medicine can be useful to analyze the evolutionary origins of presumably abnormal states, such as antisocial personality disorder (Nesse & Williams, 1994). Is it a disorder created only in modern environments, a product of infection, a tradeoff, a constraint, or is it an adaptation? The first question is whether it has had some kind of utility, but it is essential to keep all the alternative hypotheses on the table and not to jump to one conclusion.

Linda Mealey has argued that sociopathy might be a frequency dependent alternative strategy that gives a selective advantage when it is rare (Mealey, 1995). She notes a variety of supportive evidence, including the additional matings garnered by some psychopaths, however it is important to note several reasons why the hypothesis is not widely accepted.

Most alternative strategies are mating strategies, such as those used by dominant and subordinate male orangutans. Other morphs, such as benthic and limnic morphologies in fish, are alternatives for living in different ecological niches. These alternative strategies need not be associated with genetic differences. For instance, many fish change sex depending on the social environment. Such facultative adaptations, shaped by natural selection, monitor the environment and express one body type or another, depending on the situation. Locusts change from solitary to swarming morphs depending on the circumstances. The difference need not be categorical; early exposure to heat in infancy increases the number of sweat glands.

These principles may be useful for understanding antisocial personality disorder. It is hard to see the benefits of a genetically determined psychopath behavioral morph, when more flexible regulation would be more efficient. Why be stuck playing one strategy when flexibility is superior? Furthermore, substantial evidence now shows that genetic tendencies to sociopathy are not deterministic, they interact with early events to cause the disorder in some individuals. For instance, Caspi et al. found that rates of conduct disorder and criminal conviction increased dramatically with exposure to child abuse in all genotypes, but the increase was greater in those with low MAOA activity (Caspi et al., 2002). This suggests considering antisocial personality as an alternative strategy that emerges depending on early experience.

A more basic question is whether antisocial personality disorder is one condition, multiple related conditions, or positions on a continuum. Experience talking with psychopaths reveals their enormous diversity. Some use violence to get what they want. Others never use violence, but manipulate others, and get their greatest

pleasure from deceiving others, not from what they get out of the deception. Others have extraordinary seduction skills that allow them to get what they want, and then abscond. Others are simply socially incompetent; they cannot manage relationships, and they flounder in all kinds of ways. And, of course, others are simply shrewd political manipulators who become powerful leaders. Despite this observed variation, a review of five studies of antisocial personality disorder, finds four that support a view of antisocial personality as a distinct category instead of a dimension (Haslam, 2003).

This brings up the question of why such diverse characteristics as low empathy, impulsivity, using violence, and inability to have close relationships should occur together so reliably. The most obvious possibility is that all aspects arise from a common proximate cause, perhaps impulsiveness, or a low ability to learn from punishment. Another possibility is that they occur together because they are all useful aspects of a strategy for social influence. This could reflect a frequency dependent genetic strategy, as Mealey proposed, but there are several alternatives. They could also occur together as parts of a facultative adaptation that emerges in response to certain early experiences. However, they could also merely reflect what people fall back on when some defect in the cognitive and emotional apparatus makes normal social relationships impossible. Etiological heterogeneity is likely (Silberg, Rutter, Tracy, Maes, & Eaves, 2007) and there is no need to posit just one kind of deficit. Some might lack empathy, others may be unable to learn from punishment, others may simply be too impulsive to be reliable relationship partners, others may simply believe that others are untrustworthy. If a person is, for whatever reason, unable to create and benefit from ordinary social relationships, he or she will fall back to simpler strategies. Early experience with using violence and deception to manipulate others soon results, by simple learning, in more and more effective psychopaths who are locked into one strategy of social influence. The coherence of the syndrome may arise not from within, but from doing what works when you are incapable of maintaining and benefitting from enduring social relationships.

Neuroscience

Several chapters in this book demonstrate that moral decision making and moral emotions arise from brain mechanisms. Of course, we have long known that this had to be true. Nonetheless, because we are all innate dualists (Bloom, 2004), the simple fact can still seem shocking.

The next task is to find out what parts of the brain carry out moral tasks. Is one locus specialized for the task? Is there a circumscribed module to take care of it? As has been the case for other capacities, from language, to pain, to emotion, moral tasks are not processed in any one location. They may even be handled in different loci by different people. Several chapters in this book take on the challenge of trying to discover where the brain processes moral information. They amply demonstrate that these tasks are not carried out everywhere, but also that they are not carried out by a specific locus; they are carried out in diverse regions that are hard to specify.

It is disappointing that we cannot point to one brain locus, and say "morality happens here." However, what we observe is exactly what an evolutionary perspective leads us to expect. A few very specific responses, such as vomiting and panic, have functions so specific and universally essential that they have been conserved for tens of millions of years; they have specific loci devoted largely to managing their expression. Equally old tasks that are not so tightly constrained, such as balance, are more distributed. New tasks, such as language, have been grafted onto existing structures in whatever way works, resulting in a hodge-podge of loci and connections.

This evolutionary view is very different from the massive modularity that is often associated with evolutionary psychology (De Schrijver in this volume). From my perspective, it is important to recognize that specific kinds of situations have posed the adaptive challenges that shaped brains with capacities for moral reasoning. The resulting mechanisms deal effectively with those situations. However, this by no means implies that selection would shape separate mechanisms to deal with each situation. On the contrary, we should expect modules with massive overlap in their arousal, processing, and output, and in the brain loci that mediate their functioning.

Moral capacities are very recent and very nonspecific. They require input of many kinds from many sources and outputs to many effectors. Selection shaped them by acting on relevant variation wherever it was available to co-opt old structures to new uses. For instance, disgust is ancient and has obvious adaptive utility – it motivates avoidance of pathogens (Curtis & Biran, 2001). Individuals with even a slight tendency to experience disgust after betrayal would avoid the betrayer. In just a few tens of thousand more years, it should not be surprising that moral violations arouse the same brain areas aroused by disgust (Greene & Haidt, 2002). Similarly, see Chapter 2 about how communal relationships and the moral emotions that sustain them seem to be have arisen from the mother infant attachment system (Moll & Oliveira-Souza in this volume). Both examples seem likely to be correct. They illustrate how evolutionary thinking can help inhibit tendencies to think of the brain as a machine with newly minted modules for each challenge, and how it can help us to accept the messy reality of functions carried out by multiple interconnected multifunctional loci.

Conclusion

Evolutionary analyses to understand the origins of moral capacities are coming along, but no one thesis is dominant at this point except for the general conclusion that natural selection has shaped capacities for coping with the situations that arise in reciprocity relationships, and additional moral capacities that make communal relationships possible.

Even at this stage, however, an evolutionary perspective can help to guide neuroscience research about antisocial personality disorder by encouraging attention

to how new functions have been grafted onto structures with multiple other functions, and attention to likely constraints that make such systems vulnerable to failure. Moral capacities are evolutionarily new and almost completely restricted to humans. This means that the substantial genetic variation in moral traits may best be explained because the phenotype is in transition. A related possibility is that optimum may vary markedly across groups and times. Even in a stable setting, fitness may be about the same across a wide range of the distribution.

What trait? Social selection shapes extraordinary concern about what others think about us, and motivations to please others. The benefits of such tendencies may explain why social anxiety disorders are vastly more common than psychopathy. The costs of excess social sensitivity are, however, large. It seems entirely possible that reproductive success will be roughly the same for individuals across a wide range of the distribution. Our expectation that there is some sharp peak that defines "normal" may be incorrect. This does not sit well with our human wish to define categories and declare some normal and some abnormal. However, it may reflect a more realistic view that can help us better understand morality and immorality.

References

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science, 297*(5582), 851–854.

Curtis, V., & Biran, A. (2001). Dirt, disgust, and disease. Is hygiene in our genes? *Perspectives in Biology and Medicine, 44*(1), 17–31.

Darwall, S. L., Gibbard, A., & Railton, P. A. (1997). *Moral discourse and practice: Some philosophical approaches*. New York: Oxford University Press.

Dewsbury, D. A. (1999). The proximate and the ultimate: Past, present and future. *Behavioural Process, 46*, 189–199.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517–523.

Hamilton, W. D. (1964). The genetical evolution of social behavior i, and ii. *Journal of Theoretical Biology, 7*, 1–52.

Hammerstein, P. (2003). *Genetic and cultural evolution of cooperation*. Cambridge, MA: MIT Press in Cooperation with Dahlem University Press.

Haslam, N. (2003). The dimensional view of personality disorders: A review of the taxometric evidence. *Clinical Psychology Review, 23*(1), 75–93.

Hume, D. (1985 [1740]). *An treatise of human nature*. London: Penguin Classics.

Katz, L. (2000). *Evolutionary origins of morality : Cross disciplinary perspectives*. Devon: Imprint Academic.

Mealey, L. (1995). Sociopathy. *Behavioral and Brain Sciences, 18*(3), 523–599.

Moore, G. E. (1903). *Principia ethica*. Cambridge, UK: Cambridge University Press.

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature, 1*(3), 261–289.

Nesse, R. M. (2006). Why so many people with selfish genes are pretty nice-except for their hatred of *the selfish gene*. In A. Grafen & M. Ridley (Eds.), *Richard dawkins* (pp. 203–212). London: Oxford University Press.

Nesse, R. M. (2007). Runaway social selection for displays of partner value and altruism. *Biological Theory, 2*(2), 143–155.

Nesse, R. M., & Williams, G. C. (1994). *Why we get sick: The new science of Darwinian medicine*. New York: Vintage Books.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press.

Silberg, J. L., Rutter, M., Tracy, K., Maes, H. H., & Eaves, L. (2007). Etiological heterogeneity in the development of antisocial behavior: The Virginia twin study of adolescent behavioral development and the young adult follow-up. *Psychological Medicine, 37*(8), 1193.

Stevens, J. R., Cushman, F. A., & Hauser, M. D. (2005). Evolving the psychological mechanisms for cooperation. *Annual Review of Ecology, Evolution, and Systematics, 36*(1), 499–518.

Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*, 410–463.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35–57.

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

West-Eberhard, M. J. (1983). Sexual selection, social competition, and speciation. *Quarterly Review of Biology, 58*(2), 155–183.

Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, N.J.: Princeton University Press.