

Darwin and the Golden Rule

L. Karl Branting

6 July 2008

Opening Reading

If science lost us our Western paradise, our place at the center of the world ... if we have been left adrift near the edge of just another humdrum galaxy, perhaps it is time to take heartened stock of our situation... Perhaps we are at home in the universe in ways we have not known ... I am heartened by a view of evolution as a marriage of spontaneous order and natural selection. I am heartened by the possibility that organisms are not contraptions piled on contraptions all the way down, but expressions of a deeper order inherent in all life. ... I am heartened by the possibility that our social institutions evolved as an expression of deep natural principles...

We have only begun to invent the science that will account for the evolving emergent order..., from a spider weaving her web, to a coyote crafty on the ridge top ... to all of you making your ways by your own best efforts and own best lights. ... We are all part of this process, created by it, creating it. In the beginning was the Word—the Law. The rest follows, and we participate.

Stuart Kauffman¹

Presentation

Two of the most important forces in contemporary life are *science* and *religion*. Each is a powerful lens for viewing both the world and ourselves. Ideally, scientific and religious views should complement one another or, at least be consistent. However, we live in an era in which the full implications of science for our understanding of ourselves, our lives, and our place in the universe haven't yet been completely absorbed into most people's religious thinking. The friction between science and religion is particularly intense at the point at which the theory of evolution rubs up against religious doctrines on the origins of human beings and the meaning and purposes of our lives.

This morning I would like to talk about a specific aspect of this friction point: the challenge of understanding how natural selection could produce moral beings. This talk will consist of four 'P's: some *preliminaries*, a *puzzle*, a *partial solution*, and a *pitfall*.

Let's start with the preliminaries: what is religion, what is the theory of evolution, and do they conflict?

Religion is many different things to different people, but for most people religion contains three key elements:

- Transcendence – connection to things more important than one's day-to-day concerns and lasting longer than one's individual life
- Community – membership in a group united by shared beliefs, values, and practices
- Morality – an imperative to treat others in a compassionate, fair, and charitable manner

¹ Stuart Kauffman, "At Home in the Universe: The Search for the Laws of Self-Organization and Complexity", Oxford University Press (1995).

The theory of evolution is based on the insights of Charles Darwin and Alfred Russel Wallace that the diversity and development of life can be explained from just three assumptions:

- Traits can be inherited
- Traits have random variations
- Some traits lead to better chances of survival and reproduction than others.

If these three requirements are satisfied, then more adaptive traits will tend to displace less adaptive traits. Since the days of Darwin and Wallace, we have learned that heritable traits are associated with genes, which are segments of DNA molecules. This theory provides the best explanation of where we came from and why we are so similar in anatomy and behavior to other animals.

Is there a contradiction between the theory of evolution and religion? Some religious fundamentalists reject the Darwinian model because they believe that it contradicts the creation stories of their religion. This threatens the transcendent element of their religion. However, more sophisticated theists generally recognize that evolution could just be a deity's procedure for creating us. For example, the late Pope John Paul II stated that there is “no opposition between evolution and the doctrine of the faith about man and his vocation.”² The Dalai Lama has said of science and religion “There is no contradiction between the two. Each gives us valuable insights into the other. Both science and the teachings of the Buddha tell us of the fundamental unity of all things.”³ Apparently, many Hindus believe that the sequence of Avatars of Vishnu corresponds to stages of evolution. And of course for a religious humanist like myself, who believes in nothing supernatural, evolution doesn't contradict my creation myth, it *is* my creation myth. So, the theory of evolution needn't conflict with the first element of religions, transcendence. But what about the third element, morality?

This question is my segue to the second 'P', the *puzzle*: if natural selection inevitably rewards genes that lead to increased likelihood of survival and reproduction, how can genes that promote altruism ever become more prevalent than genes for selfish behavior?

Let me be clear about what I mean by altruism. Compassion, fairness, and morality all mean valuing, and sometimes subordinating oneself to, the interests of others. We deeply esteem altruism, both in ourselves and in others, and we try to transmit it to our children. We glorify those who sacrifice themselves for their family, friends, community, or nation, even if we sometimes have doubts about our own capacity for such self-sacrifice, and we revile cheaters who exploit others for their own benefit.

So the puzzle is this: If we are the product of natural selection, and natural selection inevitably selects whatever traits favor survival and reproduction, why shouldn't genes for self-sacrifice reduce the survival and reproduction chances of organisms that carry them and therefore die out? For example, if whenever a community is threatened by hostile animals or neighbors the altruistic men go forth to a glorious death and their faithful wives nobly waste away in sorrow, children will tend to be born to cowardly, malingering men and unfaithful women.

Another way to look at this puzzle is to consider the reaction to the publication of Richard Dawkins' book *The Selfish Gene*⁴. Many readers reacted with horror to the book's contention that all living things can be viewed as DNA containers engineered by blind evolution to maximize the likelihood of the DNA's duplication:

2 Pope John Paul II, Message delivered to the Pontifical Academy of Sciences, 22 October 1996.

3 Dalai Lama Tenzin Gyatso, Nobel Prize acceptance speech University Aula, Oslo, 10 December 1989.

4 Richard Dawkins, “*The Selfish Gene*,” Oxford University Press 3rd ed. (1989).

A teacher from a distant country wrote to me reproachfully that a pupil had come to him in tears after reading the same book, because it persuaded her that life was empty and purposeless. He advised her not to show the book to any of her friends, for fear of contaminating them with the same nihilistic pessimism.⁵

So, does evolution prove that altruism is a fake and that beneath the surface we are actually totally selfish DNA containers? Well, people certainly can be selfish, but there is nothing unusual about genuine self-sacrifice. To choose an obvious example, there have been numerous instances of soldiers in Iraq and Afghanistan who sacrificed their lives for their comrades and their communities. Smaller acts of altruism are a vital part of life both for those who perform them and those receive them, from helping a stranger, to comforting a child, to performing an anonymous act that makes the world a little bit better place. Donating blood, leaving a big tip in a restaurant in a city you won't visit any time soon, and giving a ride to a stranger whose car has broken down are all altruistic actions.

Science has much to say about the role of morality and fairness in human behavior. For example, the field of behavioral economics has produced extensive research on this topic. One popular approach to exploring the role of fairness in human behavior experimentally is through such tools as the Ultimatum Game, in which:

one participant gets a large sum of money to divide between himself and another participant, and the second one can take it or leave it. If he leaves it, neither side gets anything. A selfish proposer would keep the lion's share; a selfish respondent would accept the remaining crumbs, no matter how small, because part of a loaf is better than none. In reality, the proposer tends to offer almost half of the total sum, and the respondent doesn't settle for much less than half, even though turning down a smaller share is an act of spite that deprives both participants.⁶

For example, suppose an experimenter gives me \$100, and I give Chuck \$20 and keep \$80 for myself. If Chuck is like most people, he will say, "I'd rather get nothing than go along an your unfair division." In the United States, most respondents won't accept less than around \$40, and most proposers offer offer at least \$40.

In hundred of experiments Ultimatum Game participants have shown themselves willing to pay to punish cheaters rather than see cheaters get more than their fair share. A recent article in Science magazine summarizes a long line of behavioral research across multiple cultures, stating that "[s]ubstantial fractions of most populations adhere to moral rules, willingly give to others, and punish those who offend standards of appropriate behavior, even at a cost to themselves and with no expectation of material reward."⁷ Industrial and pre-industrial people, theists and atheists, all share similar moral intuitions about sharing, fairness and altruism. In summary, *our genes may be selfish, but we are not.*

So if altruism is real, how did it evolve?

This question leads us to the third 'P', a partial solution.

Recent research has suggested at least 3 ways that altruistic behavior can evolve.

- a. Kinship
- b. Reciprocity

5 Richard Dawkins, "Unweaving the Rainbow," Houghton Mifflin Co., New York, New York (1998).

6 Steven Pinker: The Blank Slate: The Modern Denial of Human Nature. New York: Penguin, 2003, pg. 256.

7 Samuel Bowles, "Policies designed for self-interested citizens may undermine "the moral sentiments": evidence form economic experiments, Science, vol.320, 20 June 2008, pp1605-1609.

c. Commitment

A. Kinship.

Bee hives and ant hills are models of altruism. Each worker bee or ant will unhesitatingly sacrifice itself for the benefit of the hive or anthill. This is the easiest form of altruism to explain in Darwinian terms because it consists of altruism to kin. Each worker is a sibling of every other worker and the offspring of the queen. From the standpoint of the worker's genes, every worker is equivalent to any other worker, and the most effective way for a worker to insure the transmission of its genes is to preserve and nurture the queen. Self-sacrifice is therefore completely consistent with the Darwinian model.

Does similar behavior occur in humans? Absolutely. Parents are typically much more willing to make sacrifices for their own children than for nieces or nephews, and more willing to make sacrifices for nieces and nephews than for step-children. There are societies in which loyalties are primarily to kinship groups rather than to colleagues, friends, or the larger community.

However, this is not generally true of western industrial societies. In societies like our own, loyalties are generally not limited to relatives. Moreover, many of the varieties of altruistic behavior that I mentioned earlier can't be explained by kinship. Even primates form close bonds with non-relatives. Frans de Waal, in his book on primate altruism "Good Natured"⁸ describes the reaction of a female chimpanzee's friends when she gave birth. The friends became increasingly excited, and screamed loudly when the birth occurred. They then hugged the mother and cradled the newborn.

In summary, some altruistic behavior can be explained by kinship, but much altruistic behavior cannot.

B. Reciprocity

Social scientists have long been fascinated by a formal model of interactions between pairs of individuals called the Prisoner's Dilemma. A typical example of the prisoner's dilemma occurs when two merchants arrange a trade in which one agrees to leave a bag of money under a rock in one part of the forest, while the other agrees to leave some goods under a rock in a different part of the forest. Both profit from the trade if both cooperate. If one merchant cheats and leaves nothing and the other plays fair, the profit for the cheating merchant is even higher. But if both cheat, they both lose the cost in time and effort of going through the forest. The dilemma is that, regardless of what the other merchant does, each merchant gets a higher profit from cheating than from playing fair, and yet if both follow this same reasoning and cheat, both will end up with nothing.

The Prisoner's Dilemma is designed so that the "rational" strategy, that is, the strategy with the highest expected return, is to cheat if there is just a single transaction. Of course, in the real world, we often interact with the same people more than once. The problem of devising the best strategy when there are multiple transactions with the same person is known as the Iterated Prisoners Dilemma.

In the 1970's a social scientist named Robert Axelrod held a computer tournament in which he elicited strategies for the Iterated Prisoners Dilemma.⁹ Each of the submissions played against all the other submissions in a computer simulation of many repetitions of the Iterated Prisoners Dilemma. Many of the strategies were fiendishly clever, yet the winning strategy was the simplest. It was called Tit-for-tat, and it consisted of cooperating in the first round and thereafter copying

8 Frans de Waal, *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge and London: Harvard University Press, 1996.

9 Axelrod, Robert. (1984). *The Evolution of Cooperation*. New York: Basic Books.

the opponent's last move. This strategy could be summarized as "play fair, but retaliate proportionately if cheated."

Axelrod published the results of the tournament and several years later solicited submissions for a new tournament. This time, he attempted to simulate evolution by rewarding winning strategies with increased reproduction. Even though the contestants had several years to devise strategies against Tit-for-tat, it was once again the winner. The reason that Tit-for-tat wins is that it cooperates with other cooperative strategies. Cheats and exploiters do well at first, taking advantage of patsies who allow themselves to be exploited. When the patsies become extinct, however, the cheats can't exploit either each other or Tit-for-tat, so they gradually lose out.

In my previous career as a computer science professor, I duplicated this experiment in classes on two occasions, getting strategies from students, then simulating evolution for hundreds or thousands of generations. Cooperative strategies like Tit-for-tat always prevailed over both exploiters and patsies. It's very satisfying when science confirms the truths that our parents always told us: play fair, but don't let yourself be exploited.

Tit-for-tat-like behavior occurs in real life. In the trenches during WWI, troops who faced each other for long periods of time developed conventions to avoid hurting each other. Tit-for-tat behavior also occurs in animals. A successful vampire bat will share part of its blood with a less-successful cave-mates with whom it has a lengthy relationship. Certain birds will remove ticks from other birds who reciprocate. If another bird fails to reciprocate, however, the cheated bird will never groom it again.

Is Tit-for-tat a model of morality? Only of a very rudimentary kind. Tit-for-tat illustrates the advantages of cooperating in sustained reciprocal arrangements. It shows how subordinating one's interests to those of a partner can be beneficial in arrangements in which such subordination will be reciprocated. However, the iterated prisoners dilemma doesn't appear to be adequate for describing altruism to strangers or subordination of one's interests to a group.

C. Commitment.

Robert Frank in his book "Passions Within Reason"¹⁰ analyzes a way that that altruism to strangers or to a group can occur. Sometimes making a *commitment* to a certain behavior can be beneficial even if the behavior itself is not beneficial. Suppose that I make a commitment to protect my community. This commitment might force me to sacrifice my life to protect people I am not related to. However, the commitment may open opportunities that would otherwise never exist. For example, I may be able to marry a desirable spouse or be trusted in business opportunities that would be unavailable if I were unwilling to make such a sincere commitment. Similarly, making a commitment to fairness and responsibility towards others may lead me to others to help strangers, pass up opportunities to exploit the helpless, or leave large tips in restaurants I will never visit again. These actions are not directly beneficial, but the commitment can lead to cooperation and trust from others even if I interact with them just once.

There are negative commitments as well. In Dr. Strangelove, Soviet Premier Dmitri Kisov devises a doomsday machine that will end all human life if the Soviet Union is ever attacked. Although ending the world is clearly not beneficial to Kisov, Kisov *is* benefited if everyone knows about the doomsday machine, because everyone will have an extremely strong incentive not to attack. Once again, a commitment can produce benefits even if the behavior committed to is not beneficial.

Of course, it is a long way from describing the commitment model to showing that it constitutes a way for altruistic behavior towards strangers and groups to evolve. For example, there should be a

10 Robert Frank, *Passions Within Reason*, W.W. Norton & Co. 1989.

very strong selective pressure in favor of faking commitments. However, experiments reveal that people have an amazing ability to detect the genuineness of commitments.

Imagine that you have left \$1000 in an envelope with your name and address on it in a crowded theater. Of all the people you know, are there some who you think would be more likely to return the envelope if they found it? Of course there are. So we distinguish among our acquaintances according to how much they can be trusted. Experiments have shown that if strangers are given just 30 minutes to meet the partners with whom they will play the Prisoner's Dilemma, they are remarkably good at predicting which of the strangers will cheat and which will cooperate. Ironically, Frank's research has shown that "students who have been taught ... neo-classical economics are much more likely to [cheat] in prisoner's dilemma games than, for instance, astronomy students." Frank's commitment model shows how committing yourself to fairness, and morality is the best strategy in the long run. Acting solely on the basis of rational calculation of self-interest is, paradoxically, not in our best interest.

In summary, altruism based on kinship, reciprocity, and commitment can be adaptive and therefore the result of natural selection. The research that I have described is only the beginnings of a full explanation of the relationship between evolution and our moral sense, but it is certainly sufficient to dispel the misconception that Darwin and Golden Rule are incompatible and that one or the other has to go.

This leads to the final 'P', a pitfall that can arise in trying reconcile evolution with morality. This pitfall was named the *naturalistic fallacy* by the British philosopher G.E. Moore, but was first identified a century and a half earlier by David Hume. Hume observed that in discussions of ethics, writers often start by discussing what IS, but at some stage of the discussion suddenly begin talking about what OUGHT to be. At first glance this seems just to be a comment about language, but actually it expresses a deep point: no amount of factual statements can every, in themselves, add up to a moral statement; no amount of IS statements can every, in themselves, justify an OUGHT statement. The view that one can justify what OUGHT to be from what IS is the Naturalistic Fallacy.

In the context of evolution, the naturalistic fallacy is that because natural selection IS the process that produced human beings, natural selection therefore OUGHT to be encouraged. Herbert Spenser coined the phrase "survival of the fittest" and argued that trying to improve the lives of the least fortunate interfered with the process of evolution and therefore was contrary to the greater good of the human species. Social Darwinism—the idea that the existing distribution of wealth reflects the greater good of human kind—has, naturally, always been much more popular with the wealthy than the poor. Social Darwinism had a particularly sinister application in the genocidal pseudo-science of Hitler and Stalin.

Another example of the naturalistic fallacy is to conclude from the fact that men and women have had different roles in many societies throughout history, that men and women SHOULD have different RIGHTS. In reality, our commitment to the equal dignity and worth of every human being is based on what we are today as moral agents and does not depend on how we evolved or on the choices that others made in the past.

In summary, I have described to you the puzzle that natural selection guided by selfish genes managed to produce human beings that, at their best, are capable of compassion, generosity, fairness, and self-sacrifice. I have described ongoing research by evolutionary and behavioral scientists that provides an initial, partial explanation of how this happened, and I pointed out the pitfall of concluding that because something is the case, it therefore should be the case.

The essence of morality is compassion, treating others as we should wish to be treated ourselves, and

acting in a manner that brings the greatest good to the greatest number. Understanding how evolution can produce morality should only increase our reverence and awe for the universe that has imbued us with our capacity for love, creativity, fairness, and compassion. So be it.

Closing words:

A human being is part of a whole, called by us the Universe, a part limited in time and space. We experience ourselves, our thoughts and feelings, as something separated from the rest—a kind of optical illusion of our consciousness. This illusion is a kind of prison for us, restricting us to our personal desires and to affection for a few persons nearest us. Our task must be to free ourselves from this prison by widening our circles of compassion to embrace all living creatures and the whole of nature in its beauty.¹¹

- Albert Einstein

¹¹ <http://www.spaceandmotion.com/Albert-Einstein-Quotes.htm>.