The Evolution of Cooperation: Darwin’s Unsolved Problem and its Relevance to Environmental Concerns.¹

Robert M. May
British Science Festival Presidential Address of 9 July, 2009.

Page 1.

Darwin’s Three Big Problems

Page 2.

An iconic event in popular accounts of the history of ideas about evolution and the origin of species is the Huxley-Wilberforce debate in 1860. The Bishop of Oxford, “Soapy Sam” Wilberforce, was a gifted, though markedly self-satisfied, orator. Speaking first, he unwisely mocked Huxley by asking “is it through grandmother or grandfather that you descend from a monkey?” This opened the door to Huxley’s memorable riposte: “would I rather have a miserable ape for a grandfather, or a man highly endowed by nature and possessed of great means and influence, and yet who employs these faculties and that influence for the mere purpose of introducing ridicule into a grave scientific discussion – I unhesitatingly affirm my preference for the ape”. After that, the meeting dissolved into Pythonesque[2] chaos. First, Fitzroy (who had captained the Beagle, gone on to become the Governor of New Zealand, but by 1860 was mentally unstable, and soon afterwards committed suicide) came raving down the aisle, brandishing a bible, and demanding that all return to “The Book”. The meeting ended as Lady Brewster[3] – in the words of one of Darwin’s biographers, DeBeer – “employing an idiom now lost, expressed her sense of intellectual crisis by fainting”.[4]

What most accounts of these events overlook, however, is the fact that Wilberforce, had he possessed a thorough knowledge of the science of his day – had he been better briefed – could have won the debate. [5] The Darwin-Wallace theory of evolution, at that time, had three huge problems.

---

¹ Excerpted by W. M. Schaffer. Material in square brackets [ ], including footnotes, added.

² The reference is to the comedian, Monty Python.

³ Wife of physicist, David Brewster, best known for his work on polarization of light.


⁵ This is something of a stretch. The problems to which May alludes, paucities of variation and time, respectively raised by Fleeming Jenkin and William Thomson (later Lord Kelvin), did not become part of the evolutionary debate until several years later.
The first problem concerned the time available for evolutionary processes to operate. Fifty years were to elapse before the first glimmers of awareness of weak and strong nuclear forces were to appear. Of the four fundamental forces recognised by today’s physics, only gravitational and electromagnetic ("chemical") forces were known in Darwin’s day. But if the sun’s energy source was gravitational, it could not have been burning for more than about 20 million years. And chemical fuels would give an even shorter life. A different calculation showed that it could not have taken more than roughly 20-40 million years for the earth to cool from molten rock to its present temperature. These two calculations meant that either the earth was at most a few tens of millions of years old, or that Victorian physics was fundamentally deficient. Faced with these arguments, Darwin removed all numerical references to geological time spans in the third and later editions of the *Origin of Species*, and you will look in vain for any explicit chronology in the later *Descent of Man*. Of course, subsequent discovery of nuclear forces showed Victorian physics was indeed inadequate: the sun burns nuclear fuel; and the heat generated by decay of radioactive elements inside the earth invalidates simplistic calculations about cooling rates.\textsuperscript{6} We now understand that evolutionary processes on earth have all the time they need.

The second problem stemmed from the conventional wisdom of the day, namely that inheritance worked by a blending of maternal and paternal characters. The essentials of this issue can be grasped by considering a trait (such as height or weight) that can be described by a single variable. Suppose

the mother departs from the population average in this respect by an amount $x$ and the father by an amount of $y$. Then, under a scheme of blending inheritance, the progeny will depart from the mean by $\frac{1}{2} (x + y)$. ... It is then straightforward to show that, with blending inheritance, the variance of this trait in the next generation is halved. But persisting variability is the raw stuff upon which natural selection works to produce descent with modification; it was critical to Darwin’s ideas. This fundamental difficulty was pointed out to him, most notably by the engineer Fleeming Jenkin. He [Darwin] acknowledged it as a problem, but – given the observed persistence of variability in natural populations – he simply put it aside. The resolution of this major difficulty lies, of course, in the fact genes are inherited in particulate Mendelian fashion, not by “blending”. And, as shown in 1908 independently by Hardy and by Weinberg, under Mendelian inheritance variability remains unchanged from generation to generation, unless perturbed by factors such as selection, mutation, statistical drift, or nonrandom mating.

In short, these first two of Darwin’s three truly major difficulties have been entirely swept away by advances in our understanding of the natural world.

\textsuperscript{6} More recently it has been argued that the heat generated by radioactive decay, if added to Thomson’s calculation, would have had little effect on his estimate. A more fundamental error was his assumption that heat from the earth’s interior is transported to the surface by conduction. Today, we know that a substantial amount of heat is transported convectively – think plate tectonics.
Darwin’s third major unsolved problem, which he himself arguably saw as the most important, is not yet solved. This problem was, and still is, explaining how cooperative or altruistic behaviour among animals evolved.

Page 5.
At first glance, the answer seems easy. You pay some small cost to gather a much larger cooperative benefit. For example, a prairie dog takes a personal risk in giving an alarm call, but all the colony benefits and, by taking turns as alarm giver, each individual’s group benefit exceeds the occasional risk. But any such arrangement is immediately vulnerable to cheats who enjoy the benefits without paying the risk-taking dues. In evolutionary terms, such risk-avoiding cheats have a selective advantage. Today we would say their enhanced probability of survival, and consequent greater reproductive success, means their uncooperative behaviour is more represented in the next generation (possibly via their genes, or alternatively by teaching their offspring – Dawkin’s memes). It is thus unclear how such observed cooperative phenomena can arise, or if it does how it can be maintained. Following work on “kin selection” by Hamilton and others, a century after Darwin, we now understand how such cooperative associations can evolve and persist in relatively small groups of sufficiently closely related individuals. This would seem to solve the problem for many non-human groups of animals, which are indeed found in such small kin groups. In particular, haplo-diploid systems of genetic inheritance, where siblings share more genes than do parents with their offspring, further facilitate such kin selection, which can help explain apparent altruism among some social insects.

Page 6.
During the hundred-thousand years and more when humans existed as small bands of hunter-gatherers, such considerations of kin selection could well have promoted cooperative behaviour. But for large aggregations of essentially unrelated individuals, as developed once agriculture appeared some ten millennia ago and cities began, the origin of cooperative associations – with group benefits which exceed the “cost of membership” – remains almost as puzzling today as it was for Darwin.

Page 10.
The Evolution of Cooperation
The failure to attain a satisfactory resolution to what might be called “Darwin’s Third Problem” is not for lack of effort. The problem has received increasing attention in the scholarly literature in recent years, employing a variety of metaphors: for evolutionary biologists, the Prisoner’s Dilemma; for ecologists, the Tragedy of the Commons; for economists, the Free-Rider problem. These metaphors are allied to artificial games in which the subjects (usually undergraduates) trade small sums of money to test limits to altruism and tolerance of cheating. ...

Page 11
Particularly interesting, in my opinion, are experiments in which each of the players start with a given sum (say 10 players each starting with 20 currency units). There now follow 10 rounds, in
each of which a player can choose to put 0, 1 or 2 units in to a common pool. At the end of each round, the accumulated size of the pool is announced. If, after 10 rounds, the pool equals (or exceeds) 100 units, all players keep the currency units they did not put into the pool. But if the pool fails to attain 100 units, no-one keeps anything.

Obviously the cooperative, “fair” strategy is for each player to contribute one unit in every round. What tended to happen in these experiments, however, was that some players initially “cheated” by contributing nothing, so that the pool initially grew at less than the linear trajectory to 100 required. At this intermediate point, most players then contributed, but not enough to catch up. In the final rounds, “oversubscription” of 2 units tended to accelerate pool growth, but nevertheless in more than half the runs the target was not achieved, and everybody lost. This pattern invites gloomy parallels with the world’s current efforts to address climate change.

A variety of recent experiments include the possibility of “punishment”. This is done by showing what each player did in a previous round, and allowing any player – at some modest cost (i.e. deduction of some currency units) – to inflict a (larger) deduction of units upon another player. The idea here is to provide an incentive to “play fair”, by punishing “cheats”, although the enforcement mechanism is itself not cost-free; in effect it requires a degree of altruism toward the group from the punisher. These experiments do, in general, tend to entrain everyone into “fair play” and achievement of the group target. Even so, the players who do best are those that avoid the cost of punishing “cheats”, thus retaining elements of the Free Rider problem. One notable set of such studies of “costly punishment”, by Herrmann et al, compared the results obtained from groups of players, each group drawn from one of 16 different cities: Boston, Melbourne, Nottingham quickly converged on cooperative behaviour; at the opposite end of the spectrum, Minsk, Samara, Riyadh were slower, with significant levels of “cheating”, resulting “punishing” and quite a bit of perverse “punishing punishers”. At the foot of the table, in Athens and Muscat, many games descended into madness, with angry “punished” players retaliating rather than cooperating, locking into spirals of mutual punishment.

Against this academic and analytic background, my own musings about how cooperative human societies may have evolved begins with the observation that many of the problems of “costly punishment” could be avoided if the enforcer were some omnipotent deity. This has parallels with Dominic Johnson’s thoughtful recent essay on supernatural punishment and cooperation.

Once we move out of the mists of pre-history, we find stories of dreamtime, creation myths, ceremonies and initiation rites, spirits and gods, with a unifying theme that all seek simultaneously to help explain the external world and also to provide a “stabilization matrix” for a cohe-

---

sive society. There are, moreover, some striking and unexplained similarities in belief systems and rituals from different times and places. Conscience, a simple word for a complex

Page 13.
concept which helps foster behaviour in accord with society’s professed norms, has been memorably defined by H. L. Mencken as “the inner voice which warns us that somebody might be looking”. And how helpful it is if the “punisher” of departures from societal “norms” is an abstract but all-seeing, all-knowing supernatural entity.

Common to these conjectured “stabilizing forces” in essentially all earlier societies are hierarchical structures, whose representatives serve and interpret – and derive their power from – the divine being or pantheon. In such systems, there is unquestioning respect for authority. Faith trumps evidence.

But if indeed this is broadly the explanation for how cooperative behaviour has evolved and been maintained in human societies, it could be very Bad News. Because although such authoritarian systems seem to be good at preserving social coherence and an orderly society, they are, by the same token, not good at adapting to change.

We saw earlier that the solution to Darwin’s Second Problem, “blending inheritance”, was provided by the Hardy-Weinberg Law. This invites comparison with Newton’s First Law in physics, which in effect says if nothing is changing, a body’s dynamical behaviour does not change (it stays at rest or in uniform motion in a straight line). Newton’s Second Law describes what happens when things do change, that is when forces act. This Second Law of physics also has its analogue in Fisher’s Fundamental Theorem, a basic principle emerging from the

Page 14.
Neo-Darwinian Revolution of around a century ago, which states that a population’s potential rate of change of gene frequency (which measures its ability to adapt to changing circumstances) is proportional to the variance in gene frequency, which will usually be small if essentially all individuals are especially well-adapted to their current environment. That is, there is an inherent tension between adaptedness and adaptability.

If there is any substance in my speculations about the answer to Darwin’s problem in explaining cooperation in human societies, we again have a fundamental tension – at the level of the entire society – between on the one hand authoritarian “ties that bind” and permit stably cooperative aggregations, and on the other hand the ability to respond effectively to changing environmental circumstances. ...