MORAL PHILOSOPHY AND THE DARWINIAN PROBLEM OF SOCIAL EVOLUTION

MORAL PHILOSOPHY AND THE DARWINIAN PROBLEM OF SOCIAL EVOLUTION

By YUSSIF YAKUBU, B.Sc. [Hons], Dip. Ed, M.Sc., M.A.

A Thesis Submitted to the School of Graduate Studies in Partial
Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University

McMaster University DOCTOR OF PHILOSOPHY (2015) Hamilton, Ontario

(Philosophy)

TITLE: Moral Philosophy and the Darwinian Problem of Social Evolution

AUTHOR: Yussif Yakubu, B.Sc. [Hons], Dip. Ed, (University of Cape Coast),

M.Sc., (University of Guelph), M.A. (McMaster University)

SUPERVISOR: Professor Richard T. W. Arthur

NUMBER OF PAGES: vii, 195

ABSTRACT

Social behaviour is common in nature. Yet, for over a century and half, scholars have struggled in vain to offer a satisfactory account of its evolution under Darwinian natural selection. In this thesis I propose that three fundamental assumptions in the Darwinian explanation of social behaviour are at the root of the problem. They are: 1. The basic Darwinian philosophy that evolutionary change occurs by one inherent trait replacing another in an organism. 2. The collapse of social behaviour, in its entirety, into a single, narrow concept called "altruism." 3. The assumption that such "altruism" arises from a mutation at a single gene locus, where it supplants "selfishness" as an alternative allele.

The thesis identifies some insights from Hume's analysis of human morality and sociality that suggest the proper circumstances of social interactions in humans. We see from Hume's analysis that nothing inherent in human nature needs to change in order to move beyond parenting to sociality. Hume identifies two principles in human nature — selfishness and empathy — that are the ultimate basis of human sociality. Empathy expands self-interest to include relatives and associates, but not strangers. And that suffices to form small, primitive human societies. For large, cosmopolitan societies, Hume suggests they are maintained only through human inventions such as governments and justice. Hume's explanation precludes the need for a weaker "altruistic gene" to supplant a

fitter “selfish gene” as a condition for social evolution, which has been the basis of the Darwinian explanatory difficulty.

Acknowledgements

My greatest gratitude is owed to my supervisor, Dr. Richard T. W. Arthur, who never wavered in his pledge to ensure I become a philosopher. Dr. Rama Singh has the special status of being my gateway to McMaster and has continued to support me ever since. Dr. Barry Allen has caught some key inconsistencies in the preliminary drafts of the thesis, made very important constructive criticisms and above all, offered very extensive text editing. I will also like to thank my wife, Evelina, for her patience and encouragement during this process.

TABLE OF CONTENTS

Chapter Four. Modeling Modern Society

Overview

The thesis deals with the broad topic of the Darwinian problem of social evolution. Social behaviour is common in nature, including humans. However, there exists a worry (Darwin 1859; 1871; Hamilton 1964) that natural selection, the most widely accepted evolutionary explanation today and for the past century and half, may not be able to account for such cooperative behaviour within groups. The motivation for this thesis is that I am confident that there is an insight from moral philosophy which, if taken into consideration in current biological modeling, could contribute significantly towards a resolution of the theoretical difficulty.

The theory of evolution postulates the transmutation of organisms from one species to another. Darwin’s theory of Natural Selection postulates that this transmutation occurs through the supplanting of existing traits by new ones, which occurs in imperceptibly small but numerous steps. By Darwinian logic, a stable evolutionary change occurs where the new trait confers a heritable fitness advantage on its bearers. Social behaviour, as perceived by Darwinians, appears to violate this fundamental Darwinian logic because it is a trait that appears to put its bearers at a fitness disadvantage, but is nevertheless prevalent in nature. Consequently, it is considered to be a puzzle under Darwinian theory of evolution.

In his Inclusive Fitness theory, Hamilton (1964) suggested that the Darwinian puzzle may be resolved if we expand our conception of fitness to include that conferred directly on the organism by the trait and an additional (inclusive) fitness it confers on the organism through its genetic relatives. This proposal by Hamilton tends to limit social behaviour to groups consisting only of genetic relatives. However, social groups tend to include genetically unrelated individuals. Consequently, Trivers (1971) proposed the theory of Reciprocal Altruism to explain social behaviour in groups of unrelated individuals. Today, five decades after these modern Darwinian theories were proposed and the myriad of models of social evolution that have been spawned under them, evolutionary biologists are not convinced they have satisfactorily explained the evolution of social behaviour. What has Darwinian explanation got wrong and how fundamental is the error?

Hume's (1740) analysis of human sociality and morality suggests that the fundamental Darwinian philosophy that evolution occurs by the supplanting of one heritable trait by another may not be applicable to social behaviour at all. There is already some softening of this fundamental Darwinian evolutionary philosophy, particularly among social scientists who model on Evolutionary Game theory. They suggest that some social behaviour may be culturally transmitted through learning and imitation. The problem with those models, however, is that they tend to describe social behaviour in existing social groups and when they venture to explain the actual mechanism of social evolution, they tend to fall back on the Darwinian trait supplanting philosophy.

In order to show how Hume's insight might help contemporary evolutionary explanation, however, some groundwork needs to be done. First, I contextualize the problem within the general practice of science and the frequent occurrence of such theoretical anomalies. Using Popper's and Kuhn's views on scientific theories, together with the Aristotelian cosmological paradigm, I show how such theoretical anomalies arise in science, and how resolutions have often come about. I follow that with a sketch of the evolution of the Darwinian explanation, and show that it fits that general pattern of paradigm development in science, of which puzzles are a regular feature. That is the subject of chapter I.

In chapter two I examine the Darwinian explanatory paradigm more closely. I identify the features of it that I believe are at the root of the explanatory difficulty with social evolution. The first of the two principal ones is that, under contemporary modeling of social evolution, social behaviour in its entirety is subsumed under a single concept called "altruism," which is described as "other-regarding" behaviour. More specifically, it is a behaviour performed at a cost to the agent, but which benefits some other organism(s). The second modeling conception of social behaviour that I believe to be another part of the puzzle is that this generalized trait called "altruism," is assumed to arise from a mutation at a single gene locus, which then competes with "selfishness" as an alternative allele, in order to establish itself in a population.

Under this modeling assumption, there arises what Wilson and Wilson (2007, 336) have described as the "the fundamental problem of social life," which is that, "in virtually all cases, traits labeled cooperative and altruistic are selectively disadvantageous within the groups." (335). Consequently, as Cavalli-Sforza and Feldman (1978, 268) assert, "the theory of the evolution of <u>genetically determined</u> altruistic behavior faces, at the outset, the central problem that altruists are, by definition, at greater risk … of elimination by natural selection." The problem is that evolutionary biologists have not fully appreciated the key role the two assumptions I have identified play in the puzzle of social evolution. Instead, they have tried to resolve the puzzle with those assumptions intact. Little do they realize that the hypotheses they propose to resolve the anomaly actually undermine those core assumptions they hold regarding social behaviour. I suggest in this thesis that those two principal assumptions ought to be renounced, as they are completely erroneous depictions of social behaviour.

In chapter three, I discuss Hume's thesis on morality and social evolution. For this, I rely mostly on his *Treatise on Human Nature*. Hume's main objective in that work is to establish the pre-eminence of emotion in human moral and social life. So Hume sets out principally to debunk the long standing view of a battle between reason and emotion, in which a victory of reason is a mark of civility and moral rightness. It is the view of morality Kant will later come to be the best known proponent of. For Hume, every moral decision is a choice between competing emotions rather than a choice between emotion and reason. That does not mean reason plays no role in moral judgments, but its role is not to override the emotions as the moral rationalists hold. Reason, according to Hume, only gives us a “cause and effect” analysis that enables us to see the consequences or outcome for each of the competing emotions we may choose to act upon, but reason does not actually dictate which of them we should choose or override any that we may choose to act upon. As an analogy, sight presents to us the appearance of each apple in a bunch from which we want to pick the best apples. But that is the extent of its role. Sight does not tell us which appearance we should desire and which we should not. That is in our nature and sight does not override or change that. In the same way, reason does not oppose or override our passions.

Hume argues, in fact, that “we always consider the usual and natural force of the passions when we determine concerning virtue and vice” (*Treatise*, 483). So not only do we not submit our passions to the control of anything other than the passions themselves, we actually determine the virtue of any action based on whether or not it is consistent with the usual force of the passions, while vice is action that deviates from the usual force of the passions.

Now, if all our moral sense conforms to the natural force of our passions, how does that inform our quest to explain social behaviour? Hume identifies two principles in human nature that are relevant to social behaviour. They are our inherent selfishness and empathy. Empathy expands our self-interest to include the interests of some other individuals besides ourselves. In fact, Leibniz, for example, held a broader view of self-interest that included individuals covered by our empathy. That is why he thought self-interest does not oppose the formation of society. Hume on the other hand, notes one critical feature of empathy that puts it in opposition to the formation of society. It is the fact that empathy is not an open, universal or non-discriminatory affection. It is highest for family members, less so for other relatives, friends, acquaintances and associates, and virtually non-existent for strangers. That is why Hume refers to it by the synonymous terms, “partiality” and “limited generosity.” Hume emphasizes this narrowness of empathy as a critical opposing force to the formation of society.

But Leibniz is not completely wrong. By promoting mutual self-interest amongst certain individuals, empathy may indeed foster social cooperation. However, given its narrowness, empathy cannot do so for groups more sophisticated than the small tribal bands of pre-industrial peoples, as they are the groups that usually do not include strangers. Similarly, I argue in this thesis, cooperation within such small primitive groups is what Darwinian models, if they are ever successful, may explain. However, for Hume, those primitive groups are not societies. They are seed from which society will evolve when they accumulate a critical mass of wealth, but until then they are not society. By society, Hume means political society, which by nature is large and includes unrelated strangers.

If political society exists, in spite of the “selfishness and limited generosity” of our nature, it is not because we have evolved some other trait to oppose or replace those that oppose the formation of society. This is contrary to the modern Darwinian approach, in which the primary goal is to show how “altruism” evolves to replace selfishness as the inherent human nature in order to make social cooperation possible. For Hume, no level of human social cooperation is sustained by replacing our natural selfishness and partiality by some other human nature. The same selfishness and partiality keep small groups together. Large groups (political society) are kept together by human artifice, particularly, justice and government.

The stage for this particular analysis is set in chapter three and brought to a head in chapter four.

Chapter One

Puzzle Solving: Analysis versus Synthesis

I. Introduction

The puzzle of social behaviour is one of the most enduring in the history of science. In this chapter, I propose an explanation as to why certain theoretical anomalies, including that of social behaviour, endure for so long. Such a general understanding should give us some insight into why the puzzle of social evolution has been so intractable, and the right way to approach the problem. To provide the broader context in which the problem of scientific puzzles is situated, I will synthesize a picture of scientific practice from the two orthogonal views proposed by Karl Popper and Thomas Kuhn. I use the evolution of our theories about the cosmos to illustrate the picture of scientific practice that I present. The centrepiece in this part of the discussion is the Aristotelian cosmological paradigm and the prolonged difficulty it had with planetary retrograde motion. The analysis reveals that the reason why such puzzles endure is that they often arise from the errors that occur where we least expect them, i.e. axiomatic assumptions deep in the core of the paradigm. Identifying such erroneous axioms requires teasing apart the paradigm and examining all of its assumptions, definitions and reasoning. That would be the analytic approach to puzzle solving. As the discussion here shows however, the normal practice of science, what Kuhn (1962) calls "normal science," consists in the articulation and building of incomplete paradigms which largely is a process of synthesis. The analytic approach to puzzle solving is precluded in the practice of “normal science” because the preoccupation of scientists, within a paradigm, is working to improve the fit between the paradigm and the observational data, rather than a continual re-evaluation of the axioms and assumptions of the paradigm.

My thought is that scientific progress is best characterized neither as the falsification of theories nor problem solving within paradigms, or successions of paradigms, which are the competing positions of Karl Popper and Thomas Kuhn respectively.[1] Rather, as it appears to me, our understanding and explanation of nature improves (which indeed is scientific progress) as we discover and replace erroneous assumptions we hold about nature. A theory, as well as a paradigm, often consists of several assumptions; and replacing an assumption does not necessarily bring down a theory or paradigm. For the purposes of this thesis, I suggest what specific assumptions need to be replaced in the Darwinian explanatory framework for social evolution to become solvent.

II. Scientific Practice and the Scientific Ideal

In order to understand why the problem of social evolution has endured for so long, we need to step back and look at it in the broader context of scientific theories, their nature and the attitude of scientific practitioners towards anomalous data. Two entrenched, standard views on the nature and practice of

---

[1] I am mindful here that a paradigm is a much broader concept than a theory.

science were proposed in the mid twentieth century, and are still considered popular. They are Popper's (1959) proposed criteria of *demarcation* between science and pseudoscience, and Kuhn's (1962) *paradigm* conception of scientific theories. The former prescribes the demarcating criteria between science and pseudoscience, which the later shows to be largely idealistic by rendering a descriptive account that portrays a gulf between actual scientific practice and the normative standards Popper advocates.

The scientific approach to discovery and knowledge has always been held to be distinct and more reliable than other approaches. Karl Popper was the first to articulate quite precisely what it is that sets science apart from other approaches to inquiry, and how scientific knowledge acquires its higher reliability. He was of the view that the prevalent notion that science sets itself apart from non-scientific epistemology by its "empirical method, which is essentially inductive," was insufficient, as it allowed certain theories which were pseudoscientific to masquerade as genuine scientific theories. Popper felt that there was something unscientific about theories such as Karl Marx's *theory of history*, Sigmund Freud's *psychoanalysis* and Alfred Adler's *individual psychology,* which were all claiming to be scientific. Popper was concerned that these theories seemed to find corroboration in all kinds of data, including contradictory sets. As he notes, admirers of such theories "saw confirmed instances everywhere. … Whatever happened always confirmed [them]" (1963, 24).

This observation led Popper to the conclusion that corroboration is not a sound criterion for distinguishing between science and pseudo-science. He notes that “it is easy to obtain confirmations or verifications for nearly every theory – if we look for confirmations” (1963, 25). If a theory allows no possible way to show that it is wrong, it is not a good scientific theory, Popper thought. Instead, he suggested that “every good scientific theory is a prohibition: It forbids certain things to happen. The more a theory forbids, the better it is” (1963, 26). A scientific theory thus, must make bold, unequivocal claims or predictions, which can be tested with relevant empirical data. Popper’s position summarily, then, is that a theory is scientific if it is formulated in such a way as to generate predictions that are testable, falsifiable or refutable. He deplores instances in which some genuinely testable theories are maintained by their admirers in the face of falsifying instances. This is done often “by introducing ad hoc some auxiliary assumptions, or by reinterpreting the theory ad hoc in such a way that it escapes refutation” (1963, 26). This final observation (or caution) will prove to be problematic for Popper’s prescriptive account of scientific theories, as we shall see. For contrary to what Popper may have thought, that practice would prove to be the norm rather than the exception in science.

Testability and falsifiability were long recognized as essential features of scientific hypotheses well before Popper (see for example Whewell (1847/1967) and Peirce (1931/1858)). However, it was Popper who first articulated them as criteria for the demarcation between science and pseudoscience. That seemed to strike a chord with popular sentiment and many do take the Popperian tenets as the rule of thumb for demarcating between science and pseudoscience. For example, in one of the many court battles in the United States over mandating the inclusion of the so-called *creation science* in the high school science curriculum along with the Darwinian theory of evolution (McLean v. Arkansas Board of Education, 1982), the judge, in ruling that *Creation Science* is not science, accepted expert testimony that (Ruse 2009, 14):

> The ‘essential characteristics’ of what makes something scientific are:
>
> 1. It is guided by natural law;
> 2. It has to be explanatory by reference to natural law;
> 3. It is testable against the empirical world;
> 4. Its conclusions are tentative, i.e. are not necessarily the final word; and
> 5. It is falsifiable.

As we can see, the last three out of the five criteria are Popperian. In fact, in Ruse’s account, falsifiability was instrumental in determining the case. Ruse reports that “the judge accepted that evolutionary thinking is falsifiable,” and Creation Science “is not falsifiable and hence not genuine science” (2009, 15-16).

However, not all scholars think such a sharp line of demarcation is possible between science and non-science, particularly not falsifiability. Ruse (2009, 16) recounts that there was strong objection from some prominent scholars to the invocation of falsificationism in the trial. Even more damning to the falsificationist strategy of the anti-creationism expert team was the fact that Karl Popper (1974) himself – the architect of the falsifiability test – did not think Darwinian natural selection itself met the falsifiability requirement. In fact, Ruse, who was on the anti-Creationism expert team in the trial, admits that the anti-creationists did not have satisfactory responses to the objections from scholars, other than to say, rather disingenuously (his own characterization), that the constitutional challenge was not about “the teaching of false science. It bars the teaching of non-science, especially non-science which is religion by another name.” Ruse is content that for the purpose of the legal case at hand, the strategy was “good enough.” Entailed in this response is the concession to the critics that Darwinian natural selection (one of the most respected scientific theories) may not pass the falsifiability test. This note leads me into the descriptive accounts of scientific theories, and the gulf it exposes between scientific practice and the normative ideals set by Popper.

III. The Descriptive View of Science

One of the most comprehensive and best known descriptive accounts of the practice of science was rendered by Thomas Kuhn (1962) in his now classic book *The Structure of Scientific Revolutions*. Kuhn demonstrates, contrary to the prevalent view, that the practice of “normal science” involves not the discovery of new theories, but the articulation of paradigms. Kuhn’s descriptive account of actual scientific practice paints a picture that is quite contrary to what Popper perceived science to be. Kuhn explains that even though scientific theories do fall, they do so not on account of single falsifying instances. In fact, Kuhn (1962/1996, 77) declares, in direct contradiction of Popper, that “no process yet disclosed by the historical study of scientific development at all resembles the methodological stereotype of falsification by direct comparison with nature.” He adds (1962/1996, 77), contrary to Popper, that one thing “that scientists never do when confronted by even severe and prolonged anomalies (falsifying data)” is renounce the paradigm. Instead, he observes, “they will devise numerous articulations and ad hoc modifications of their theory in order to eliminate any apparent conflict” (1962/1996, 78). It is the very thing Popper suggests scientists ought not to do. However, it is true that we never throw a theory out simply on account of counter-instances and then operate in a theoretical vacuum. Usually we hang on to it in the face of anomalies until a better explanation comes up to replace it. As Gould (1996, 351-352) notes, “the barrel of theory is always full” and “science advances primarily by replacement, not by addition.” Gould adds (1996, 352) that scientists do not debunk only to cleanse and purge. They refute older ideas in the light of a different view about the nature of things. As Kuhn noted, “there is no such thing as research without any paradigm, and to reject one paradigm without simultaneously substituting another is to reject science itself” (1962/1996, 79).

In Kuhn’s view, the preoccupation of normal science is puzzle solving, and those puzzles are supplied by the “incompleteness and imperfection of existing data-theory fit” (1962/1996, 146). Thus, there are always anomalous data and were such anomalous data grounds for theory rejection, “all theories ought to be rejected at all times,” according to Kuhn (1962/1996, 79).

Kuhn may have succeeded in showing that scientific theories are not rejected in the falsificationist fashion proposed by Popper, and the “paradigm” model he proposes in its place may be quite tenable, but the historical record seems to suggest that the paradigm boundaries may not be as distinct as Kuhn suggests. As Kuhn’s own account shows, paradigm changes generally involve changes to no more than one or two assumptions or suppositions, with the bulk of the assumptions unchanged between paradigms. As a matter of fact, the primary activity of paradigm practitioners – puzzle solving – actually consists in replacing those carried over assumptions that are thought to be responsible for observational anomalies. But these are usually the peripheral ones, while the core assumptions that define the paradigm are sacrosanct, as Lakatos (1978) noted. The goal of these activities is to improve the fit between the paradigm and the observational data. Kuhn describes the practice as “mopping up” activities, and suggests that it is the main driver of scientific progress. Therefore, regardless of whether we see scientific progress as falsifications and refutations, or successions of paradigms, the kind of change that is undisputed is the replacement of assumptions and suppositions, and this, I see, is the source of explanatory progress in science.

Kuhn for example identifies Copernican and Ptolemaic astronomies as contiguous paradigms in succession. However, it is equally conceivable and in fact more expedient to view them as alternative models under the *Aristotelian Cosmological Paradigm*, whose core axiom was the *Principle of Uniform Motion* that was proclaimed by Aristotle. In fact, Copernicus did not see his model as anything more than an alternative arrangement of the planets that would rectify some of the anomalies facing the principle of uniform motion (Commentariolus, 57-58). The heliocentric model of Copernicus was not in itself a novel proposition, and Copernicus himself noted that (*Revolutions* 1543). Also, it did not realize its objective of resolving the anomalies. As Kuhn himself noted, “Copernicus's system, for example, was not more accurate than Ptolemy's until drastically revised by Kepler more than sixty years after Copernicus's death” (Kuhn 1973, 357).

What Kepler did, in fact, represented the actual shift away from the fundamental paradigm that dated back to Aristotle. I say this for a number of reasons. The first is that the core assumption of Aristotelian cosmology was the *principle of uniform motion*, which held that the planets have circular orbits and uniform motion. Copernicus is on record as defending that principle and severely castigating scholars who opposed such handed down wisdom. Secondly, Copernicus made it clear that the cosmological views he was proposing, however radical they may appear to us today, were aimed at preserving those Aristotelian axioms. In the third place, Copernicus’s insistence that the orbits of the planets are their real paths in space (and not simply their angular positions as observed from Earth), together with his heliocentric arrangement of the planets, were notable departures from the Ptolemaic system. However, those were not sufficient to resolve the puzzles of the Aristotelian system with the principle of uniform motion, the central dogma of Aristotelian cosmology, intact. Enter Kepler, and he, with the first two of his celestial laws, overturned Aristotle's principle of uniform motion, which governed the practice of all prior planetary scientists. In fact, Copernicus viewed his own work, including his heliocentric proposal, as part of (to speak in Kuhnian lingo) the "articulation" and "mopping up" of the Aristotelian paradigm. His heliocentric model was proposed as a desperate radical attempt to counter models that he thought blatantly violated the principle of uniform motion. It appeared under those arrangements of spheres, he writes, "that a planet moved with a uniform velocity neither on its deferent nor about the centre of its epicycle" ((*Commentariolus*, 57). Copernicus was troubled by this, saying that "a system of this sort seemed neither sufficiently absolute nor sufficiently pleasing to the mind" ((*Commentariolus*, 57), thus, motivating his proposal (*Commentariolus*, 57-58) :

> Having become aware of these defects, I often considered whether there could perhaps be found a more reasonable arrangement of circles, from which every apparent inequality would be derived and in which everything would move uniformly about its proper center, as the rule of absolute motion requires.

The business of experimenting with alternative arrangements of the celestial spheres was the ordinary puzzle solving activity within the Aristotelian paradigm and antiquity. By the time of Copernicus, however, the Ptolemaic arrangement had become widely accepted. In order to soften the blow of his contravention of the Ptolemaic system, Copernicus, in his letter dedicating his *De Revolutionibus* to Pope Paul III, pointed out that alternative arrangements of the heavenly spheres had always been permitted. He cited several past scholars whose planetary systems involve the motions of the earth. Then he adds:

> Therefore, having obtained the opportunity from these sources, I too began to consider the mobility of the earth. And even though the idea seemed absurd, nevertheless I knew that others before me had been granted the freedom to imagine any circles whatever for the purpose of explaining the heavenly phenomena. Hence I thought that I too would be readily permitted to ascertain whether explanations sounder than those of my predecessors could be found for the revolution of the celestial spheres on the assumption of some motion of the earth.

For Copernicus, success would have meant an arrangement that preserved the circularity and uniformity of the motion of the planets as required under the Aristotelian paradigm. There is the issue of whether the orbits of the planets are their actual paths in space or simply as their angular positions as viewed from earth. However, that was not a core principle of the paradigm. The defining and immutable principle of the Aristotelian paradigm was uniform motion. Any model that went contrary to that would have constituted a departure from the Aristotelian paradigm. I assert in fact, that Copernicus would have rejected Kepler's improvement of his model for the same reason that he rejected Ptolemy's and the other models of his time, i.e. violations of the principle of uniform motion. All these go to illustrate that drawing paradigm boundaries is quite arbitrary and subjective. What is uncontroversial and quite obvious is that axioms, suppositions and conceptions upon which theories are built often turn out to be erroneous, and such erroneous assumptions have in fact been the chief source of the mismatch of theory and empirical data, which according to Kuhn (1962), is the source of the puzzles scientists strive to solve in the practice of “normal science.” In my view therefore, the best approach to solving theoretical puzzles in science is analyzing theories to identify and eliminate erroneous assumptions that engender the anomalies. However, such an *analytic approach* has not been the practice, contrary to our common illusion. The analysis below of the way puzzles are dealt with under “normal science,” as rendered by Kuhn, shows it to be anything but analytical. This is on top of the fact that the way scientists interpret research data, as Kuhn (1962) and Feyerabend (2008) have noted, is already theory laden.

IV. Analysis versus Synthesis

In the disagreement between Popper and Kuhn, one thing they took for granted is the common view of a scientific theory/paradigm as a unity, to be rejected (Popper) or protected (Kuhn) in the event of anomalous data. As shown in the discussion above, Popper recommends an outright rejection of the theory rather than engaging in what he calls *ad hoc stratagems*, which he explains as “introducing ad hoc some auxiliary assumptions, or by reinterpreting the theory ad hoc in such a way that it escapes refutation” (1963, 26). Kuhn on the contrary, reports that in reality what Popper forbids is actually what happens. In other words, protecting scientific theories via *ad hoc stratagems* is the substantive activity of “normal science.” It consists in the main of adding auxiliary assumptions or hypotheses to plug the holes that leak puzzles into the theory. Thus, it is a synthetic approach to solving puzzles, in the sense that, rather than analyzing the paradigm for faulty components or faulty assumptions to be thrown out, it looks outside the paradigm (most certainly, outside the core assumptions of the paradigm (Lakatos 1978)) for what needs to be corrected in order to eliminate the anomaly. And as the following examples will illustrate, the practitioners of a paradigm that is faced with a particular anomaly will tend to propose, prior to any empirical evidence, a view of the world that might eliminate the anomaly.

Such additional assumptions and hypotheses, the so-called *ad hoc hypotheses*, serve to prop up the theory or paradigm and keep it intact. This synthetic approach to puzzles via ad hoc hypothesis is opposed to the analytic approach to theoretical anomalies which I shall put forward later on. In the analytical approach to investigating theoretical anomalies, the components of the theoretical paradigm (axioms, assumptions, logic, etc.) are each examined for fit with observation. Adherents of a paradigm are very reluctant to adopt this approach, in spite of all pretentions of scientific objectivity. For, as Kuhn (1962) explains, they are in the business of articulating the paradigm, rather than investigating and judging it. In other words, they accept the paradigm and work to make it the best it can be. That is why any alterations to the paradigm by its practitioners are superficial, because renouncing the core principles of the paradigm would mean rejecting it. Those who take the analytic approach to investigating an anomaly are often taken to be critics of the theory, especially when they point out faults.

To illustrate the synthetic approach to puzzle solving, consider the Aristotelian astronomical paradigm when it faced the puzzle of planetary retrograde motion. Ptolemy, Copernicus and others within the Aristotelian worldview responded with ad hoc hypotheses, most notably epicycles, which they introduced to resolve apparent retrograde motion or angular positions. Aside from epicycles, however, the core puzzle solving activity was experimentation with different arrangements of the celestial spheres. Even though most modern text books identify a single geocentric system associated with Ptolemy and an alternative heliocentric model associated with Copernicus, there were actually other variants of either system that were proposed by different astronomers.[1] The important thing to note about the Aristotelian astronomers is that none of their puzzle solving activities, i.e. introducing epicycles or rearranging the planets, involved challenges to any of the core assumptions or axioms of the Aristotelian system itself. Instead, they sought to construct reality in a way that would fit the paradigm by experimenting with different arrangements of the celestial spheres.

---

[1] Copernicus himself (Revolutions, 1543) cites Cicero and Plutarch as having written about earlier astronomers who held such views.

Secondly, by postulating epicycles they were seeking to reinterpret the observational data to fit the axioms of the paradigm.

It is interesting to note that this historical fact is quite inconsistent with Popper’s view of “ad hoc stratagems.” It is true that a defining feature of the synthetic approach to puzzle solving is the introduction of additional assumptions to prop up the theory. However, as the Aristotelian example illustrates, the reinterpretations are not of the theory to fit the observational data, but rather of the observational data to fit the theory. We have to note that a theory or any assumption within a paradigm makes a certain broad claim about nature, which is not empirical but speculative. Observational data also assert something about nature, which unlike the theory or assumption, is empirical. So when observation and theory (or assumption of a paradigm) conflict, we are forced into a position to reject on or the other. For Popper, the choice is obvious, reject the non-empirical speculation in favour of the empirical data. He was concerned, however, that some scientists may be tempted to adjust the theory to fit the data. In this on-going example, however, the Aristotelian astronomers, by introducing epicycles and rearranging the celestial spheres, were not altering the Aristotelian paradigm as Popper would suppose. Rather, they were interpreting the empirical data on nature to fit the view of nature as postulated by the paradigm. Specifically, they preserved the principle of uniform motion while trying to interpret the observed retrograde motion as really not retrograde motion but an illusion caused by epicycles. In a way, this may be seen as part of the influence of theory on data as observed by Kuhn (1962) and Feyerabend (1975). Thus, the Aristotelian astronomers were not rearranging elements of the paradigm; rather, they were adjusting the model of nature to accommodate the paradigm. They were rearranging the features of nature to fit the predictions of the paradigm. I emphasize this clarification because it is critical, as it will help in the development of more effective analytic tools for problematic theories.

The goal of ad hoc stratagems to alter the perception of reality rather than temper with a paradigm is indeed more pernicious than what Popper is worried about. Popper’s main concern was the tendency of the adherents of a scientific theory to render it pseudoscientific by shielding it from scrutiny. Thus, auxiliary assumptions or ad hoc hypotheses are a problem because they change our view of nature in a way that normalizes the anomaly rather than altering the theory to fit the data. The former, indeed, is more effective in immunizing the theory from the impact of the anomaly than altering the theory to fit the data. I think that any manipulation of a theory, even if only by interpretation, to fit the observational data represents responsiveness of the theory to the empirical data and I see no epistemic impropriety in that. It seems to me that sensitivity of a theory to observational anomalies is an epistemic virtue rather than vice. I think what Popper was actually worried about was responses (to anomalies) that preserve the theory, and that only happens when the response alters interpretations of the world rather than interpretations of the theory.

There are other illustrations of this. In the response to the anomaly in the orbit of Uranus under Newton's theory of celestial mechanics, the ad hoc hypothesis that was introduced was that there is probably another planet in the vicinity of Uranus that is causing the disturbance in its orbit. This, like the Aristotelian examples, involved altering our view of the world rather than altering the theory itself. In another example, in modern physics today, there is a 95 percent variance between the theoretical model of the composition of the universe and the empirical data. Here again, what the adherents of the current cosmological paradigm have done in response is to propose the ad hoc theories of "dark matter" and "dark energy." This again clearly involves proposing an alteration of the world to fit the theory rather than the converse. The overarching point here is that the alteration — whether the introduction of new objects or new orbits or their rearrangements — occurs on the observation side rather than the theory side

By proposing "dark matter" and "dark energy," for example, we are pointing elsewhere outside our theoretical paradigm for the source of the theory-observation mismatch. Supposing this problem were presented to somebody who is not a cosmologist, he might insist that we go back and re-evaluate our theoretical model and its assumptions, no matter how many times we come back to report that we found nothing wrong with the model. I do not see what would bring any non-practitioner of the *Standard Model* paradigm, the theory which generates this anomaly, to the thought that perhaps there is some material everywhere around us which registers no effect on us or our instruments despite having substantial mass. This is not to say that the existence of such a substance is impossible, but to any dispassionate analyst of the problem that may well be considered most improbable. Irrationality, as contemporary empirical psychology is increasingly revealing, is not a psychological affliction of only infants and the religious, but rather a very general human psychological disposition to which scientists are not immune.

The examples above illustrate an important characteristic feature of the synthetic approach to puzzle solving, which is that it consists of looking outside the paradigm to explain the anomaly, which guarantees the preservation of (at least, the core assumptions of) the paradigm. In the synthetic approach, we try to construct a view of nature that will fit the paradigm rather than the converse, as Popper thought.

Ad hoc hypotheses have generally been the first recourse in the face of theoretical anomalies. While they have sometimes worked in minor variances between theory and observation, it often takes the alteration of some fundamental assumption within the theoretical paradigm to rectify the deeper and more enduring anomalies. As the Aristotelian example illustrates, all the numerous epicycles and alternative arrangements of the celestial spheres did not resolve the key anomalies of the paradigm until some core assumptions of the paradigm itself were altered. In proposing that the orbits of the planets are elliptical, Kepler directly contradicted the two most fundamental assumptions of the Aristotelian paradigm namely, that the motion of the heavenly bodies is uniform and circular. As Copernicus explained, the objective of his (and of course, the other Aristotelians') alternative arrangement of the celestial spheres, which included the introduction of epicycles, was to preserve the *Aristotelian principle of uniform motion*. Thus, even though Copernicus' system conflicted with Aristotelian physics, Copernicus did not seek to overturn those principles. Instead, he offered an arrangement of the celestial spheres that would fit those Aristotelian principles. This is consistent with Kuhn's (1962) and Lakatos' (1978) observation that adherents of a paradigm never question its core assumptions. To do so according to Lakatos is to opt out of the paradigm. That is why such core assumptions are always dislodged only by challenges from without, the consequence of which is often a paradigm change. It is for this reason that it is generally held that the Aristotelian cosmological paradigm ended with Kepler, who opposed the principle of uniform motion. On the part of Copernicus, we can say that he opted out of the Ptolemaic paradigm of geocentrism.

In the remaining sections of this chapter, I illustrate how the synthetic approach to puzzle solving shows up in the Darwinian paradigm, and how that is the reason that the paradox of altruism remains intractable. One predictive consequence of the Darwinian evolutionary explanation is that social behaviour should not have evolved. It is however quite rampant in nature, and thus a puzzle for the Darwinian paradigm. Here again, the response has been to preserve the axioms of the paradigm and instead turn outward to nature, which the modern

Darwinians reinterpret in a way that fits the paradigm. The puzzle of social evolution in the Modern Darwinian Paradigm has created one of the most impressive collections of ad hoc hypotheses yet. This is due not only to the longevity of this puzzle but also because of its contemporaneity. Perhaps in future only a couple of key ad hoc hypotheses will be remembered.

The overarching assumption in all the auxiliary hypotheses in response to the anomaly of social behaviour within the modern Darwinian paradigm is what Michod (1982, 25) calls a "structured population" in which factors such as "kinship recognition, geographical structure, dispersal systems, and mating systems," etc. operate. It proposes that social populations are structured in such a way that the effects of certain behaviours do not affect individuals randomly in the population, but instead fall disproportionately on some individuals relative to others. Thus, under the inclusive fitness/kin selection hypothesis, the benefit of altruism falls disproportionately on individuals who are "identical by descent (ibd)" to the altruist (Hamilton 1964). Under group selection, the benefit falls on members of the altruist's home group (Darwin 1859; 1871). Under the reciprocity/cooperation models, the benefits of altruistic behaviour fall disproportionately on fellow reciprocators/cooperators (Trivers 1971; Axelrod and Hamilton 1981). You will notice in the detailed discussion of the Darwinian paradigm below that the core assumptions of the modern Darwinian explanation, whose predictive consequences social behaviour contradicts, are not altered by the auxiliary hypotheses. What I do subsequently is cast an analytic beam upon the paradigm itself to see which, if any, of its assumptions and axioms are in conflict with empirical reality.

V. The Darwinian Social Evolution Puzzle

Social behaviour has been the bane of Darwin's theory of evolution ever since its inception over a century and a half ago. Darwin treated the problem of the evolution of human morality at length, and ultimately left it with no satisfactory account under natural selection. The Darwinian problem of morality, and in fact social evolution in general, crystalize in the "paradox of altruism," which has become a major research focus in theoretical evolutionary biology (Grafen, 2007).

In fact, to date altruism remains a puzzle for the evolutionary biologist (Bshary & Bergmuller, 2007; Boyd & Richerson, 2009; Johns et al, 2009; Boehm 2012), as there is no account of its evolution that is completely satisfactory. Instead, there is a patchwork of explanatory models, none of which has universal appeal among biologists. The models include kin selection (Hamilton, 1964; Maynard Smith, 1964), group selection (Wynne-Edwards 1962, 1986; Wade 1977; Wilson and Wilson 2007; more), reciprocal altruism (Trivers 1971), and cooperation (Axelrod and Hamilton 1981). Each of the models explains some forms of altruism and is contradicted by other instances of altruism. Why is there such theoretical chaos in the attempt to apply natural selection to explain social evolution?

The *modus operandi* of Darwinian natural selection is the maxim of "survival of the fittest" (1859, 108). Basically, according to Darwin, "as more individuals are produced than can possibly survive", a "struggle for existence" inevitably ensues (Darwin 1859, 91), in which those individual differences and variations that are favourable are preserved and those that are injurious are eliminated (Darwin 1859, 108). This dual process of preservation of some and elimination of others results from the differential rates of survival and reproduction amongst individuals according to their differential fitness for the given conditions of life.

However, natural selection works only if individuals with favourable traits are able to pass such traits on to their offspring. Hence, according to Darwin (1859, 168), natural selection relies on "the strong principle of inheritance." Consequently, Darwin, in talking about evolution by natural selection, makes it clear that "any variation which is not inherited is unimportant for us" (Darwin 1859, 31). Thus, any trait that is not heritable is not amenable to explanation by natural selection. This is the broad paradigm and immutable core of classical Darwinian explanation. This is now considered the bare bones of Darwinian explanation. Nesse (2000, 228), explains that when the logic of natural selection "is combined with our intuitive notion that altruism consists of costly acts that benefit others, and genes are seen as the ultimate currency, then altruism is impossible." This places the source of the problem of social evolution squarely within the frame of the Darwinian paradigm. Yet none of the many proposed solutions to the puzzle alters any of the Darwinian principles outlined here. They all postulate features of the social group that will make altruistic behavior sustainable. (See pp 24-25 of the next chapter).

In the century and half since its inception, the Darwinian paradigm has expanded enormously, and in its current form — the so-called modern synthesis — it has developed considerable internal tension (with the traditional Darwinian explanation, some think).

VI. Evolution and Structure of the Modern Darwinian Paradigm

In this section, I describe the nature of the modern Darwinian explanatory paradigm, particularly how it views social behaviour and the fundamental axioms upon which it models the evolution of social behaviour. The core principles of the contemporary Darwinian explanation of social behaviour are best appreciated if presented in a historical context. The historical development of the Darwinian explanation can be shown to fit the pattern for scientific theories in general as described particularly by Kuhn above. Kuhn (1962/1996) explains that every paradigm comes incomplete and is “at the start largely a promise of success.” The practice of “normal science,” as Kuhn sees it, involves in the main, efforts to extend the fit between the paradigm’s predictions with the empirical data, as well as further articulation of the paradigm. In this regard, the Darwinian paradigm is not different. It is easy to see how Darwin’s annunciation of his theory of natural selection launched that particular paradigm of evolutionary explanation. In this discussion, we shall see quite clearly how incomplete it was at its nascence, the daunting puzzles that immediately confronted it, and the “mop up” activities that have brought it to its current form. In the process, we shall also see how the practitioners’ articulation of the Darwinian paradigm took it on a rigid tangential trajectory away from Darwin’s original flexible curve.

The two major deficiencies of the nascent Darwinian paradigm were its weak theory of heredity and its difficulty in explaining social behaviour. It is now well over a century and half since the birth of the Darwinian paradigm, and those two problems are still unresolved. The efforts to solve the hereditary problem have given the Darwinian paradigm the unique feature of being an amalgamation of two distinct paradigms – Darwinian natural selection and Mendelian genetics. That is, it coopted or absorbed Mendelian genetics, which was an entirely separate paradigm, into its explanatory scheme, hence, the name *the Modern Synthesis*. Prior to this integration, Darwinian theory struggled through its first half century without a robust theory of heredity. *Blending inheritance*, which was the default theory of heredity then, was ill-equipped to deal with some fundamental puzzles in the Darwinian paradigm. The most acute of these was the maintenance of variability, which the action of natural selection appears intuitively to depress. In addition to this, the mechanism of blending inheritance predicted the exponential attenuation of heritable variation down the generations. The loss of variability at such a rate was simply not supported by the empirical estimates of the mutation rates that begun to appear at the turn of the twentieth century.

There is evidence that Darwin was quite unsatisfied with blending inheritance and sought a more robust mechanism of inheritance (Fisher 1929). However, in the absence of any better alternative theory, he continued to work with blending inheritance, which explains his inability to give a convincing account of the maintenance of variability. All he did was to assert without justification at several places in his *Origin of Species* (156, 169), that natural selection increases variability.

In spite of these difficulties faced by natural selection under blending inheritance, there was a great deal of scepticism and hesitation about adopting *Mendelian Particulate Genetics* – a new hereditary theory that appeared in the 1900s, following the rediscovery of the work of the Austrian monk Gregor Mendel, which had been neglected for nearly four decades. The biometricians, who were working on the mathematical models for the detailed mechanism of evolution by natural selection, felt that Mendel's particulate genetics was incompatible with the imperceptible gradualism that characterized evolution by natural selection. However, crippling anomalies generated by their *blending inheritance* theory, coupled with the spectacular empirical successes of Mendelian genetics, forced their capitulation.

At the time of its adoption however, the Mendelian paradigm was hardly two decades old, nascent and inchoate, its articulation or the "mop up" activities that Kuhn described was still underway. Thus, the complexity of some of the Darwinian puzzles was compounded by puzzles from the nascent Mendelian paradigm that it incorporated. In fact, the version of Mendelian genetics that Darwinism adopted was the earliest and most rudimentary, and was laid down less than a decade into the development of the Mendelian paradigm. In that version, traits are cast as binary phenotypes, which are controlled by a pair of contrasting alleles at a single locus. That would form the inalterable bedrock of the modern Darwinian paradigm. It was articulated by the founding fathers of the modern synthesis (Fisher 1930; Haldane 1932; Wright 1932) and reaffirmed consistently (Hamilton 1964; Trivers 1971; Grafen 1984; Queller 1992) down to the present (Krebs 2012, 62; Van Veelen et al. 2012, 68).

The structures for the mathematical modeling of Mendelian genetics were developed independently by G. H. Hardy and W. Weinberg in 1908 in the now famous Hardy-Weinberg ratio. The Hardy-Weinberg ratio states what the genotypic and allelic frequencies of an offspring generation will be in a population in which mating is random and in which the evolutionary forces of natural selection, mutation, migration and drift are absent. The fundamental heuristic of the Mendelian genetic system is to suppose a pair of contrasting alleles, say (A_1), (A_2) to control each trait. If the respective frequencies of the alleles in the population are (p), (q), then the following frequencies emerge for the three possible genotypes in the offspring generation; it is the so-called Hardy-Weinberg proportions:

$p^2:2pq:q^2$

This formed a perfect toothing stone for the Darwinian system, which is interested in gene frequency changes caused by natural selection. The Hardy-Weinberg principle showed that in the absence of disturbing forces, allele frequencies will remain constant. Fisher, Haldane and Wright used the Hardy-Weinberg principle as their starting point (Halliburton, 2004). Since these proportions are the state of affairs in the absence of selection, a model for natural selection needs to determine how these gene frequencies will change under selection, recognizing that the main driver of such frequency changes would be the relative fitness (*w*) of the phenotype associated with each allele. From the Hardy-Weinberg proportions above, we derive the following equation to represent frequencies of the two contrasting alleles at any time, that is, in the absence of selection, migration or drift:

$$p_{t+1} = \frac{p^2 w_{11} + pqw_{12}}{w} \qquad 1$$

$$q_{t+1} = \frac{pqw_{12} + q^2 w_{22}}{w} \qquad 2$$

To ascertain whether and how much evolution has occurred, we calculate the change in gene frequencies by the following equation, which is derived from equations (1) and (2) above:

$$\Delta p = \frac{pq[p \ \ w_{11} - \ w_{12} \ \ - q \ \ w_{22} - w_{12} \ ]}{w} \qquad 3$$

This is the fundamental model of evolution by natural selection under population genetics, which in turn is the germ of the modern synthesis. All the other branches of biology within the modern synthesis joined in after this point by demonstrating explanatory consilience with this core principle. The Δp term in the equation is the change in the frequency between generations of the gene that expresses the trait in question, hence the neo-Darwinian conception of evolution as a change in gene frequencies.

However, as we were soon to learn, many of the traits whose evolution we wish to model do not fit this simple mathematical scheme. Mendelian genetic principles were established in work with very simple physical traits such as flower colour and seed shape of peas. A single pair of alleles was assumed to control each trait. In the early twentieth century, at the time that Darwinism adopted such Mendelian genetics, those principles held because the early experiments largely replicated Mendel’s results by investigating similarly simple traits. As time marched on, however, it became increasingly clear that many traits were much more complicated than those from which Mendel drew his insight. In fact, this was the main reason why the early Darwinians were reticent about the interpretation of natural selection in Mendelian terms. They did not feel that the idea that an evolutionary change, especially for a complex trait such as behaviour, could occur by the flip of a single hereditary factor, was compatible with the gradualism that Darwin emphasized for natural selection. In fact, it was in cognisance of this apparent incompatibility that Fisher (1918) proposed that we consider complex traits to be controlled by several Mendelian factors, i.e., multiple alleles at multiple loci. However, extending the basic two-allele single-locus model over such multiplex Mendelian systems, which feature rampant cross-locus interactions, becomes inordinately complicated; and as Halliburton (2004, 487)) laments, “if we cannot comprehend the potential complexity of two-locus systems, how can we hope to understand the interactions of hundreds or thousands of loci?”

Simply, certain traits are just too complex for the rather simplistic Mendelian mechanism for evolutionary explanation. As a matter of fact, Hugo DeVries who is largely credited with reviving Mendel’s work, had cautioned the early Mendelians about the limits of Mendel’s theory. He wrote in a 1911 correspondence with Bateson (Provine 1971, 68): “I prayed you last time … it becomes more and more clear to me that Mendelism is an exception to the general rule of crossing. It is in no way the rule.” The modern synthesis however was in full flight as an explanatory paradigm before this difficulty could sink in. Its early adherents were (in Kuhnian terms) in the full “articulation” and “mopping up” mode, trying to persuade sceptics and fending off detractors. To reopen the Mendelian component of the synthesis would have meant and appeared to be nothing more than going back to the drawing board, and the de facto demise of the synthesis. Thus, to date we still define and model population genetics on the basis of Mendelian genetic principles (Grafen 1984; Wade 2008, 2; Lynch 2007), and then try to get around the anomalies that arise out of that “by introducing ad hoc some auxiliary assumptions, or by reinterpreting the theory ad hoc in such a way that it escapes refutation.” The words in the quotation marks are from Popper (1963, 26), and describe what he calls “pseudoscience.” However, this remark from Popper is another vindication of Kuhn, who asserts such practices to be the mainstay of “normal science.”

Thus, under the neo-Darwinian system, once we determine that some feature in a population is amenable to “Darwinian” explanation, the inevitable starting point is to suppose an allele that is responsible for the trait’s expression and a corresponding allele for its absence (Grafen 1984). If any anomalies arise out of that supposition, as Hamilton (1964) noted for social evolution, we propose a plausible ad hoc condition under which that anomaly does not arise. If other anomalies remain, other plausible circumstances are further invoked. As many such plausible circumstances as needed are introduced to address all the anomalies, reminiscent of the introduction of epicycles under the Aristotelian paradigm. Thus, the condition invoked under group selection makes the benefit to the group to which the altruist belongs greater than the cost to the altruist. Under kin selection, the benefit to the altruist’s relatives must be greater than the cost to the altruist; and under reciprocal altruism, the condition is that the altruistic act be reciprocated.

VII. The Ascendency of Population Genetics

Mayr (1993, 31) notes two major features of the modern synthesis. The first is what he describes as “a convincing refutation of the three major anti-Darwinian paradigms – the typological-saltationist, the teleological- orthogenetic and the transformationist - Lamarckian.” These were some of the detractors of the modern synthesis I alluded to earlier, aside from Darwinians who were skeptical or decidedly against the synthesis. The second feature of the modern synthesis that Mayr notes is the synthesis itself, of which he identifies two broad views of evolution as constituents – the naturalist’s tradition of Darwin on one hand and the geneticist’s mathematical reductionist view on the other. What Mayr is pointing out in this second feature is that the product of the synthesis was not a simple unity. It had internal components and complexity. There appears to be a fault line keeping the major components apart, indicating that there is little intertwining or intermeshing of the founding components. Depew (2013, 384) has also noted that some scholars see the modern synthesis as “a matrix of inter-field practices and assumptions, not a theory.”

The modern synthesis, thus, is a collection or loose association of parts, rather than an organic fusion. For example, Mayr (1993) describes the synthesis internally as “the thinking of three major biological disciplines – genetics, systematics and paleontology.” If we superimpose the bipartite structure Mayr described earlier on this tripartite view, systematics and paleontology will fall under the traditional Darwinian explanation, whilst the discipline of (Mendelian) genetics aligns with what Mayr describes as the “mathematical reductionist view.” He adds (1993, 31) that “It was a synthesis between an experimental-reductionist philosophy (strongest among the geneticists) and an observational holistic philosophy (strongest among the naturalists), and finally between an Anglophone tradition with emphasis on mathematics and adaptation and a continental European tradition with emphasis on populations, species and higher taxa." But I think the true synthesis, and what most profoundly altered Darwinian evolutionary explanation, was the incorporation of Mendelian genetics and the formulation of evolutionary change as discrete, quantifiable genetic steps.

The original concern about incongruence between the two major components of the synthesis has now evolved into considerable tension between them today, chiefly in the form of the population genetic view of evolution as change in gene frequencies, on the one hand, and the critics of that view on the other. Looking at a review by one of its architects (Mayr 1993), it is clear that the two traditions have never completely harmonized in all those years. As a matter of fact, they have actually been pulling apart since at least the 1960s. As Mayr (1993, 32) admits, "conspicuous differences still remained. To begin with, evolution continued for the geneticists to be a change of gene frequencies, with the gene considered the target of selection, while for the naturalists, evolution was a series of processes resulting in adaptedness and diversity, with the individual being the target of selection."

It appears however, that the geneticist/mathematical reductionist camp has gained the upper hand and is evermore becoming the face of the modern synthesis, while the "naturalist evolutionist tradition of Darwin" has now become its critic. Population genetics is widely regarded as the cornerstone of modern evolutionary explanation – the modern synthesis (Wade 2005; Lynch 2007; Pigliucci 2008; Okasha 2012). How did this happen?

A number of prominent theoretical biologists (Alexander 1974; Axelrod and Hamilton 1981, 1390; Wilson and Wilson 2007; Forster 2009) have noted that the early 1960s were a watershed period in the development of the modern Darwinian paradigm. Until then, group selection, which Darwin proposed in 1859 and 1871, was generally accepted as key to the explanation of social behaviour, even though the problem of the theoretically predicted decline of altruistic individuals within groups remained unresolved. Even the population genetic explanations of social evolution prior to the 1960s were modeled with the framework of group selection. It proposes that "most traits associated with eusociality [for example] do not evolve by increasing in frequency within colonies, but by increasing the colony's contribution to the larger gene pool" (Wilson and Wilson 2007, 340). Actually, the frequency of the altruistic trait within the group should be falling, and the efforts were directed largely at resolving the problem of the decline of the altruistic allele within groups. Wright (1945), for example, proposed a model for the maintenance of the altruistic gene by suggesting that if a species consists of small nearly isolated populations, the sociality gene may drift to fixation in some of them and then spread into the rest of the population through increased migration. In Williams and Williams' (1957) model, the sub-groups are "sibships" within a single large panmictic (random mating) population.

So what changed in the 1960s? The discussion in the previous paragraph should dispel the fairly widespread perception that apart from brief references by R. A Fisher and J. B. S. Haldane, the problem of social evolution was largely neglected prior to the 1960s (Ratnieks and Helantera 2009; Michod 1982; Axelrod and Hamilton 1981). Axelrod and Hamilton (1981, 1390), for example, assert that prior to “about 1960, accounts of the evolutionary process largely dismissed cooperative phenomena as not requiring special attention.” However, what follows next in their analysis indicates that it was not actually the case that the problem was not recognized or that nothing was done about it. Instead, what was being done about the problem of social cooperation was later discounted due to the upheaval in evolutionary explanation that took place in the 1960s. Axelrod and Hamilton, for example, suggested that the problem was not seen as acute because scholars took comfort in “a misreading of theory that assigned most adaptation to selection at the level of populations or whole species” (1981, 1390). Hamilton was more elaborate on this in his (1964, 19) paper, in which he writes: “That this phenomenon [namely, social behaviour] presents a difficulty, namely an apparent absence of positive selection, is obvious as soon as we reject the pseudo-explanations based on the ‘benefit to the species.’” They dismissed group level selection as weak, and instead emphasized kinship and reciprocity. Dawkins (1976 [1989], 7) similarly denigrated “group selection theory” by suggesting that it is “long assumed to be true by biologists [who are] unfamiliar with the details of evolutionary theory.” It seems, then, that the more plausible characterization of the transition in the 1960s is what Wilson and Wilson (2007, 327) describe as “the rejection of group selection and the development of alternative theoretical frameworks to explain the evolution of cooperative and altruistic behaviors” (also see Okasha 2010).

So Hamilton did not fill an explanatory vacuum, as Ratnieks and Helantera (2009) and others suggest. Rather, Hamilton budded off from an explanatory tradition that upheld the true spirit of the modern synthesis as consilience between the two broad traditions of Darwinian naturalism and abstract mathematical genetics. The consilience is evident in the works of Wright (1945) and Williams and Williams (1957) that were explained above. Hamilton’s inclusive fitness model was the first major explanation of social behaviour that did not appeal to the “between groups” selection advantage of groups with altruistic individuals to compensate for the “within group” disadvantage of altruists. Instead, the model relied solely on abstract mathematical genetic principles to explain social evolution. That however, is only one of the two major constituents of the modern synthesis (Mayr 1993). Hamilton was the first to model social evolution exclusively on selection between alleles with no group level selection, which in fact had been a constant feature of prior Darwinian explanations of social evolution. In fact, Wilson and Wilson (2007, 334) assert that “Inclusive fitness theory (also called kin selection theory), evolutionary game theory (including the concept of reciprocal altruism), and selfish gene theory were all developed explicitly as alternatives to group selection.”

Thus, the significance of Hamilton in evolutionary biology is that, with his exclusion of group level selection from his model of social behaviour, he initiated a reversal of the narrowing gap between traditional Darwinian naturalism and mathematical genetics, which had been the project of the modern synthesis. Hamilton’s 1964 paper triggered a movement within the modern synthesis that sought to reduce evolutionary explanation to the mathematical manipulation of gene frequencies. It is a brand of evolutionary explanation which, as Axelrod and Hamilton (1981, 1390) say, “has increasingly taken a gene's-eye view of natural selection.”

It is interesting to note, however, that within a few years (by 1970) after his watershed publication, Hamilton had considerably moderated his stance on the issue of group selection (Forster 2009). Under the influence of his collaboration with mathematical biologist George Price, Hamilton had come to accept the importance of group selection in social evolution. In his autobiography, Hamilton (1996) comments on his collaborative relationship with Price:

> I am pleased to say that, amidst all else that I ought to have done and did not do, some months before he died I was on the phone telling him enthusiastically that through a “group-level” extension of his formula I now had a far better understanding of group selection acting at one level or at many than I had ever had before. (Quoted in Forster 2009, 405)

However, the train had left the station following Hamilton’s game changing publication of 1964, and publications reflecting his modified stance in the 1970s and beyond could not reverse the momentum.

The modern synthesis originated with population genetics, as its pioneering architects, R. A. Fisher, J. B. S. Haldane and Sewall Wright, were mathematical biologists and geneticists. The other branches of biology were slower to sign on, and some, such as developmental biology, have never been able to find an entrance and have remained outside the synthesis. Even for the branches of biology that joined in, a fault line remained between them and the original population genetic nucleus. The exclusionism (of other Darwinian explanations) to which population genetics took in the 1960s triggered a widening of the fault line between Darwinian naturalism and mathematical genetics into a gulf, and a battle for the soul of modern evolutionary explanation ensued. Thus, the debate in evolutionary biology changed from *what arrangement of the different levels of selection best explains the evolution of a particular trait?* (As was the case from 1859 to 1964), *to which level/unit of selection to adopt as the sole or primary driver of evolutionary change?* (From 1964 on).

Other than the denunciations of group level selection, the other exclusionist feature of the Hamiltonian school has been the exaggeration of the status of the gene. Often, the gene or DNA is cast as a conscious strategy-plotting entity. Consider the following: “A gene, in effect, looks beyond its mortal bearer to interests of the potentially immortal set of its replicas existing in other related individuals” (Axelrod and Hamilton 1981, 1390). The Hamiltonian (gene-centric) school was also characterized by reification of genes and teleology (see also Dawkins (1976, and 1982)). The literature is replete with evolutionary explanations that take off from the base assumption of genes seeking to "maximize their inclusive fitness." Hames (2001, 6946), for example, states that "human behavioral ecology attempts to explain behavioral diversity as a consequence of environmentally contingent responses made by individuals in their attempts to maximize their inclusive fitness." The exclusionist population genetic brand of evolutionary explanation has been on the ascendency since its inception in the 1960s and has now become the dominant force in modern evolutionary explanation. Its view of evolution as changes in gene frequencies is now nearly the standard perception and in fact, definition of evolution (Lynch 2007; Futuyma 2006; Carroll 2005; Wade 2005). Lynch (2007, 8598) asserts that "the field of population genetics is now so well supported at the empirical level that the litmus test for any evolutionary hypothesis must be its consistency with fundamental population genetic principles, [which are] grounded in basic Mendelian processes and sampling theory." This will jolt Karl Popper in his grave. To assert that a scientific theory is "so well supported at the empirical level" that no other hypothesis can contradict it is quite blatantly contrary to some of the most basic principles of scientific discovery.

Many biologists and philosophers of biology now view science in a similar manner to Lynch above, i.e. seeing some of its theories as so well corroborated that they become certified truths not to be contradicted by any new theory or data. There are many pronouncements that reveal the feeling on the part of some theoretical biologists (and in fact, many field researchers as well) that empirical research findings have to conform to theoretical predictions in order to be credible. According to science reporter Michael Marshall (2010), Oxford evolutionary biologist Allan Grafen had explained during a conference that (quoting Marshall): “theoretical biologists have always known that inclusive fitness was an approximation, though this seems not to have filtered through to experimental biologists, who have tended to take it as gospel.” I cannot speculate on what experimental biologists did that suggested that they were missing the nuances in their theoretical claims. However, what we can gather from Grafen’s claim is that experimental biologists have been following theoretical guidelines that they apparently sometimes misunderstand. Such a state of affairs turns the principles of scientific discovery upside down, particularly as Karl Popper would see it. Unfortunately, experimental biologists, most of whom lack any exposure to the philosophy of science, fail to see the epistemic impropriety in such a framework of scientific discovery.

As a point of clarification, it is fine for experimental scientists to set up experiments to test theoretical claims, but the validity of empirical research findings should not be judged on the basis of conformity to theoretical claims. However, it is common in recent decades to see authors of experimental research articles invoke conformity with kin selection, and particularly inclusive fitness, to validate their research results. If there is this perverted thinking that experimental research must conform to the prevailing theoretical paradigm in order to be credible, then we have to wonder just how much research data is being thrown out on grounds of lack of conformity to theory. That is why we have to take with advisement claims such as Lynch's (2007, 8598), that "no principle of population genetics has been overturned by an observation in molecular, cellular, or developmental biology." Lynch appears to be suggesting that population geneticists would humbly concede and abandon any principle that conflicted with the empirical data. That, according to Kuhn (1962), would not likely be the response. What has often happened in the practice of "normal science" is that contrary data do not result in the revision of the assumptions of a theoretical paradigm. If the adherents of the paradigm are unable to re-interpret the theory to fit the data, they tend to simply ignore it. As a matter of fact we are living through one such period of denial at this very time. Population geneticists have long known, for example, that their most fundamental modeling assumption for social evolution, the "phenotypic gambit," is actually a caricature of reality (Grafen 1984; Gardner et al. 2011) and very clearly contrary to the empirical data from behavioural ecology (Yakubu 2013), but that has not caused them to abandon it. Instead, we are simply told that the models are abstractions from reality (Grafen 1984).

There are many who see Hamilton's 1964 work as a successful and complete account of social evolution and therefore a reason to be disdainful of other models that do not fit into Hamilton's model. Proponents of group selection on the other hand have been on the defensive since the 1960s. Their theoretical claims tend to be more modest. In fact, very few scholars today will model the evolution of any trait exclusively on group selection. Instead, in what is dubbed multi-level selection, group selection supplements, or is supplemented by, other levels of selection. The effort to demonstrate the equivalence between group selection and kin selection may also be a consequence of this state of affairs. Even though there are still a few evolutionary biologists pitching for multilevel selection, it is clear from developments since the 1960s that population geneticists are quite determined to go it alone.

This current character of population genetics seems to abandon all the compromises that made the modern synthesis possible. Mendelian particulate genetics suggested some kind of saltationism, which was antithetical to Darwinian gradualism. The compromise Fisher brokered in his 1918 paper was to imagine the process of evolution to be represented by "multiple Mendelian factors." However, today, with the "phenotypic gambit,[1]" we are back to modeling evolution of complex traits such as behaviour, for example, as a single giant evolutionary step, whereas Fisher (1918, 135) proposed that complex "features such as stature are determined by a large number of Mendelian factors." Clearly, contemporary population genetics has gained no traction in upholding the compromise Fisher proposed, as we continue to model a trait as complex as social behaviour, as a two allele single locus trait.

---

[1] The term "phenotypic gambit" is explained in chapter II pages 1-3 and 13.

There have been a number of criticisms of the modern synthesis, which stem from population genetics as its explanatory cornerstone (Okasha 2012; West-Eberhard 2003). Some scholars have argued that “the axiomatic definition of evolution as a change in allelic frequencies dictated by the population genetic perspective” leaves a lot out of the evolutionary picture (Pigliucci 2008, 322; Carroll 2005; West-Eberhard 2003). Pigliucci (2008, 321) notes particularly that population genetics “is at its core a theory of changes in gene frequencies and cannot be a theory of change in form.” In this work, I point out another limitation similarly imposed on the modern synthesis by population genetics, which I will show to be the basis of the protracted difficulty the synthesis has had in explaining social evolution. The modern synthesis, as a fusion of Darwinian natural selection and Mendelian genetics, models evolution as a change in gene frequencies (Lynch 2007; Futuyma 2006; Carroll 2005; Wade 2005). This in itself has attracted some criticism (refs). That view of evolution sets a particular framework for modeling evolution into which every trait whose evolution is to be modeled must fit.

Today, social evolution remains the most contentious research area in modern evolutionary biology, and as we have seen above it was the bid to bring social evolutionary explanation under the exclusive domain of population genetics that has resulted in the current state of virtual dissolution of the modern synthesis.[1] However, there are some who, concerned about the reputation of social evolutionary biology, try to downplay or veil the theoretical disarray. Thus, Forster (2009, 403) writes of his bid to "offer some reassurance that the appearance of ongoing controversy is mostly illusory." Okasha (2010, 653) also writes: "I contend that there is little to argue about. Much of the current antagonism stems from the fact that different researchers are focusing on different aspects of the same phenomenon, and are using different methods." Again, this is an attempt by Okasha to broker a compromise, due to his express concern that such inter-tribal warfare within social evolution research will "cause serious damage to the field." However, it is not true that the different models apply to different aspects of the problem of social evolution. They are in most cases rival accounts of the same things. Whichever fault line we consider, whether genic versus higher levels of selection on one hand, or population genetics versus traditional Darwinian naturalism on the other, no side makes modest claims. In the next chapter, I begin the process of sorting out the theoretical jumble.

The situation that triggered Okasha's remark, for example, was a particularly rancorous debate following an article (Nowak et al 2010) that questioned the very legitimacy of the flagship theory of the population genetics stream of the modern synthesis – Hamilton's inclusive fitness theory. The threat

---

[1] As the other components of the synthesis are increasingly being alienated.

to the theory was considered so severe that a hundred and thirty-seven scholars who are sympathetic to the theory appended their names to a collective rebuttal to the article.[1] That is not the kind of response one would expect for an inconsequential disagreement.

The battle between the gene-centric movement and its critics is real. My point, however, is that it is just a symptom of a root problem, which is the unresolved problem of heredity. A testament to this is the surprising[2] concession of Dawkins (2004, 387) to his critics, when he says "a Darwinian replicator does not have to be specified as DNA, but can be a Maynard Smithian 'strategy' defined in a minimalist 'like begets like' fashion. Presumably DNA is involved in practice, but it is not a specified part of the reasoning." Dawkins' ultimate point in that discussion is that the iconic "selfish gene" is not the Mendelian gene that most of us take it to be, but rather some "unspecified" abstract concept that has no specified relationship with DNA. However, this sounded like a bombshell perhaps only because it came from Dawkins, the populariser of *genism*. In fact, other key scholars of the gene-centric movement have always professed the abstraction of their models from reality. Grafen (1984, 64), for example, writes in reference to gene-centered evolutionary modeling, that "taken literally, the *gambit* [i.e. the modeling assumption] is usually false." The question we should be

---

[1] See *Nature* volume 21, 24 March 2011.

[2] Dawkins himself said he expected us to be surprised that he holds such a view.

pondering then is: what is our contemporary theory of heredity in evolutionary explanation? If all we have now is some principle such as “like begets like,” a conceptual object (as Dawkins suggests) that has no specified relationship with the DNA of contemporary empirical genetics, can we say we have a satisfactory and credible theory of heredity?

Further development of inclusive fitness, since Hamilton (1964), has largely been its adaptation or application to specific cases. For example, Hamilton (1970) extends inclusive fitness to explain spite, by linking it with Price’s (1970) equation, and Ohtsuki et al. (2006) and Grafen (2007b) show how inclusive fitness works in social networks. There have also been attempts to reformulate it as an optimising principle (Grafen 2002, 2006, 2007). This is what has led to the frequent talk, these days, of “inclusive fitness maximization” (Abbot et al 2011). Grafen (2006, 541) explains that “the chief difficulty in accepting the optimization principles of Fisher and Hamilton has arisen from misunderstanding of what maximization of fitness means.” He suggests two approaches to clarifying this, the first of which is the reestablishment of “the connection between inclusive fitness and gene frequency change” (542). He develops a version of inclusive fitness to include uncertainty and conditional behaviour, explaining that. “biologists take for granted that animal behaviour is conditional and that it must be average fitness that is maximized, as there is no such thing as unconditional behaviour and the life of no organism possesses certainty” (542). As we shall see in the next chapter, this is a frequent assertion of sociobiologists in denying accusations of genetic determinism.

However, the bulk of the articulations of inclusive fitness have been its various responses to the steady criticism it has been under since the 1970s. There are those, here cited, from Gardiner et al (2011, 1021), that suggest that Hamilton's rule "requires restrictive assumptions, such as genetic variation segregating at only a single locus, weak selection and absence of gene interactions within and between individuals" (Cavalli-Sforza & Feldman, 1978; Charlesworth, 1980; Uyenoyama & Feldman, 1980, 1981; Colwell, 1981; Uyenoyama et al., 1981; Michod, 1982; Queller, 1984, 1985; Avile´ s, 1993; Bulmer, 1994; Wolf et al., 1998, 1999; Gintis, 2000; Wolf, 2003; Wilson, 2005; Doebeli & Hauert, 2006; Fletcher & Doebeli, 2006, 2009; Fletcher et al., 2006; Fletcher & Zwick, 2006; Killingback et al., 2006; Nowak, 2006; Queller & Strassmann, 2006; Traulsen & Nowak, 2006; Wenseleers, 2006; Wilson & Wilson, 2007; Bijma & Wade, 2008; Goodnight et al., 2008; Traulsen et al., 2008; Wilson, 2008; Nowak et al., 2010a,b; Traulsen, 2010). According to Gardner et al (2011, 121), others such as Doebeli & Hauert (2006), Fletcher et al (2006) and Nowak et al., (2010), accuse inclusive fitness practitioners of "redefining the cost, benefit and relatedness terms as a *post hoc* fix and departing from original formulations of kin-selection theory." Gardiner et al's (2011, 1021) response was to "clarify the generality of Hamilton's rule, as a statement of how natural selection drives changes in gene frequencies." Gardiner et al (2011) point out that their analysis does not require any assumption of the absence of gene interactions, both within and between individuals, as charged, for example by Michod, (1982). Fletcher & Doebeli, (2006), Fletcher & Zwick (2006), Queller & Strassmann, (2006), Wenseleers (2006) and Nowak et al. (2010). The claim that Hamilton's rule is ineffective unless selection is vanishingly weak (Wenseleers 2006; Traulsen et al. 2008; Nowak et al. 2010; Traulsen 2010) has also been rebutted by Gardiner et al (2011), Lehmann & Keller, (2006), Gardner et al. (2007). .

So, to the questions as to what has changed since Hamilton (1964), I would answer that at the fundamental level, nothing. Consistent with the ideas of Kuhn and Lakatos (discussed in the earlier sections), all these reformulations and rancorous debates are about peripheral matters. The core assumption or basis of inclusive fitness is not challenged in any of the discussions. As Huneman (2014) explains, the Modern Synthesis (MS), which is the modern understanding of natural selection, is rooted in "population genetics, which unraveled the processes of selection as an aggregated dynamics of allele frequencies. In this perspective, evolutionary theory is much closer to statistical mechanics, as made famous by Fisher's (1930) analogy." That is why Mayr (1963) has called it "beanbag genetics."

In his reformulation of inclusive fitness as an optimizing principle, (Grafen 2006, 542) suggests that "Mendelian-type genetics will be taken as the known process underlying evolutionary change, and accepted as fundamental.

Mendelian-type means that there are haploid sets each of which contains one copy of each locus" "The Formal Darwinism project (Grafen, 1999, 2000, 2002,) therefore aims to link the mathematics of motion (difference and differential equations) used to describe gene frequency trajectories with the mathematics of optimization used to describe purpose and design" (549). So between Hamilton (1964) and now, we are still counting the frequencies of selfish genes and altruistic genes. And that is my concern in my meta-theoretical analysis. It is irrelevant whether they are single genes or multiple per trait and it does not matter if the environment plays a role or not. Those are some of the peripheral adjustments you may see in some of the models.

Fundamentally, however, we are still modeling how altruistic genes might replace selfish genes. Citing a long list of tradition, up to West and Gardner (2013), Bourke (2014), asserts that "Conceptually, [inclusive fitness'] fundamental contribution has been to identify genes as self-promoting strategists whose evolutionary interests are conditional on the relatedness class in which they reside ," and, that by this principle, gene behaviour explains all higher levels of organization, including social groups.

VIII. Concluding Remarks

What I have tried to get across in this chapter is that two (related) difficulties are at the core of the theoretical crisis in evolutionary biology. The problem of heredity, which confronted Darwin right from the onset, remains unresolved, and the problem of social evolution is a consequence of that. Population genetics, which is the extant genetical explanatory tool, is constructed rigidly on a now antiquated Mendelian genetics, which is incommodious for complex traits, of which social behaviour is a model case. We will never be completely satisfied with our models, nor will our explanation be fully convincing until we find a way to model social behaviour based on assumptions that are consistent with contemporary empirical genetics, rather than abstractions that are unconnected to the empirical data, and which are instead dictated by the heuristics of our explanatory paradigm. Let us not rest our laurels on the corroborations of predictions from such abstract models. As Popper (1963, 26) observed, “it is easy to obtain confirmations, or verifications, for nearly every theory,” even a pseudoscience.

Chapter Two

The modern Darwinian explanation and Social Evolution

I. Introduction

In chapter one, I grave a historical sketch of the evolution of the modern Darwinian explanation called the Modern Synthesis. I described it as a loose association of several biological disciplines, and I also pointed out that it features a persistent fault line between traditional Darwinian naturalism and abstract mathematical genetics. I noted that the latter, as population genetics, has now come to be the face of modern evolutionary explanation. I also discussed how developments of the 1960s precipitated new explanatory traditions that veered away from Darwinian naturalism toward abstract mathematical explanations of evolution. In this chapter, I turn to the treatment of the evolution of social behaviour under the modern synthesis. Here we see a new flash of explanatory plurality, all of which emerged subsequent to the events of the 1960s and the ascendency of population genetics. Recall that the significance of W. D. Hamilton and the change in direction of evolutionary explanation came about as an effort to model social evolution solely in terms of gene frequencies. The proliferation of theories of social evolution that followed that theoretical redirection was due mainly to the inadequacies of abstract mathematical models of population genetics in dealing with social evolution. The genetical explanation of social behaviour has been dubbed sociobiology. In this chapter, I examine the specific conceptions and modeling assumptions of social behaviour that engender the explanatory difficulties for modern evolutionary biology. One such assumption, which in fact is very fundamental in the modelling of social evolution, is the view of social behaviour simply as “other-regarding behaviour,” which we call “altruism.” The other fundamental problematic assumption is what Grafen (1984) has dubbed the “phenotypic gambit.” This is where genetic models assume social behaviour (or altruism) to be controlled by a pair of contrasting alleles at a single locus. We are to date still unable to satisfactorily explain social evolution on these modeling assumptions. I assert, however, that even if we found a way to successfully model the evolution of “other regarding behaviour” as a trait controlled by a pair of alleles at a single locus, we would not have explained the evolution of social behaviour, and this for two reasons: i) social behaviour is more than just “other regarding behaviour,” and ii) even if it were only that, it is highly unlikely that such a complex phenotype will be determined by a single Mendelian factor. Thus, all our current modeling endeavours are in pursuit of a theoretical phantom.

II. Modeling Conceptions and Assumptions

One of the key goals of my research is to argue that the difficulty of explaining social evolution stems largely from erroneous assumptions that form the foundation of the modern Darwinian evolutionary explanation. I will state the erroneous foundational assumptions first and then elaborate on each of them.

The following are the pillars upon which current explanations of moral/social evolution are constructed.

1. For the purposes of evolutionary modeling, the modern Darwinian paradigm regards social behaviour (and in fact, moral behaviour) simply as altruistic behaviour. The three concepts (sociality, morality and altruism) are applied synonymously in evolutionary explanation. Thus we explain the evolution of moral and social behaviour by explaining the evolution of altruism.
2. In explaining the evolution of altruism, we assume altruistic behaviour and selfish behaviour to be two contrasting phenotypes which are controlled by two contrasting alleles at a single locus.

In the rest of this chapter, I examine the implications of these two cardinal modeling assumptions for the evolutionary explanation of social behaviour.

II. 1. Does Altruism Encompass All Social Behaviour?

Forster (2009) defines a social behaviour as one that has a fitness effect on another individual. He identifies "four types of social actions based on their average effect on the direct fitness of the actor and recipient to include altruism, mutualism, selfishness and spite." Thus, if social behaviour includes these other forms, why do we talk of its evolution in terms of only one of the behaviours, namely, altruism? Wilson and Wilson (2007) also assert that "social behaviors need not be *pro*social; aggression fits the definition as does cooperation." Clearly, aggression, selfishness and spite are diametrically opposed to altruism as defined. So if social behaviour consists of such diverse and contrasting forms, wouldn’t a simplification that reflects only one of the diverse forms be a misrepresentation? Especially, if there is no claim that the other forms are insignificant.

Three key concepts that need to be clarified in this discussion are “altruism,” “sociality” and “morality.” Traditionally, evolutionary biologists have treated these as synonymous. In his discussion of social behaviour (as Wilson and Wilson (2007), for example, have called it), Darwin talks generally of the evolutions of morality and hardly uses the term “altruism.” Today’s evolutionary biologist uses the word “altruism” in all of Darwin’s mentions of :”morality.” On Forster’s analysis, to which I allude above, altruistic behaviour is only one of many kinds of social behaviour. Yet every model that sets out to explain moral or social evolution ends up explaining how altruism might evolve, without considering the other behaviours. Thus, under modern evolutionary explanation, the totality of social behaviour is modeled with the single concept of altruism (Bshary and Bergmuller 2007; West et al 2007). This may be justified to some extent, but it is also conceptually problematic. I will discuss the justification first.

There is one feature common to all three concepts (sociality, morality and altruism), which is of interest to evolutionary biologists, namely, that they all involve what we generally refer to as “other regarding behaviour.” This is behaviour that, to all appearances, seems to benefit another individual at a cost to the actor. Evolutionary biologists generally use the term “altruism” to describe all such acts. We see moral acts and social behaviour to be of this nature, hence our concern for their evolution. As Alexander (1987, 179-180) asserts, “the concept of morality implies altruism or self-sacrifice.” Our interest in such “other regarding” behaviour stems from the evolutionary paradox its prevalence presents. It is a predictive consequence of the theory of natural selection that a trait could not evolve in one species of organism if it is solely for the benefit of another (Darwin, 1859, 255). That is why other-regarding behaviour has been a challenge to model under the Darwinian paradigm. Thus, the evolution of morality and social behaviour become a matter of concern for evolutionary biology solely on account of their “other regarding” nature, and the interest of theorists in those subjects is focused exclusively on that particular feature.

Moral behaviour is seen to be “other regarding” because it represents a cost to the agent to forgo the advantage that he could gain in not behaving morally. Take the moral rules as specified by Gert (2004, 21)) for example:

1. Do not kill
2. Do not cause pain
3. Do not disable
4. Do not deprive of freedom
5. Do not deprive of pleasure
6. Do not deceive
7. Keep your promises

8. Do not cheat
9. Obey the law
10. Do your duty.

Or W. D. Ross's "prima facie moral duties," which are as follows:

1. Duties of fidelity
2. Duties of reparation
3. Duties of gratitude
4. Duties of justice and fairness
5. Duties of beneficence
6. Duties of self-improvement
7. Duties of non-maleficence

The agent accrues some advantage if he behaves contrarily (immorally) under any circumstance in which he is required to follow any of the above rules (apart, perhaps, from the "duty of self-improvement," whose inclusion as a moral rule has been questioned). People kill for advantage, they lie for advantage, cheat for advantage and so on for all the rest of them. A trait that would make an individual desist from taking these advantages, in the view of the evolutionary biologist, would put its bearer at a fitness disadvantage (prima facie) relative to others who have no compunction about violating those rules. It is that feature of morality that fits the evolutionary biologist's conception of *altruism* and makes moral behaviour seem essentially altruistic. Referring to Rifkin (2010), Zwick and Fletcher (2014) for example, write: "higher manifestations [of altruism] might simply be called ethical behavior or—focusing on proximal mechanisms—empathy." So we may grant that when evolutionary biologists use the word "altruism" interchangeably with the words "sociality" and "morality," it is in respect of the common feature they share, which is of theoretical significance to them, i.e. they *prima facie* entail a fitness cost to the actor. In fact, many philosophers also hold morality to require altruism. Doris and Stich (2012, 26) cite a few of them.

However, we are not completely non-discerning in our everyday usage of those concepts – altruism, morality and sociality. Even though they all have an "other regarding" component, there are many instances of "other regarding" behaviour that are neither moral nor social behaviour. For example, if a Canadian donates money to an orphanage in Nepal, the act is best described as altruistic rather than ethical, moral or social. No plausible moral system would require individuals or groups to sacrifice for external groups or individuals, especially those in faraway lands. Moral behaviour has more to do with not harming others (the so-called golden rule) than seeing to the positive advancement of others (except Ross' duty of beneficence, though failure to uphold that duty is not as morally repugnant as a failure in the duty of non-malevolence). It may be immoral to do nothing if it is within your means to prevent your neighbour's house from burning down. On the other hand, if your neighbour lives in a crumbling shack beside your mansion, there is no moral rule that says you ought to downsize your domicile in order to put up a liveable house for that neighbour. If you did that, it would certainly be altruistic, but not a moral requirement. Other than individuals who dedicate their lives to religious asceticism, hardly does any society find it practical to have a moral requirement that its members ought not to have surplus resource while there are others who have insufficient means. If there is a wealthy person who never gives to the poor at all, we may call him non-altruistic but not immoral, as long as he does not cheat, exploit or harm anyone. We do not consider people who do not give to the poor unethical. Moral norms appear largely to prohibit harm rather than impose responsibility for the positive welfare of others, which would be altruistic behaviour.

Social behaviour is best captured by the broader term cooperation, which requires different people to perform different functions or roles in order to accomplish a task. Cooperating with others in the community is not an ethical requirement. Some roles in cooperative actions may carry more cost than benefit to the agent. Evolutionary biologists regard such roles as altruistic. An extreme example of this is found in social insect societies, in which some members dedicate themselves to foraging and defending the nest, while others only reproduce. But cooperation has also been defined in a way that reductively equates with to altruism. Taylor and Nowak (2007, 2282), for example, write: “In a broader sense, ‘cooperation’ can be used as referring to an action where one individual pays a cost for another individual to receive a benefit." Doebeli et al (2004) also assert that "cooperative acts incur costs to the acting individuals, whereas benefits only accrue to others." Also, we have Boyd and Richerson (2009, 3283), who explain that they use "cooperation to mean costly behaviour performed by one individual that increases the payoff of others" (also see West et al, 2007 and references therein). Notice that these definitions of cooperation are exactly the usual definition of altruism in evolutionary biology.

However, a cooperative or social behaviour need not necessarily be altruistic. In a social club, such as a country club or fraternity, for example, membership is sustained by clear reciprocity rather than altruism. Nevertheless, evolutionary models still bring such situations under altruism as they view reciprocity as returning altruism for altruism. The point, however, is that a person joins such social groups for self-interested reasons, and would leave it if he thinks he is losing more than he gains in the group.

Thus, morality and sociality are not exactly congruent with altruism. There is more to social and moral behaviour than their other regarding feature. Altruism, construed as providing benefits to others at a cost to oneself, occurs in the context of organized group living. In other words, it is always found in a social context in association with other social factors. It is never all there is to any society (recall the other social actions Forster (2009) described at the beginning of this section). In fact, in situations where one organism simply benefits another at a cost to itself, in the absence of any other social factors, it is usually construed as “parasitism.” In their review of the subject, West and Gardner (2010) identify four kinds of social interaction – altruism, reciprocity, *green beard*[1] and spite. Other regular concomitant features of altruistic groups include approbation/disapprobation, sanctions and parochialism. As we shall see later in this discussion, the view is now nearly universal among scholars of social evolution that altruism, as standardly construed and defined, cannot possibly evolve in any group of organisms unless certain other mechanisms are also at play (Allen et al. 2012; Nowak 2012; Burton et al. (2012; Godfrey-Smith 2009; Taylor and Nowak 2007; Wilson and Wilson 2007; Hamilton 1964). Those other mechanisms, some of which I have mentioned above in this paragraph, should be considered together with altruism as integral parts of an organic unit – social behaviour.

The problem for researchers of social evolution comes about when they extract “other regarding behaviour” from its social and moral contexts, give it the label “altruism,” and try to explain its evolution as a genetic trait. In doing so, they purport to be explaining the evolution of social behaviour, but in reality they are not. Altruism, construed simply as “other regarding behaviour,” does not reflect the entirety of social behaviour. In the coming sections, we shall see the effect this narrow view of sociality simply as altruism has on the effort to model the evolution of social behaviour.

---

[1] Explained on page 27 below.

The fundamental theoretical concern of the evolutionary biologist regarding altruistic behaviour is “how [it might] evolve if the genes promoting it are at such a disadvantage in competition with genes that oppose it” (Wilson, 2005, 159). This statement leads us to the second of the two problematic modeling assumptions I identified above, that is, the genetic view of altruism.

III. The Fundamental Genetic Modelling Assumption

Wilson (1975, 3) has stated that “the central theoretical problem of sociobiology [is] how can altruism, which by definition reduces personal fitness, possibly evolve by natural selection?” Almost every article or book today that aims to explain the evolution of social behaviour is motivated by this perceived puzzle, which has now spilled over into psychology and philosophy. The crux of the puzzle comes through in Wilson’s (2005, p. 159) alternative, or more precisely his phrasing of the problem thus: ‘‘how might such a behavior evolve if the genes promoting it are at such a disadvantage in competition with genes that oppose it?” The puzzle that social behaviour poses for natural selection is a problem that Darwin (1859; 1871) himself dealt with as a special case that required a modified view of natural selection. The new (genetic) rendition of the problem is simply the same one Darwin confronted, but in a new incarnation triggered by the Mendelian genetic character of the modern synthesis. This is why some modern scholars (Ratnieks and Helantera 2009; Michod 1982; Axelrod and Hamilton 1981) view social evolution rather inaccurately as a hitherto unknown problem that was first discovered and solved by Hamilton. Indeed, it was Hamilton (1964) who first called attention to the fact that social evolution was not possible under the modern Darwinian explanation that was based on Mendelian genetics. At the same time, he proposed his inclusive fitness hypothesis as a way to account for social behavior under Mendelian genetic principles.

The implication of Hamilton's explanatory gambit, as I explained in the previous chapter, was that the prior explanation for social evolution – Darwin's hypothesis of group selection – was to be rendered obsolete. As it happens, Hamilton had attracted some following, and that movement worked to assert the supremacy of the Hamiltonian system over the traditional Darwinian system. This they did by discrediting group selection and emphasizing the gene as the most important unit of selection.

This marginalization of group selection was facilitated by the fact that its articulation and application had been altered in several significant ways by its later adherents. For example, in explaining eusocial evolution under group selection, Darwin (1859) conceived of the group as a family of genetically related individuals. This provided a basis for explaining the maintenance of eusocial behaviour within such family groups. Consider the following conception of eusocial group selection by later adherents as expressed by Wilson and Wilson (2007, 340): "Most traits associated with eusociality [for example] do not evolve by increasing in frequency within colonies, but by increasing the colony's contribution to the larger gene pool." This view of group selection seems to be pointing to the numeric increase in the number of altruists due to the expansion of their home group, as the mechanism for the evolution and maintenance of altruism under group selection.

That view of group selection, however, has never been able to adequately address the fact that the frequency of altruists within the home group will actually tend to fall due to the lower fitness of the trait. This theoretically predicted progressive diminution of altruistic individuals within social groups has come to be known as the “within group selection problem,” and the inability of contemporary group selection models to effectively deal with this theoretical problem has greatly undermined them. However, the same cannot be said of Darwin’s model of group selection. In eusocial evolution for example, Darwin supposes the group to consist of a family of individuals all of whom share the altruistic genotype. So that obviates the problem of lower relative fitness of the altruists within the group, which is the basis of the predicted fall in the frequency of altruists.

Thus, the modern conception and application of group selection are less judicious than Darwin’s. The groups in Darwin’s (1859) eusocial model were family groups, usually a mother and her offspring, whereas the individuals in a group under modern group selection need not be related. So Darwin’s group selection was more consistent with what would today be called “multi-level selection” than pure group selection. Also, Darwin (1871) recognized the problem of “within group” selection in groups of non-related individuals, and offered credible mechanisms for the evolution and maintenance of altruism within such groups, rather than suggesting that altruism is maintained by group level advantage, as later models do.

Darwin's family-group concept has a greater explanatory power than even Hamilton's inclusive fitness hypothesis in which the group consists of random individuals, some of whom may be kin, and in which case some of the benefits of the altruistic behaviour may still fall on non-kin. The "inclusive fitness" is thus probabilistic and therefore weaker than Darwin's.

Another way that group selection was rendered vulnerable to attack was the very poor case for group selection that was put forward by Wynne-Edwards (1962) in his book entitled *Animal Dispersal in Relation to Social Behaviour*. As Forster (2009) points out, it is a book replete with poor examples and arguments for population level selection. Wynne-Edwards (1962) argued for example, that organisms evolve altruism in order to control their population growth so as to avoid overpopulation.

As a consequence of these circumstances and events, modern evolutionary explanation is now centered essentially around population genetics, which has been described as the cornerstone of the modern synthesis (Okasha 2006), and gene frequency change has been described as the "hard core" of modern evolutionary theory (Michod 1982, 25), even though prior to the 1960s population genetics was only one of several components of the synthesis (Mayr 1993).

The modern conception of evolution as a change in gene frequencies commits us to explaining the evolution of phenotypes by tracking underlying genotypic changes. Lynch (2007, 8597) quotes Carroll (2005) as saying that the view of evolution under population genetics as change in gene frequencies "forces the explanation toward mathematics and abstract descriptions of genes, and away from butterflies and zebras or Australopithecines and Neanderthals." Recall from the discussion in chapter one that the modern synthesis consists of Darwinian naturalism and abstract mathematical genetics in a strained relationship. We discussed the turn of events in the 1960s, which tipped the balance away from Darwinian naturalism and towards the abstract mathematical modeling of evolution under population genetics. Thus, at the very foundation of contemporary evolutionary explanation are abstract reductionist mathematical assumptions that ramify extensively into most contemporary explanatory models, including some that would claim to shun reductionism.

The basis of the pervasiveness of the reductive genetic view of social behaviour is that when Darwinism, after initial resistance, ultimately adopted Mendelian particulate genetics, under the so-called modern synthesis, a binary conception of genotypes as well as phenotypes (as Van Veelen et al. (2012, 68) term it) emerged as its fundamental heuristic. Thus, according to Grafen (1984, 64), in modeling the evolution of behaviour (altruism, for example), the behavioural ecologist takes a leap of faith and goes with a "phenotypic gambit" that there is an allele for one phenotype and a contrasting allele for the other.

Thus, we suppose an allele for altruistic behaviour and then set as its binary alternative, selfishness, which represents any behaviour that is non-social or non-cooperative.

This is how the two-allele, single locus assumption has come to be standard in contemporary genetic modeling of social behaviour. Because of the minimal contact this standard modeling assumption has with the empirical data (Yakubu 2013; Carroll 2005; Wade 2005; Grafen, 1984) and the way it largely caricatures reality (Grafen 1984; Gardner et al. 2011), many scholars these days do not want to be explicit about the role of this foundational assumption in their models. Nevertheless, there are others who are quite plain about that modeling principle. Cavalli-Sforza and Feldman (1978, 268), for example, explain their modeling assumptions, saying that "in these models altruistic behavior is determined at a single locus with alleles A_1, and A_2. The altruistic phenotype is denoted by A while N represents the non-altruistic or selfish phenotype." Trivers' reciprocal altruism model sets the following as the genetic assumption in the model: ''Assume that the altruistic behavior of an altruist is controlled by an allele (dominant or recessive), a_2, at a given locus and that (for simplicity) there is only one alternative allele, a_1, at that locus and that it does not lead to altruistic behavior" (Trivers 1971, 36). Bowles (2006, 1569) supposes in his group selection model that ''(A) individuals are bearers of a hypothetical 'altruistic allele'; those without the allele (Ns) do not behave altruistically.'' Other references to the contrasting altruistic and selfish alleles at a single locus include Haldane's (1932, 208), Rousset and Roze (2007, 2321), Sober (1984, 184).

The conception of evolution as a change in gene frequencies puts us in a modeling predicament with regards to social evolution for two reasons. In the first place, with behavioral traits we do not know what genes are involved or in what manner they relate to the phenotype. As Queller (1992, 557) notes, "in the real world, the ps and qs of population geneticists are not usually available to us; we do not know the frequencies of genes affecting social behaviours." So it is only a leap of faith, as Grafen (1984, 63) concedes, when we assume that "there [is] a haploid locus at which each distinct strategy was represented by a distinct allele." This kind of modeling latitude, of course, has no empirical justification, and that brings us to the second part of the predicament; that is, even if we could identify every pair of alleles and every locus involved, for a complex behavioural trait like social behaviour there would be scores of them, if not hundreds, and that renders our current Mendelian factorial modeling mathematically unwieldy to the point of impossibility.

The complexity introduced just in extending the Mendelian model from one-locus to a two-locus selection is enormous. As Halliburton (2994, 87) laments, "If we cannot comprehend the potential complexity of two-locus systems, how can we hope to understand the interactions of hundreds or thousands of loci?" Halliburton (2994, 87) also notes that describing evolution by gene frequency changes is useful only where multi-locus interaction is minimal. If that is the case, why would anyone be modeling a complex trait such as altruism, which is most likely controlled by scores of loci, on gene frequency changes? Indeed, Grafen (1984, 65) recognizes that "the sorts of character studied by behavioural ecologists are likely to be controlled by many loci." In fact, "one approach to multi-locus population genetics is to (temporarily) abandon the explicitly genetic approach and to study the statistics of phenotypic variation" (Halliburton 2004, 487). This suggests a quantitative genetics approach to such multi-locus traits. That is why there has been a proliferation of quasi-genetic models, especially in evolutionary game theories, which talk largely of phenotypic strategies while leaving their genetic basis ambiguous.

As Queller (1992) notes, however, if we want to relate our models to the real world, we need to make the genotype-phenotype-fitness connection explicit, as in the basic selection equation of quantitative genetics:

$$R = Sh^2$$

Where, R, the evolutionary response to selection, is the product of a selection differential S, which describes the relationship between phenotypes and fitness, and heritability h^2, which describes the relationship between phenotypes and genes. Queller suggests we can do this by observing phenotypes and estimating fitnesses, heritabilities, and relatednesses. Grafen (1984, 64) also suggests that "it is essential to know the underlying genetics in order to understand the distribution of phenotypes observed in a population." At the same time, Grafen (1984, 65) explains, however, that even though the behavioural ecologists depends on population genetics for the soundness of their models, their “methods are designed to avoid doing genetics.” Consequently, he concedes that “the behavioural ecologist, though, does not usually know the genetics underlying the character he studies” (1984, 63).

I am also quite certain that we may never get a statistically significant estimate of the heritability (h^2) of altruism in empirical research. The available data is overwhelmingly contrary to the standard modeling conception of social populations as consisting of two contrasting genetic factors that express altruism and selfishness (Yakubu 2013). In that work, I suggested instead that the field research data point to altruism and selfishness as dimorphic phenotypic expressions of a single flexible genotype. If that is the case, empirical estimates of the heritability (h^2) for altruism should be close to zero. However, as we learn from Grafen above, we are not interested in knowing or working with this fact. We do not bother to make any empirically informed estimates about genotype-phenotype relationships. Instead, as Grafen (1984) explains, we take a leap of faith, and proceed as if there was a haploid locus at which each distinct phenotype is represented by a distinct allele. The evolutionary biologist ought to know that such an assumption could not be consistent with the genetics of a trait as complex as social behaviour. However, it is a well-known feature of theoretical models that they often take a flight of fancy into unrealistic abstractions. As Wade (2008, 2) describes it, “population genetics tends toward a branch of applied mathematics founded on Mendelian axioms, often with minimal contact with data." Depew (2013, 184) on the other hand has explained that Theodosius Dobzhansky, one of the architects of the modern synthesis, "thought of population genetics as empty mathematical exercises until and unless they are shown by experimental and field evidence to solve the real biological problems of naturalists." But how can our models solve real problems in nature if we build them on assumptions that have no bearing on the reality of nature?

IV. Modeling Disarray

The discussion so far has established two problematic fundamental assumptions in the genetic modeling of social behaviour. The first is the reduction of social behaviour to altruism, and the second is that altruism and selfishness are controlled by two underlying contrasting alleles. The discussion also showed how unrealistic the assumptions are. We shall see in further discussion below that it takes more than sacrificing or incurring cost for the welfare of others to sustain moral behaviour and maintain social cooperation. These two problematic assumptions at the core of modern evolutionary explanation have made a genetic account of social behaviour impracticable. As a consequence, there has been a palpable reticence among scholars these days about making explicit genetic claims in their models of social behaviour.

Genetic models have remained largely abstract, and adherents to that approach are constantly fighting off charges of genetic determinism. Krebs (2013) has written at length about this problem. The usual and now trite defence has been that "notions of genetic determinism—behaviors caused by genes without input or influence from the environment—are simply false" (Buss 2004), and that "kin selection theory [for example] does not deny the truism that all traits are affected by both genes and environment" (Okasha 2009). Wilson and Wilson (2007, 331), similarly, explain that the early group selection models assumed that altruistic and selfish behaviors to be "caused directly by corresponding genes, which means that the only way for groups to vary *behaviorally* is for them to vary *genetically.* Hardly anyone regards such strict genetic determinism as biologically realistic." He adds.

Some have suggested that the persistence of the genetic determinism charges, in spite of the historical string of rebuttals, is "because the genetic determinist is too convenient a straw man to be discarded" (Alcock 2001, 44). However, the standard rebuttal to the charge of genetic determinism has also been superficial. It commits the same offence it accuses its critics of, by targeting a straw man version of the charges and avoiding the kernel of the problem. Buss's (2004) and Okasha's (2009) denial of genetic determinism, for example, are typical. Declaring that "we do not deny that all traits are affected by both genes and environment" buries the bone of contention under vague genetic talk. The issue is whether genetic difference contributes in any way to individual differences in altruistic behaviour in a social group, as expressed by Michod (1982, 25), who writes: "In the genetic approach, differences in social behavior are assumed to result, in part, from differences in genotype". If you hold this belief, even together with the belief that the environment also contributes to the phenotypic difference, then you do believe in some form of genetic determinism of behaviour. There is a clear difference between this position and that of those who criticize the position. The critics simply do not believe that genes play a significant role in individual differences in altruistic behaviour in any social population.

As Jablonka (2004, 354) explains, what biologists mean when talking about “a gene for trait X” is the “genetic basis of X” and when geneticists talk about “genes for” they are talking about genetic differences that make a difference to the phenotype. “No geneticist thinks about a gene for eyes, nose or intelligence.” The following two examples will illustrate the position further.

Take human height for example. We all agree that it is a “genetic trait” which is also affected by the environment. We know that taller parents tend to produce taller offspring, but at the same time nutritional differences can cause an average variation of about 4 inches in human height. Is the sociobiologist saying that altruism is a genetic trait in the same way that human height is, such that altruistic parents will tend to produce altruistic offspring, and the environment adds or subtracts a few shades of altruism? If this is the gene-environment model of behaviour the sociobiologist has in mind when he thinks it is unfair to accuse him of genetic determinism, I would like to inform him that yes, it is that genetic notion of altruism his critics are opposed to and call determinism.

Now consider the alternative scenario I gave in Yakubu (2013) to illustrate the issue. There are two phenotypic forms of the water flea *Daphnia lumholtzi*, one with a helmet and a long tail spine, while the other morph lacks both of those features. In experiments using clones of *Daphnia*, Agrawal (2001) has demonstrated that whether an individual grows into one morph or the other depends on whether predacious fish are present in the environment in which they grow. Yakubu (2013) then argued, citing field research data on altruism, that the expression of the altruistic and selfish phenotypes is more consistent with that of the two morphs of *Daphnia* than the human height example above. Queller and Strassmann (2012) point out, for example, that “the division into sterile and reproductive castes is almost universally based on differences in the physical, nutritional and social environment” (also see Wheeler 1986). No empirical study has ever demonstrated genetic difference (with influence from the environment) as the basis for altruism in a social population.

If that is the case, then we cannot model altruism and selfishness as competing contrasting alleles, as current genetic models do. If the sociobiologist is opposed to this latter genetic view of altruism and favours the former, then he would require empirical support for that position rather than rhetorical denials of genetic determinism.

This quagmire around the genetic modeling of social behaviour has led to a wave of models that emphasize phenotypes rather than genotypes. Some scholars reason that genetic modeling is not the way to go for social behaviour and therefore urge non-genetic alternatives. Campbell (1978, 52) for example, suggests that:

> [I]n man, genetic competition precludes the evolution of . . . genetic altruism. The behavioral dispositions that produce . . . self-sacrificial altruism must instead be products of culturally evolved indoctrination that has had to counter self-serving genetic tendencies.

This seems to suggest that we do not genetically evolve altruism. Instead, we remain genetically selfish in our nature and altruism is acquired through cultural indoctrination. If this were the unequivocal stance of evolutionary biologists, there would not be as much controversy on the issue as there is today. The reality is that neither kin selection nor reciprocal altruism, the most widely accepted explanations of sociality today, reflect this sentiment. Even game theory is largely vague on this

On the contrary, however, the ESS/game theoretic models, the early ones in particular, explicitly cast the competing strategies as genetic strategies (Maynard-Smith 1972; 1982; Dawkins 1989; Axelrod and Hamilton 1981). In fact, that is what the phenotypic gambit (Grafen 1984) unambiguously declares. However, following heavy criticism of such a view, including charges of genetic determinism, there has been some pullback from that position. But to what? Many, for example, now tend to talk loosely of the interactions of phenotypic strategies, leaving the genetic picture vague. Some of the recent models talk of both heritable and non-heritable components superficially, without any indication as to which is what and how. Taylor and Nowak (2007, 2282), for example, explain the transmission of phenotypic strategies in their models, saying that "reproduction can be genetic or cultural. Genetic reproduction means that individuals leave genetic offspring that inherit their strategy. Cultural reproduction means that individuals are imitated by others; thereby strategies reproduce by imitation or learning.” Similarly, Sigmund and Nowak (2003; 2005, 1292) have noted that “in evolutionary game theory it is not assumed that players are rational but only that successful strategies spread — by being inherited, for instance, or copied through imitation or learning.”

However, these models are vague as to whether every strategy is sometimes heritable and copied at other times or whether each strategy is partly heritable and partly copied or, even further, whether some strategies are by nature heritable and others always copied. As (Bshary and Bergmuller 2007, 409) explain, “we do not distinguish between genetically determined strategies and learned strategies (tactics) because [among other things], we usually lack information on how genes and learning interact to produce behaviour in a specific situation.”

Thus, these scholars cleverly avoid the charge of genetic determinism by claiming to be discussing only phenotypic traits or strategies, with no claim for any genetic association. However, we cannot talk of the evolutionary fitness of one phenotype over another, as such models do, without implying differential reproduction in some underlying associated heritable factors. As Gardiner et al (2011, 1021, and 1023) note, the character of interest in natural selection is not the phenotype per se, but the “change in the heritable component of the phenotype.” Thus, unless we know the exact genotype-phenotype relationship, a "phenotypic gambit" of some sort is unavoidable. So unless a model specifies otherwise, we would have to take it that each of the competing phenotypes it describes is independently heritable.

Also, before an organism can have the capacity to copy and express a trait wouldn't the trait have to somehow evolve in the organism first? Not every organism can copy altruism, reciprocity or TIT FOR TAT. So those that can, must have some inherent (evolved) capacity to do so. Given such a scenario, we can infer further that if a trait spreads by imitation and learning, then all individuals in the population who are capable of learning the behaviour have the genetic capacity to do so, and if it is inherited, but is such that as far as we know any individual in the population has the capacity to express it under the appropriate circumstance, then the gene is present in every individual in the population. So how do we model the underlying genetics for the evolution of such a trait? Can the evolutionary model assume an allele (or genotype) x for individuals in the population who exhibit the trait, and genotype y for those in the population who do not? That is the challenge facing the *phenotypic gambit* assumption under the modern synthesis.

There are some scholars who still discuss the game theoretic models in terms of the genes associated with the phenotypic strategies. Krebs (2012, 62) writes that "neo-Darwinian evolutionary theorists have found it helpful to model the selection of social strategies as games in which players who possess different genes, or more exactly alleles that induce them to adopt different strategies, compete against one another to win fitness-increasing points." He adds: "In evolutionary games, winners contribute more copies of the genes that code for their strategies than losers do, and thus more copies of their strategies, to future generations, which affects the distribution of strategies in the population." Even more explicitly, Gardner et al. (2011, 1029–1030), in setting up the assumptions for their game theoretic analysis, write: "We assume an infinite population of haploid individuals engaged in two-player games. A single locus controls the cooperation phenotype, with a proportion p of individuals carrying an allele A which encodes the cooperator strategy, and the remaining 1 – p carrying an allele a, which encodes the non-cooperator strategy."

As a matter of fact, we cannot talk of evolution solely in terms of phenotypes without reference to how they ultimately relate to genotypes, since a phenotype cannot evolve unless it is heritable. As Queller (1992) notes, if we want to relate our models to the real world, we need to make the genotype-phenotype-fitness connection explicit. After all, isn't it the case that Darwinian natural selection applies only to phenotypic traits that are heritable?

Unless maybe those ESS theories are not Darwinian, for if they are, Darwin, in talking about evolution by natural selection, makes it clear that, "any variation which is not inherited is unimportant for us" (1859, 31). In any case, how is a behavioural strategy evolutionarily stable if we do not assume it to be transmittable by an underlying heritable factor? Taylor and Nowak (2007, 2282), for example, refer to the evolutionary game strategies as phenotypes, but they indicate that natural selection favours the higher payoff strategies through reproduction (which they specify may be genetic or cultural). Even though a genotypic dichotomy is not explicitly asserted, it does not take much to show that it is their basis of all the considerations.

In his seminal paper on *Evolutionarily Stable Strategies* (ESS), Maynard-Smith (1972, 21) explains that “a strategy qualifies as an ESS if, in a population in which most individuals adopt it, there is no alternative strategy which will pay better.” Maynard-Smith goes on to explain what “pay better” means: “The utility of an outcome is simply the contribution that outcome makes to the fitness of the individual – that is, to the expected number of future offspring born to that individual” (21). So the pay-off of each strategic phenotype is cashed out in terms of transmittable genetic fitness that underlies that phenotype. Ultimately, therefore, the game theoretic models assume the strategic phenotypes to have corresponding genotypes. It is just not to be mentioned.

V. The Genetics of Conditional Altruism

There are three broad explanatory models of social evolution in the Darwinian tradition. They are: *group selection* which originates from Darwin himself (1859, 1871), *kin selection* (Maynard-Smith 1964), which derives from *inclusive fitness* (Hamilton 1964), and *reciprocal altruism* (Trivers 1971). These constitute the “primary colours” of traditional evolutionary explanation. All other models have roots in these three, and are therefore secondary or derivative models. They include, among others, models such as *cooperation*, which is now essentially the *game theoretic/ESS* models (Axelrod and Hamilton 1981), *indirect reciprocity* (Nowak & Sigmund 2005), *strong reciprocation* (Gintis et al), and *gene-culture coevolution* (Boyd and Richerson 2009) etc. These secondary (derived) models are for the most part tweaks and re-blends (using several aspects) of the primary models.

Of course, there are other ways contemporary evolutionary explanation may be compartmentalised. Bshary and Bergmuller (2007), for instance, render the following categories: (1) basic social evolution theory, which explores the evolutionary pathways that select for helping. This is largely the inclusive fitness (kin selection) approach. (2) The ecological approach, he explains, still focuses on direct and indirect benefits as mechanisms for the evolution of helping, and is based on life history parameters. (3) The game theoretic approach, which ‘translates’ ecological conditions into a game structure (n interactions, payoff matrix), and mechanisms such as reciprocity and punishment. (4) The social scientists’ approach, he explains, identifies psychological and physiological mechanisms that promote helping in humans. Lately, this latter method has come to involve quite a lot of social and moral neuropsychology. The first three of Bshary and Bergmuller’s categories are quite congruent with my three primary categories. Their fourth category, as they indicate, is outside the purview of evolutionary biology, and is more of an endeavour from the social sciences. This points to the problem of disciplinary silos in contemporary evolutionary explanation. Past scholars such as Hume and Darwin had greater successes than any of today's splintered models because of their cross-disciplinary approach.

With the exception of pure group selection, nearly every other model requires some conditionality in the expression of the altruistic trait in order to be efficacious. This however ought not to be misconstrued to suggest that group selection is successful. It is not, and this is primarily due to the evolutionary instability of the altruistic trait within groups. In fact, since the 1960s, group selection as a standalone explanation for altruism has fallen into disfavour. Instead, it is now more often proposed as part of multilevel selection models, which ultimately involve conditional altruism. The multilevel selection models that Boyd and Richerson's (2009) propose, for example, feature kin selection within the groups under group selection. It appears, then, that there is no significant model of social evolution that does not rely on conditional altruism in some way. Conditional expression of the altruistic trait therefore seems to be key to the success of any model under the current evolutionary paradigm. In fact, theoretical analysis suggests that "because unconditional selfishness is always a potential player in evolutionary games, we must conclude that no species will ever evolve to behave in unconditionally altruistic or unconditionally cooperative ways" (Krebs 2012, 130).

Thus, the models generally proposed to resolve the problem of social evolution mostly hinge on the capacity of the altruistic gene to flexibly express both altruism and selfishness (the so-called conditional altruism). Wilson and Wilson (2007, 331) suggest that one of the reasons why group selection failed in the past is that “all of the early models assumed that altruistic and selfish behaviors are caused directly by corresponding genes, which means that the only way for groups to vary *behaviorally* is for them to vary *genetically.* Hardly anyone regards such strict genetic determinism as biologically realistic, and this was assumed in the models primarily to simplify the mathematics.” Alger and Weibull (2012, 42) have suggested that Hamilton’s rule is ineffective if a behavioural trait “is an action always to be taken” (i.e., unconditional), as opposed to the situation in which the trait expresses one behavior or another contingent upon some exogenous factor.

However, this often invoked “conditional altruism” is inconsistent with the two-gene model of altruism and selfishness that evolutionary theorists of today tend to favour. The phenotypic flexibility required under conditional altruism is more consistent with altruism and selfishness as plastic expressions of a common genotype, and therefore, a case of *phenotypic plasticity*. If that is so, then the genetic difference between the altruistic and selfish individuals dissipates, and the concern about the unsustainability of the altruistic trait within social populations is allayed.

VI. Conceptual Reorientation

All the elements for a successful model of social evolution are present in the current theoretical disarray. The best theoretical direction now is to sift through the jumble, isolating the useful elements and constructing a robust model that is independent of the existing explanatory traditions. It is now a truism of Darwinian evolutionary modeling that altruism, simply by virtue of its definition, cannot evolve by Darwinian natural selection (Cavalli-Sforza and Feldman (1978, 268; Wilson 1975, 3). The following statement from Cavalli-Sforza and Feldman (1978, 268) summarizes the perception of altruism under the neo-Darwinian paradigm and why its evolution is a paradox. They write: “The theory of the evolution of genetically determined altruistic behavior faces, at the outset, the central problem that altruists are, by definition, at greater risk to their lives than non-altruists and are therefore in greater danger of elimination by natural selection.” Nesse (2000, 228), explains that when the logic of natural selection “is combined with our intuitive notion that altruism consists of costly acts that benefit others, and genes are seen as the ultimate currency, then altruism is impossible.”

The need to provide some countervailing mechanism to this theoretically predicted depression of the frequency of altruistic individuals within social groups is the driving force of modern social evolutionary theory today. Hamilton (1964, 1), in his pivotal paper declared that the existing mathematical models did not allow for the evolution of sociality and consequently, that some mechanism was needed that would offset the apparent decline of the altruism allele. Wilson and Wilson (2007, 329) assert that ‘‘something more than natural selection within single groups is required to explain how altruism and other group-advantageous traits evolve by natural selection.’’ Similarly, Taylor and Nowak (2007, 2281) state, from the view point of ESS models, that “cooperation is always vulnerable to exploitation by defectors; hence, the evolution of cooperation requires specific mechanisms, which allow natural selection to favor cooperation over defection’’ (see also Allen et al. 2012; Nowak 2012; Taylor and Nowak 2007, 2282).

Thus, there is nearly a perfect consensus among scholars that something other than mere altruism must be going on in a society for altruistic behaviour to be sustainable. What is it? A compendium of some of the broad mechanisms that have been suggested is provided by Yakubu (2013). They include: (1) altruists associating exclusively with other altruists (Maynard-Smith 1998; Sober and Wilson 1998); and (2) conditional deployment of the altruistic behavior, i.e., only towards genetic relatives (Hamilton 1964), or towards other altruists (Trivers 1971). Dawkins (1976/1989, 89) introduced the metaphorical “green beard,” in which we are to imagine altruists identifying other altruists by a characteristic green beard. For group selection models, Godfrey-Smith (2009, 174) explains that if social groups are formed randomly, ‘‘the A (altruistic) type is lost regardless of the details.’’ However, the altruist can be maintained, he explains further, ‘‘if groups are formed in a way that ‘clumps’ the two types, so like tends to interact with like [and] the benefits of having ‘As’ around tend to fall mainly on other As’’

(174). For ESS game-theoretic models, Burton et al. (2012, 55) state: ‘‘one general answer is that interactions need to be assortative, so that individuals carrying genes coding for cooperation interact, on average, more often with cooperating individuals than individuals carrying genes coding for defection.’’ Similarly, ‘‘clustering’’ is called for in the latest ESS models using ‘‘evolutionary graph theory’’ so that ‘‘the benefits of cooperation are received mostly by other cooperators’’ (Allen et al. 2012 and references therein).

Prior to Hamilton’s seminal work, group level selection was generally considered to be the counterbalancing mechanism to the theoretically predicted decline of altruism within groups (Wright 1945; Williams and Williams 1957). This stated with Darwin (1871, 159) who wrote:

> It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and an advancement in the standard of morality will certainly give an immense advantage to one tribe over another . . . At all times throughout the world tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase.

In fact, proponents of group selection today still maintain that claim. Wilson and Wilson (2007) for example, write:

> From a multilevel evolutionary perspective, traits that cause an insect colony to function as an adaptive unit seldom increase in frequency within the colony and evolve only by causing the colony to out-compete other colonies and conspecific solitaires, either directly or through the differential production of reproductive” (341). He adds: “Slackers are fit than solid citizens within any single

> colony, but colonies with more solid citizens have the advantage at the group level" (341).

The idea of group level fitness as a counterpoise to the declining frequency of altruistic individuals within social groups remains unsatisfactory as an account for social evolution.

The other proposed mechanism is Hamilton's (1964) inclusive fitness hypothesis, which suggests that the altruistic trait may be maintained in a population if the benefits of the altruism fall disproportionately on genetic relatives, who are more likely to also carry the altruistic trait. This too has not been quite satisfactory (Nowak et al 2010). Besides, Hamilton's model, which has also been dubbed "kin selection" (Maynard-Smith 1964), would be of a limited scope, if relevant at all, because a substantial number of cases of altruism do not involve genetic relatives.

Consequently, Trivers (1971) proposed *reciprocal altruism*, which holds that altruistic behaviour may be maintained in a society if altruists direct their altruism only towards individuals who will reciprocate the altruism, essentially, fellow altruists. There have been a number of spin-off models from this. They include a whole class of models called *Evolutionarily Stable Strategies (ESS*) (Maynard-Smith 1982). This class of models have alternatively been referred to a *Game Theoretic* models, or *Evolutionary Game Theories*. Other spin-off models from Trivers' *reciprocal altruism* include direct reciprocity, indirect reciprocity (Alexander, 1987; Leimar & Hammerstein, 2001; Nowak & Sigmund, 1998), network reciprocity (clustering, or graph selection) (Taylor and Nowak 2007, 2282) as well as strong reciprocity.

Each of these models postulates an inducement or regulatory mechanism for altruistic behaviour in societies. Thus, altruism does not evolve if these regulatory mechanisms are absent and the altruistic phenotype is expressed automatically upon the presence of the altruistic gene (Alger and Waybull; Wilson and Wilson 2007). This is why the altruistic gene has to be flexible and phenotypically plastic, so that it is capable of expressing the phenotypic alternative (selfishness) under some conditions, especially conditions that do not favour altruistic behavior.

The requirement of these regulatory or triggering mechanisms for the expression of altruism suggests that social cooperation is sustained by several interacting but independent instincts or traits, rather than just a single trait that is other-regarding. Secondly, the flexible expression of the trait probably means that regulatory genes are also involved, which in turn suggests the involvement of several gene loci. This is what overturns the two fundamental modeling assumptions of contemporary revolutionary biology, which were introduced as the subject of this chapter.

Thus, the modeling difficulties of today’s evolutionary explanation are to be expected because we take phenomena as complex and multifaceted as sociality and morality, reduce them to a single one dimensional trait called altruism and assume it to have the same inheritance mechanism as simple traits such as seed colour and seed shape in peas. In chapter one, I argued that stubborn scientific puzzles are usually the result of erroneous axiomatic assumptions deep in the core of scientific paradigms. In this chapter we have identified the two deep-seated assumptions that are responsible for the paradox of social evolution. In the next chapter, we will examine David Hume's thought on morality and social evolution and see what insight we may find therein to help us deal with this Darwinian problem.

Chapter Three

Insights from Hume's Moral/Social Theory

I. Introduction

The classical textbook moral theories are: Aristotle's virtue ethics, the utilitarian ethics of Jeremy Bentham and J. S. Mill, Immanuel Kant's deontological ethics, W. D. Ross' ethical pluralism, and lately feminist ethics. All these approaches to morality are characteristically *prescriptive*, in that they tell us *what ought to be*, rather than *what is*, which would be *descriptive*. Hume is often not on this list because his moral theory is descriptive. It tells us the way things are, rather than the way we ought to make them be. It is the *naturalistic* explanation, the method of science. In that system, we describe phenomena in nature and in some cases propose explanations (hypotheses) as to how such phenomena have come to be. That is exactly Hume's approach to morality.

Kant, in the preface to his *Metaphysics of Morals*, distinguishes between two broad systems of knowledge. He explains that whereas natural philosophy describes "laws according to which everything does happen," moral philosophy has to do with "laws according to which everything ought to happen." Thus, Kant perceives moral principles to be in the normative domain. On the other hand, Hume, in his approach, explains how "a sense of morals" is generated out of the original constitution of the human mind, rather than how we ought to act. He looks at what man is naturally inclined to praise as virtuous or condemn as vice, and he then attempts to explain how man came to be of that disposition. This places Hume's theory of morality among the former of Kant's philosophical systems.

Hume's project is to explain man's moral and social nature, which he does in a very detailed and systematic fashion. Many of the ideas Hume uses as building blocks for his grand picture of moral/social behaviour may have also been asserted by prior philosophers. However, it is inaccurate, and in fact unfair to Hume, to characterize these points of commonality simply as influences or borrowings from his predecessors, because Hume doesn't just borrow these ideas; he develops them from first principles, often using more systematic and clearer original arguments than his predecessors. In fact, his analyses result in conclusions contrary to his predecessors as often as they might concur. Also, as we shall see in the ensuing discussion, Hume's analyses often reveal deeper foundations of human nature, which often show his predecessors to be on opposing sides of a false dichotomy. These are often the same predecessors many assert Hume to be borrowing from or heavily influenced by. There is a general tendency among scholars to exaggerate Hume's indebtedness to those prior philosophers, which is based on nothing more than the temporal order of the discoveries.

The quality of Hume's analysis that caught my interest is the unique and systematic picture of human sociality that he paints. The key insight in his analysis that is pertinent to my project is his clear delineation between the natural and artificial virtues, which correspond, respectively, to natural social bonds based on empathy on the one hand, and on the other, artificial virtues such as justice that make civil societies possible. On this basis, there are two kinds of human social groups, which are held together by different mechanisms. As we see here and in the next chapter, this two-tier explanation is the key to resolving the Darwinian theoretical difficulty regarding social evolution.

As we shall also see in this discussion, most other philosophers, unlike Hume, take the social virtues to be innate in humans. That social virtues are innate (genetic) human traits is also a view held by the modern Darwinian. However, unlike many of the ancient and early modern philosophers who simply assumed such social virtues to be implanted by God, Darwinians think they evolved by natural selection, but they have had a difficult time providing a convincing mechanism for that. Hume's insight that the social virtues are artificial rather than natural could provide the way for the Darwinians to get out of the theoretical quagmire. That is what my thesis aims to demonstrate.

One other relevant insight of Hume is his bold restoration of morality to emotion by wrestling it from its popularly exalted status as a product of rationality. The ideas Hume asserts as premises for his major insights often have precedents and sometimes deep histories. This project is not about those "building block" ideas and will not get mired in their histories and philosophical personalities. Instead, this project aims to move forward with Hume's key insights as possible contributions towards a solution to a contemporary theoretical puzzle.

My obligation under argumentation theory is to provide additional support only for those of Hume’s premises that are controversial. Nevertheless, corroborations, even for uncontroversial premises, hardly hurt an argument. Thus, I will mention philosophers whose views corroborate Hume’s “building block” ideas when they are readily available.

II. Terminological Map

The totality of Hume’s analysis of human nature is confined to the mind. So a layout of the architecture of the Humean mind will be didactically expedient. *Mind* in this discussion will always refer to the human mind unless otherwise specified. Hume uses the word *perception* to mean anything that registers on the mind – pain, pleasure, hunger, anger, thirst, love, hate, memories etc. Hume divides “all the perceptions of the mind into two classes or species, which are distinguished by their different degrees of force or vivacity” (EHU 2:3).[1] The less vivid or dull perceptions, he calls thoughts or ideas. Hume then decides to use the word impressions, to refer to the more lively sensations of the mind. He cautions that this use of the word impressions differs from its usual use. Thus, by impressions, Hume (EHU 2:3) writes “I mean all the more lively perceptions, when we hear, or see, or feel, or love or hate or desire or will.” So how are ideas or thoughts and impressions related? Ideas, according to Hume, are conscious recollections of

---

[1] EHU is *An Enquiry Concerning Human Understanding*. The first number is the chapter and the second is the part. If a third number is added, it will be the paragraph. Where there is only a single number, it will be the page number.

the more lively perceptions. Thus, according to Hume, they are not as vivid as the first-hand perceptions. Hobbes' (1651/2012, 2) equivalent for the less vivid or dull perceptions is "imagination," which he describes as decayed or weakened sense. Thus, Hume writes (EHU2:5) "all the materials of thinking are derived either from our outward or inward sentiment." In this case, according to Hume, all our ideas, which he calls "feeble perceptions, are copies of our impressions or more lively perceptions" (EHU 2:5).

Hume divides impressions into *original impressions*, or *impressions of sensation,* and *secondary* or *reflective impressions*. The former comprise "all the impressions of the senses and all bodily pains and pleasures." The latter "are the passions and other emotions resembling them" (*Treatise*, 2.1.1.1)[1]. They are called impressions of reflection because they are derived from ideas, which are copies of the original impressions (*Treatise* 1.1.2.2) From this quotation we can see that Hume uses 'passion' and 'emotion' synonymously, and there is further evidence of this on p. 276, where he uses the phrase: "those emotions or passions, their nature, origin, causes and effects." Cohon (2010) reads Hume similarly in his entry on Hume in the *Stanford Encyclopedia of Philosophy*. Hume explains further that the secondary or reflective impressions are derived from the original impressions either directly or by the interposition of ideas of them. In

---

[1] In this citation system, *Treatise* refers to Hume's *Treatise on Human Nature*. The first numeral is the book number within the *Treatise*, followed by the part, section and paragraph numbers respectively.

other words, not only do impressions such as pain, pleasure or smell elicit some passions, thoughts of them can also elicit those passions or emotions.

Hume splits the emotions or passions further into direct ones, which arise immediately from pain or pleasure, and indirect passions, which involve additional qualities beyond those that generate the direct passions. Hume leaves the basis for this distinction rather scanty, but goes on to enumerate the specific passions (*Treatise*, 2.1.1.4). The direct passions are “desire, aversion, grief, joy, hope, fear, despair and security.” Included in the indirect passions are “pride, ambition, vanity, love, hatred, malice, pity, generosity, with their dependants.”

Hume goes on to explain the causes of the passions. As we have seen here, the passions are caused or excited by the primary impressions as well as ideas, which are copies of the primary impressions and the passions (secondary impressions). Consequently, there are two sources exciting the passions. The first is directly by the “idea” that is presented to the mind. For example, the news that you have been offered a job for which you interviewed will elicit joy; suffering a bereavement or having your house destroyed by a natural disaster will elicit sadness; if your house is destroyed by a fire set by an arsonist, anger might be another elicited passion.

The second source of excitement for the passions is some quality in objects or the motives of actions. This brings us to another important term in Hume’s moral theory – *virtue*. There are certain qualities in objects and motives of actions that excite the passions, and it is those that Hume uses the word *virtue* to refer to. In Hume's example, if I am proud of my horse because it is beautiful, it is the quality *beauty* that excites the passion of pride in me. Thus, in the case of the arsonist who burnt down my house, if I discover that it was malicious rather than accidental, it is the quality of *malice* that excites my passion of anger. Hume (*Treatise*, 2.1.5.2) states, therefore, that it is "the good and bad qualities of our actions and manners that constitute virtue and vice."

Hume (*Treatise*, 3.1.2.6) suggests that there are so many moral norms and duties that it is unlikely, or even "impossible that our original instincts should attend to each of them, and from our very first infancy impress on the human mind all the multitude of precepts, which are contained in the completest system of ethics." For this reason, he sought to identify "some more general principles upon which all our morals are founded."

Hume realizes that these fundamental principles of our moral impulses cannot be based on what is natural and what is not. He (*Treatise*, 3.1.2.10) considers the various senses of *natural*. "In the first sense of the word *nature*, as opposed to *miracles*, both vice and virtue are equally natural." In another sense, we can say that since all our actions are performed with "a certain design and intention," and since such actions can either be virtuous or vicious, "virtue and vice are equally artificial and out of nature." Hume (*Treatise*, 3.1.2.10) concludes then "it is impossible therefore, that the character of natural and unnatural can ever, in any sense, mark the boundaries of vice and virtue." This resolves Rousseau's disagreement with Hobbes, to which Pinker (2006, 4) has drawn attention, and which concerned the question whether man is by nature a “noble savage” who is corrupted by civilization, as Rousseau thinks, or a brutish savage tamed by civilization, according to Hobbes.

This is one of the occasions, as far as I can tell, when Hume distinguishes himself as an original and independent thinker. For like Rousseau and Hobbes above, nearly every moral philosopher prior to Hume and even beyond, in fact, including those who are said to be his so-called influences, has lined up on one side or the other on the long standing debate as to whether humans are by nature selfish or benevolent, or, correspondingly, whether they are vicious or virtuous.

Grotius (2013, 796-800) for example, writing in the sixteenth century, in toeing the “natural virtue” line, traces a lineage of philosophers all the way back to Aristotle who, held man to be “an animal of a gentle nature.” In between were philosophers such as Cicero, Lactantius, and Plutarch who asserted that “man neither was, nor is, by nature, a wild unsociable creature. But it is the corruption of his nature which makes him so.” Plutarch argued that taming man through habit and culture only reclaims him to “his original gentleness.” Another form in which the same debate is cast is whether humans are inherently benevolent, as a result of which they are able to live in societies, or whether they form societies for their own inherently selfish motives Gill (2000). I will revisit this issue a little later.

For Hume (*Treatise*, 3.1.2.11), “virtue is distinguished by the pleasure, and vice by the pain, that any action, sentiment or character gives us by the mere view and contemplation [of it].” “It is from the prospect of pain and pleasure that the aversion and propensity arises towards any object” (*Treatise*, 2.3.3.3). This is one of the conclusions Hume reaches, which has abundance of precedence and may have originated with Hobbes (1651). It is interesting however, to notice how Hume arrives at such conclusions and where he goes with them. Spinoza (*Ethics IVp8*), for example, similarly held that things we call good or evil correspond to whether they affect us with pleasure or pain. He also concludes that “the knowledge of good and evil is nothing else but the idea of the pleasure or pain, which necessarily follows from that pleasurable or painful emotion.” It is important to note that Hume does not assert views such as the one in the above quotation as ends in themselves. He asserts them as building blocks for his comprehensive theory on human morality or sociality. My interest, in the same vein, is not in Hume’s insight into human social morality as an end in itself, but how it might be a piece in the bigger puzzle of the evolution of human morality and sociality. Thus, in the grand scheme of things, assertions such as those relating pleasure to virtue are tangential to my analysis, and as such, this discussion should not get mired in their histories.

For Hume, considered simply, desire arises from good, and aversion from evil, and all these arise ultimately from pleasure and pain. A will is formed upon the discovery of any cause of action by which the mind or body may attain pleasure or avert pain (*Treatise*, 2.3.9.7). Thus, Hume (*Treatise*, 2.3.1.2) describes the will as “the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind." The will itself, then, is not a cause of action, other than the instincts and the direct passions, as Cohon (2010) explained. It is just another impression we feel as the mind or body reacts to the excitement of a passion. This is what I will call "total naturalism." It plants man completely in nature, and makes his actions completely determined and explained by natural processes. This, as opposed to philosophers who, as Spinoza observes, tend to view man as "a kingdom within a kingdom: for they believe that he disturbs rather than follows nature's order, that he has absolute control over his actions, and that he is determined solely by himself" (*Ethics III).*

III. Hume's Moral Thesis

The goal of Hume's moral analysis, as he states it, is to correct the predominant view "upon which the greatest part of moral philosophy, ancient and modern, seems to be founded." Thus, he opens his treatise on morality saying: "Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason, to give the preference to reason, and to assert that men are so far virtuous as they conform themselves to its dictates" (*Treatise*, 2.3.3.1). This supposed pre-eminence of reason over passion is still a popular sentiment even today. However, Hume's "affective primacy" view, as it is sometimes called, is only now being rediscovered by current empirical psychology, even though Hume argued it nearly three hundred years ago. In that iconoclastic view, Hume (*Treatise*, 2.3.3.1) argues that "reason alone can never be a motive to any action of the will, and secondly, that it can never oppose passion *in the direction of the will*."[1]

By taking this bold stand, Hume is going against the current of centuries of virtual unanimity (a probable exception being Spinoza) in a philosophical view that roots human morality in rationality. Emotion and self-interest are often seen as elements of our more primitive nature that rationality or reason, which is unique and innate to humans, has to subdue in order for morality to hold. Thus, Grotius (1625/2013, 770) criticises Roman law that assigns "one unchangeable right to brutes in common with man, which in a more limited sense they call the law of nature, and appropriating another to men, which they frequently call the Law of Nations." He argues instead that "no beings, except those that can form general maxims, are capable of possessing a right," as might arise from a law (of nature). This will later feature quite prominently in Kant's deontological ethics, in which only rational beings can be moral patients, i.e. beings to whom we owe moral duties. Grotius (1625/2013, 770) goes on to cite the ancient Greek poet Hesiod as saying that "the supreme Being has appointed laws for men; but permitted wild beasts, fishes, and birds to devour each other for food." For Grotius (1625/2013, 730), like the Kantian, "Natural right is the dictate of right reason, [out of which arises] the moral turpitude, or moral necessity."

---

[1] *In the direction the will* or *in directing the will*, properly put.

In agreement with Hume, on the other hand, Rousseau, in his *Discourse on the Origin of Inequality* (henceforth, DOI) (p. 16) declares that "the human understanding is greatly indebted to the passions," in the sense that "it is by the activity of our passions, that our reason improves: we covet knowledge merely because we covet enjoyment, and it is impossible to conceive why a man exempt from fears and desires should take the trouble to reason." In other words, reason helps us satisfy our desires and combat our aversions. Reason, however, does not determine or dictate what these desires and aversions are. They stem from our inherent nature. Thus, Rousseau points out (DOI, 16), "the progress of the mind has everywhere kept pace exactly with the wants, to which nature had left the inhabitants exposed, or to which circumstances had subjected them, and consequently to the passions, which inclined them to provide for these wants." Thus, we use reason to discover more efficient ways to satisfy wants, which are determined by the passions solely. As Prinz (2012, 302) remarks, "reason can only tell us how to act if we already know what we want to achieve." We may conclude, then, that greater knowledge or higher reason and understanding only makes us more efficient at satisfying these passions, but do not alter them. This helps clarify the role of reason in moral judgement, which Hume tries here to explain. It also serves to distinguish Hume from most philosophers prior and subsequent, who see the role of reason in moral judgment as subduing the passions.

Diametrically opposed to this view of the relationship between reason and our natural inclinations is Immanuel Kant, for example, who argues in his *Metaphysics of Morals*, that if the preservation, welfare or general happiness of a being were the purpose for its existence, nature would have made a poor choice in ‘reason’ for attaining such a goal. He suggests that nature would instead have trusted that function to instinct, which would have been more efficient. Kant is of the view that reason is intended for a much nobler end to which our natural inclinations must defer. The true function of reason, according to Kant, is to produce a will which is good in itself, and not a means to anything else.

It is outside the scope of this work to discuss Kant’s normative ethical principles. An important general observation I will make is that the discordance we are witnessing between Hume and Kant is akin to what we saw between Kuhn and Popper in chapter one. It is that normative theories such as, “this is what science ought to be”, or “this is what morality ought to be” etc. tend to prescribe standards that are ideal and often unreflective of reality or actual practice. Thus, according to Kant, an action has the highest moral worth if it is compelled by duty in opposition to our natural inclinations or desires (*Metaphysics of Morals*, First Section). Further to this notion of morality as rationality overriding passion, Grotius (1625/2013, 769) asserts that man, “possessing the knowledge of good and evil, refrains, even with inconvenience to himself, from doing hurt.” However, as we shall see, Hume observes, contrary to Kant’s “duty theory,” that what we praise and blame, or perceive as virtue and vice, tend to conform to our natural passions. As we shall also see, Hume’s descriptive account has been abundantly corroborated by current research in neuro-psychology.

In order to help us understand the nature and true source of virtue and morality, Hume sets up an analogy between moral and cognitive judgment. He explains why reason is not the source of moral judgment, even though it is often mistaken to be. He writes: “Reason is the discovery of truth or falsehood. Truth or falsehood consists in an agreement or disagreement either to the real relations of ideas, or to real existence and matter of fact” (*Treatise*, 3.1.1.9). However, our passions are not the kind of thing that agree or disagree with reality. Consequently, they are never judged in terms of truth or falsehood and are therefore not conformable to reason. “Abstract or demonstrative reasoning never influences any of our actions, but only as it directs our judgment concerning causes and effects” (*Treatise,* 2.3.3.2). For Hume, therefore, reasoning is “incapable of preventing volition or of disputing the preference with any passion or emotion” (*Treatise,* 2.3.3.4).

For those who would dispute this point, recall that Hume stated right at the beginning of his moral thesis, that he wished to dispel the myth of the opposition between reason and emotion in moral judgements, in which we give preference to moral judgments based on reason rather than emotion. He explains in detail how reason and emotion operate in completely distinct domains and are not alternatives or competing processes for making the same judgments. That is why, according to Hume, reason does not "oppose or retard the impulse of passion." Only a contrary passion may do so (*Treatise*, 2.3.3.4). Kant will later become the emblem for the traditional view that Hume emphatically opposes. In the Kantian system, our natural passions frequently oppose our duties as determined by reason, and as Ellington writes in his commentary on Kant's (1785) *Grounding for the Metaphysics of Morals*, "when there is a conflict between duty and inclination, duty should always be followed." For Hume, "our sense of duty always follows the common and natural course of our passions" (*Treatise*, 3.2.1.18). Thus, contrary to what Kant believes, duty arises from our very inclinations and the two are not in frequent opposition. Kant comes to the mistaken belief because he attributes a different source to duty, namely, reason.

Spinoza (*Ethics*, IVp7), in concordance with Hume, has also noted that "an emotion can only be controlled or destroyed by another emotion contrary thereto, and with more power for controlling emotion." Our moral choices therefore consist of choosing between competing emotions. Reason supplies the "cause and effect analysis," according to Hume, that helps in deciding which emotion to act on and which to leave unsatisfied. "Morality therefore is more properly felt than judged," according to Hume (*Treatise*, 3.1.2.1). As Deleuze (1991, 34), explains in his expositional work on Hume, "causal relations are grasped by inference," whereas "there is no inference to be drawn" in the case of ethics.

Hume notes, therefore, that moral “actions may be laudable or blamable; but they can never be reasonable or unreasonable” (*Treatise*, 3.1.1.10). Hume explains that reason may inform us of certain relationships, which can turn out to be erroneous. However, actions based on such false judgments of reason are never condemned as immoral. “They extend not beyond the mistake of fact, which moralists have not generally supposed criminal” (*Treatise*, 3.1.1.12). If through the fallibility of our reasoning we suppose mistakenly that some means could achieve a certain end, and we consequently fail in that endeavour, such an error is usually not seen as a defect in our moral character. Ineptitude is a far cry from evil; they say.

So if a mistake of fact is thus not morally blamable, what about a mistake of right? Hume suggests this could not possibly be the original source of judgment of immorality because such a mistake presupposes some real right and wrong. When we judge an action such as theft for example, to be wrong, we do so based on some antecedent morality. When we say that someone has a duty to be grateful for some kindness extended to him, we are, according to Hume, supposing that to be commanded by some antecedent rule of duty or moral, rather than simply duty as an end in itself, as Kant held. Until we find the ultimate basis of these moral judgments, we reason in circles. In other words, there is some ‘moral reality’ or object, the conformity or non-conformity to which our moral actions are judged to be right or wrong respectively. What is that “moral object?”

IV. From Instinct to Morality

Hume sets his moral theory at the deepest foundation possible – the biological being and his instincts (or what he calls the "natural passions") as set by nature, and out of which every virtue emanates. So Hume takes the passions as given or implanted by nature, and then goes on to explain how they generate the moral sense. As we saw in the previous section, however, Aristotle, Cicero, Plutarch, Hesiod, Grotius, and Kant, do not think morality arises from our passions, which they point out we share with brutes. Instead, they point to that aspect of human nature that elevates humans above the brutes, i.e., rationality, which they consider to be the source of morality. Grotius (1625/2013, 769), for example, citing Lactantius (one of the early Christian fathers), asserts that "in all animals devoid of reason we see a natural bias of self-love. For, they hurt others to benefit themselves; because they do not know the evil of doing wilful hurt." Notice that the difference between Hume and the opposing camp is not whether morality arises from some innate quality of humans or not, for both camps proclaim that. As Grotius asserts, the law of nature "proceeds from the essential traits implanted in man" (Miller 2014, 13). Hume would not disagree. Hume (*Treatise*, 3.2.2.14) acknowledges human nature to be "composed of two principal parts, which are requisite in all its actions, the affections and understanding." The difference though, is whether moral judgements stem from our rational nature (the understanding) or from our base emotions (the affects). For Hume (*Treatise*, 3.2.2.14), they are both involved, even though he gives primacy to the affects or passions, whereas the opposing camp, exemplified by Grotius (1625, 2013, 730), asserts that “the moral turpitude or moral necessity of any act [stems] from its agreement or disagreement with a rational nature.”

The relevant natural principles that govern the passions are “egoism” and “sympathy,”[1] the same duo Hume alternatively calls “self-interest” and “limited generosity.” Hume in his moral theory endeavours to show how our moral judgments are undergirded by those principles. To this effect, Hume begins by going over what is involved in moral attitudes or moral judgments. Gilles Deleuze notes in his account of Hume’s moral theory that the essence of moral conscience is the approval and disapproval of the actions of people, and that what we approve and disapprove is character in general rather than individual interest (Deleuze 1991, 37). Hume’s analysis does have some common elements with Aristotle’s virtue ethics, as for example when Hume holds that “when we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles in the mind and temper” (*Treatise*, 3.2.1.1). However, for Hume, unlike Aristotle, the virtue that impels a person’s action is not itself the goal of the action, in other words, a virtuous act cannot be said to be performed for the sake of virtue. That is the circular reasoning Hume tries to correct. The motive for the virtuous act should be some other principle (*Treatise*, 3.2.1.7).

---

[1] What Hume refers to as ‘sympathy’ is what psychologists today refer to as ‘empathy.’ So in this discussion I will mostly use the word empathy, except where I am directly quoting Hume.

For Kant, for example, the motive of virtuous acts is *duty* proximally, which is determined by reason ultimately. As he writes (1785/1993, 9), "reason recognizes as its practical function, the establishment of a good will." In doing so, he adds, reason "fulfils a purpose which in turn is determined only by reason." The will established by reason is good in itself, and "always holds first place in estimating the total worth of our actions" (9). It is this *will* that duty requires as to follow, even when it is (and it often is) contrary to our inclinations (i.e. our natural passions). Grotius (1625, 2013, 730) similarly declares that "natural right is the dictate of right reason, shewing the moral turpitude or moral necessity of any act from its agreement or disagreement with a rational nature." These examples describe the popular sentiment regarding rationality and morality that Hume sets out to dispel.

Characteristically, Hume illustrates his viewpoint by arguing from uncontroversial common experiences. On this issue, Hume points out that when we blame a father who neglects his child, it is because we expect natural affection from a father for a child. Thus, the motive for that action is distinct from a sense of duty, which itself arises from the natural affection we expect the father to have for his child. In Hume's second example, if we imagine a benevolent person who spreads his bounties even to include strangers, we regard this deed as a proof of his humanity. So the merit we see in his actions is only secondary to the antecedent principle of humanity to which we think his action is consistent and, which the source of the merit of the act is. Consequently, Hume declares the principle: “that no action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality” (*Treatise*, 3.2.1.7). Further on (*Treatise*, 3.2.1.18) Hume narrows down this motive further. He describes those motives as “impelling passions” that have a great influence on our sense of morals, and adds that “it is according to their general force on human nature that we praise or blame.” Thus, according to Hume (*Treatise*, 3.2.1.18), we always consider “the natural and usual force of the passions, when we determine concerning virtue and vice.” It is according to conformity with the general force of the passions or lack thereof that we praise or blame.

Then, reminiscent of Aristotelian virtue ethics, Hume (*Treatise*, 3.2.1.18) adds that “if the passions depart very much from the common measures on either side, they are always disapproved as vicious.” Hume is talking about individual character now, saying that human nature gives each of us a sense of its (human nature) mean for each of the passions. The closer a behaviour or character is to the mean (or the typical) of human nature, the more virtuous it is, and the further a behaviour deviates from the mean on either side, the more vicious it is (*Treatise*. 3.2.1.18). Aristotle differs from Hume in seeing this “mean” as the action that is optimal in achieving the desired good or end, whereas Hume considers this mean to be the modal human nature or the point where the majority of natures of individual humans lie. In other words, “the mean,” for Hume, is the typical rather than the ideal.

The basis of Hume’s moral theory, therefore, is simply that moral actions are those that are laudable or blamable on the basis of motives that are impelled by the passions of human nature. Spinoza (*Ethics*, IIIp29) similarly writes: “We shall also endeavour to do whatsoever we conceive men to regard with pleasure, and contrariwise we shall shrink from doing that which we conceive men to shrink from.” Spinoza explains that from the mere fact that we tend to love or hate a thing that men generally love or hate, “we shall feel pleasure or pain at the thing's presence.” Thus, Hume has established from common knowledge or uncontroversial notions of human behaviour the basis of our approbation and disapprobation. As we see in this particular case, Spinoza had reached the same conclusion arguing from a very different angle.

Again, Hume’s criterion for making these moral judgments is quite consistent with Aristotle’s view of virtue as the mean between the two extremes of excess and deficiency. Similarly, Hume asserts that “we always consider the natural and usual force of the passions, when we determine concerning vice and virtue; and if the passions depart very much from the common measures on either side, they are always disapproved as vicious” (*Treatise*, 3.2.1.18). In Cohon’s (2011, 24) exposition, “We reach a moral judgment by feeling approval or disapproval upon contemplating someone's trait in a disinterested way from the common point of view.” Thus it is from the need to act within the mean range of these passions that our sense of duty is derived. For example, the proper order of the passions is that a man loves his children more than his nephews, then his cousins and then strangers. As he reiterates it elsewhere (*Treatise*, 488), "in the original frame of our mind, our strongest attention is confined to ourselves; our next is extended to our relations and acquaintance; and it is only the weakest which reaches to strangers and indifferent persons." Thus, this partiality of empathy, or "unequal affection," Hume (*Treatise*, 3.2.2.8) writes, "must not only have an influence on our behaviour and conduct in society, but even on our ideas of vice and virtue; so as to make us regard any remarkable transgression of such a degree of partiality, either by too great an enlargement, or contraction of the affections, as vicious and immoral." Consequently, it does seem to us untoward that a man would favour a stranger over his own child. In line with this reasoning, the Stoic philosopher, Seneca suggests a sliding scale of moral responsibility. In *On the Private Life* (3.5), Seneca says: "What is required, you see, of any man is that he should be of use to other men—if possible, to many; failing that, to a few; failing that, to those nearest him; failing that, to himself" (Vogt 2015, 10-11). This lends support to Hume's position.

V. The Origin and Nature of Justice

When we borrow something from someone, we feel we need to restore it for reasons of honesty and justice. Where is this need for honesty and justice from? For Hume, a virtue, such as honesty or justice in this case, is never an end in itself. We do not perform a just or honest act simply to be just or honest. This is quite distinct from the view of deontological ethicists such as Kant and W. D. Ross, who see such moral duties as *a priori* and thus not in need of any further justification; and such a view, Hume has shown to entail circularity. For Hume there is an underlying motive or principle that drives the desire to be just or honest or any other virtue. These are not ends in themselves.

But what is this fundamental moral motive? This is one of the many occasions on which Hume presents an analysis that tears down a false dichotomy set by his predecessors. We cannot simply say that we act justly for the sake of our own private interest. For unrestrained "self-love," as Hume observes, is the root of all injustice and violence. Just actions also could not be for public interest, which Hume thinks "is too remote and too sublime to affect the generality of mankind" (*Treatise*, 3.2.1.11). In fact, Hume does not think there exists any such passion as "the love of mankind" independent of merit and circumstances (*Treatise*, 3.2.1.12). He suggests that what appears to be a general love for fellow humans and even other sentient beings "proceeds only from the relation to ourselves; which in these cases gathers force by being confined to a few persons." It is important that we pay attention to what Hume is saying. For that feature of empathy will be crucial to explaining social behaviour in the next chapter.

Hume puts nuance into what has sloppily been perceived by many as "empathy for mankind and even all sentient beings." As Deleuze (1991, 37-38) elucidates, "it is impossible for sympathy to extend 'without being aided by some circumstance in the present, which strikes upon us in a lively manner,' excluding thereby, cases which do not present these circumstances." There is no blanket, uniform and unconditional sympathy. Hume illustrates this, saying that (*Treatise*, 3.2.1.12) “an Englishman in Italy is a friend: A European in China; and perhaps a man would be beloved as such, were we to meet him in the moon.” However, we would have no such special affection for the same stranger if we met him in the street in our hometown. Thus, the sympathy instinct is naturally directed towards family members and those with whom we have some kind of association, on the basis, as Deleuze (1991, 38) thinks, of contiguity, resemblance and causality. In this way, it excludes strangers. In line with this, Spinoza (Ethics, IIIp27) has also suggested that “by the very fact that we conceive a thing, which is like ourselves, and which we have [prior] not regarded with any emotion, [were such a thing] to be affected with any emotion, we are ourselves affected with a like emotion.” In other words, based on perceived resemblance or shared identity, one individual feels the emotion of another.

The original motive for morality therefore, is not in the regard for the interest of mankind (public benevolence) or in regard for the interest of the particular individual we are interacting with, what Hume refers to as the “party concerned” or “private benevolence.” Hume’s position is that “the sense of justice is not derived from nature, but arises artificially, though necessarily from education and human conventions” (*Treatise*, 3.2.1.17). We again see a similar sentiment from Spinoza (Ethics IVp37n2), who asserts that “justice and injustice, sin and merit, are extrinsic ideas, and not attributes which display the nature of the mind.” In other words, laws have to be invented in order to generate rights or justice, and where such rights do not exist, organisms have licence with each other, a view also expressed by Hobbes and Rousseau.

In the contrary camp is Grotius (1625/2013, 770) for example, who declares that “no beings, except those that can form general maxims, are capable of possessing a right.” For Grotius, the capacity to form these general maxims stems from our rational nature, which is implanted in as by God. This puts him in the camp of those who take justice to be natural rather than artificial. Grotius (2013, 770) cites Hesiod as saying that “the supreme Being has appointed laws for men; but permitted wild beasts, fishes, and birds to devour each other for food.” Grotius (2013, 730) explains that property, for example, was initially created by the human will, “but, after it was established, one man was prohibited by the law of nature from seizing the property of another against his will.” In other words, justice, which prohibits one man “from seizing the property of another against his will,” is a law of nature, rather than an artifice of humans as Hume asserts.

Also to be found in this opposing camp is Leibniz, who, according to Arthur (2014, 170), “objects … to the Hobbesian position that there is no justice in the state of nature.” Instead, Leibniz sides with Grotius in the view that natural obligations such as justice “proceeds from the essential traits planted in man” (Arthur 2014, 171). As I explained earlier, this is also the position of Darwinians, except that they believe that innate sociality replaces, by natural selection, our original innate selfishness. The problem for the Darwinian, as I pointed out earlier, is that such a transformation from innate selfishness to innate sociality has proven to be theoretically insolvent. Thus, the Humean position that the social virtues are artifices of humans rather than implanted (by whatever means) in human nature potentially provides the key to theoretical solvency for the Darwinian.

According to Arthur (2014, 171), “Leibniz did not dispute that we act out of self-interest any more than Hobbes, but he did not regard this as incompatible with sociability or morality.” This, as Arthur (172) explains, quoting Leibniz, is because “we love a thing whose happiness causes our own pleasure.” This alludes to empathy. What Hume observes, however, is that empathy is the very thing that makes the formation of society nearly impossible because it is by nature almost entirely confined to the family and close associates and virtually non-existent for the stranger. Thus, even though very few scholars would fail to see the huge advantage that society means to man in the pursuit of his self-interest, that very self-interest stands in opposition to the formation of society. That, in fact, is the source of the theoretical difficulty for the Darwinian and also the basis of the paradox of the prisoner’s dilemma. Hume gives an excellent psychological account of how this opposition is generated, as I will briefly discuss later. What Hume observed is only now being realized in empirical psychology.

VI. The Origin of Society

“Of all the animals with which the globe is peopled, there is none towards whom nature seems, at first sight, to have exercised more cruelty than towards man, in the numberless wants and necessities with which she has loaded him, and in the slender means, which she affords to the relieving these necessities" (*Treatise*, 3.2.2.2). This is the insight on which Hume's theory of sociality or social morality is based. It is a very compelling insight, which will later become the most fundamental concept of the subject of economics – *the economic problem* – as it has come to be referred to in modern economics. Hume may have been influenced here by Stoic thought. Seneca, for example, writes in *On Benefits*, that "God has granted two things that make this vulnerable creature the strongest of all: reason and fellowship. […] Fellowship has given him power over all animals [: …] Remove fellowship and you will destroy the unity of mankind on which our life depends" (Vogt 2015, 11). This points to the indispensability of society to man.

Elucidating on this, Hume observes for all other creatures, their advantages hold in proportion to their wants. The sheep is adequately equipped for life as a herbivore, and the lion is equally equipped to meet all the necessities for life as a carnivore. It is only in man that we find this imbalance of infirmity and necessity. It is by society, according to Hume, that man's infirmities are compensated for. Through society, man is afforded the opportunity for a division of labour, which greatly augments his abilities. This is another legacy of Hume to modern economics, primarily via Adam Smith (1776). Through the coordination of efforts, man's output of is multiplied several fold so that he can accomplish tasks that required forces greater than he could ever muster as a solitary individual. He is also afforded security through mutual support in society. In these three provisions lies the advantage of society to man; and society is formed and maintained when man appreciates these advantages of society.

I would like to point out one mischaracterization in what Hume sets as the key circumstance that necessitated human sociality, i.e. man's physical ill-adaptation for his environment. It is inconsistent with modern (Darwinian) evolutionary explanation to suppose, as Hume does, that an organism could come into being with a relative fitness disadvantage and survive until a compensatory adaptation is evolved. In other words, an organism does not exist without an answer to a survival challenge before or until it finds one. That is inconsistent with Darwinian evolutionary logic. For an organism to survive or even evolve it must have all the adaptations essential for its survival. An essential new adaptation does not fill an adaptive vacuum. It replaces an inferior one which preserved the organism to that point.

Thus, it is not the case that man sought society because he was not physically equipped to survive in nature. Rather, man shed many of the features with which nature had equipped him for survival because they had become irrelevant after he had formed social groups and had also, more importantly, acquired the ability to use tools. Hume's faulty premise, however, does not derail his moral theory, since a dependency does indeed exist now, and society is inextricably part of what it is to be a modern human being. The essential premise on Hume's analysis is that society is an essential human adaptation. Few will dispute the premise that society has an enormous net benefit to humans. At this superficial level, Leibniz is right in his assertion that self-interest is not incompatible with sociability (Arthur 2014, 171), even though, as we saw earlier, in coming to this view, Leibniz overlooked the partiality of empathy. However, Hume, on a deeper analysis, thinks that unrestrained self-interest undermines sociability, with which the modern Darwinian concurs. Hume's theory of social evolution, which I render below, consists in explaining how this opposition comes about and how it is overcome.

Hume notes that the scarcity of the resources necessary for man's survival, together with the ease with which such goods can be expropriated, present an impediment to the formation of society; and there is no remedy for this inconvenience in "uncultivated nature," according to Hume (*Treatise,* 3.2.2.8). As we shall see further on, our partial affections also exacerbate this difficulty, and we should not, according to Hume (*Treatise,* 3.2.2.8), "hope for any inartificial principle of the human mind" to overcome this difficulty. In other words, there is nothing inherent (or evolved) in human nature that can resolve the difficulty other than artificial contrivances.

This is a watershed point, which is pivotal for my effort to resolve the modern Darwinian problem of social evolution. For in holding that the social virtues are inherent in human nature (i.e. evolved), the modern Darwinian is united with Grotius, Leibniz and Kant (GLK) against Hume, Hobbes and Rousseau (HHR). The project of the modern Darwinian has been to provide a mechanism for some evolutionary change in the natural constitution of humans that would have made them moral or social. Thus, while they both hold the social virtues to inhere in the nature of humans, the difference between Darwinian and the GLK positions is that whereas GLK hold the social virtues to be implanted in humans (by God?), the Darwinian holds that they evolved subsequently in human nature by supplanting selfishness through natural selection.

In arguing the artificiality of the social virtues (justice in particular), Hume considers the attraction between the sexes (which results in the family) to be “the first and original principle of human society” (*Treatise*, 3.2.2.4). This is echoed by Rousseau (1762/2010, 1), who declares that “the most ancient of all societies, and the only one that is natural, is the family.” It is important to note however that in Rousseau’s view the bond between parent and offspring terminates as soon as the offspring can survive independently. As he writes (1762, 1), “If they remain united, they continue so no longer naturally, but voluntarily; and the family itself is then maintained only by convention.” Hume does not go this far. Hume believes in a natural unselfish bond (sympathy) that exists between family members and close associates.

This is a critical point to note for the application of Hume’s insight to the biological problem of social evolution. Notice that Hume locates “the first and original principle of human society” at the attraction between the sexes, not altruistic or cooperative behaviour. Actual social behaviour, of which current Darwinian explanation picks out altruism as problematic, follows after the emergence of empathy. Current evolutionary biology perceives altruism to be inherent in our natural constitution, and seeks to explain how a gene that causes that behaviour might have evolved. As Hume's account shows, however, the idea of "regard for the welfare of others" is much more nuanced than unmitigated altruism.

Similarly, Hume's predecessors in moral and political philosophy jumbled up many issues and concepts, which Hume teases out and systematizes in his moral social analysis. Discussions of human sociality or morality, both before and after Hume, often find scholars arguing from opposing sides of a simplistic dichotomy. I have already discussed the example from Pinker (2011) as to whether humans are inherently virtuous or vicious, which Hume's analysis has shown to be a faulty dichotomy. But there are more such simplistic dichotomies in human social behaviour that Hume's analysis dispels.

In "Hume's Progressive View of Human Nature," Gill (2000) provides another example of Hume's analysis undermining the two contrasting moral views of his predecessors. Gill explains that Hume's moral theory has many elements in common with those of some of his English predecessors, in particular, Shaftesbury, Mandeville, and Hutcheson on the origin of human sociability. All three philosophers recognized the propensity of humans to form societies. However, they disagreed as to whether human sociability originates in natural benevolence, as Shaftesbury and Hutcheson maintained, or whether it originates in self-interest as Mandeville held.

Hume’s analysis is consistent with Mandeville’s position that self-interest is the motive of justice and society. However, Hume once again establishes a more nuanced position by stopping short of endorsing what he sees as an effort by some writers on morals “to extirpate all sense of virtue from among mankind” (*Treatise,* 3.2.2.24). Hume explains that even though “self-interest is the original motive for the establishment of justice, sympathy with public interest is [indeed] the moral approbation which attends that virtue” (*Treatise* 3.2.2.23). However, this particular moral sense is not natural, but is instead cultivated through experience, education and exhortations.

In constructing his social theory, Hume leads us through a systematic series of steps to the fundamental principle that serves as the motive for virtuous actions. The first principle, as we have just said, is the affection between the sexes, which no one will doubt is a primary instinct of human nature. In Hume’s reasoning, the attraction between the sexes results in the formation of unions between individuals, which are often further strengthened by the generation of offspring. The natural affection of the parents for the offspring represents another principle of union. Thus, there arises “sympathy between parents and sympathy of parents for their offspring” (Deleuze 1991, 39). Such family units provide the conditions for parenting, through which “custom and habit, operating on the tender minds of the children, makes them sensible of the advantages, which they may reap from society” (*Treatise*, 3.2.2.4). Thus, forming a society, which evolutionary biologists reduce to behaving altruistically, is not inherent in humans in the same way that ducklings swim. It is learned and contingent. Note that society in the Humean sense means "political society," which implies large and often cosmopolitan groups. So we see here how the primary passion of sympathy, which unites the nuclear family, also provides the conditions for cooperation beyond the family, and is ultimately one of two primal instincts or "natural passions" out of which virtue and morality emanate. The family, thus, is the germ of society. As we shall see below however, that which makes the family possible is opposed to the formation of larger societies. Again, to avoid confusion and seeming contradiction, let us remember that for Hume, such small, primordial social groups within which sympathy operates beyond the family are not yet societies. Hume refers to them as "rude and uncultivated nature."

Things do not go quite smoothly from family to society, unfortunately, for there are countervailing natural passions working against the apparent development of social cohesion from the attraction between the sexes and the affection between parents and offspring. In particular the partiality of sympathy. Hume notes that the effect of the selfish passion has generally been exaggerated. In fact, contrary to what most philosophers and particularly Darwinians think, the biggest obstacle to the formation of society is our natural empathy, rather than our natural selfishness. Hume asserts that "though it be rare to meet with one, who loves any single person better than himself; yet, it is rare to meet with one, in whom all the kind affections, taken together, do not over-balance all the selfishness" (*Treatise*, 3.2.2.5). Hume (*Treatise,* 3.2.2.5) points out that we frequently invest more in our wives and children than ourselves, which lends further support to his claim that empathy does have the capacity to prevail over egoism.

Rousseau (DOI, 30) also recognizes empathy (*pitié*) as a natural passion[1] that moderates self-love, and describes it as man's "innate abhorrence to see beings suffer that resemble him." Rousseau was arguing against Hobbes, and may have intended "beings that resemble him" to mean humans, and therefore a preference for humans as against non-humans. His example (DOI, 30) that "one animal never passes unmoved by the dead carcass of another animal of the same species" indicates the scope of empathy he has in mind.

However, empathy is much more restricted in scope in the human species as a group. In Hume's analysis, the resemblance, which Rousseau (and Spinoza) suggest to be the driver of empathy, represents a distinction, within humans, of relatives and associates from strangers. As we shall see, the confinement of sympathy is a stumbling block to the formation of society beyond the family.

What Hume acknowledges to pose the bigger problem for the formation of larger societies is that this empathy-driven generosity and affection between family members, which is often much stronger than self-love, "instead of fitting men for large societies, is almost as contrary to them as the most narrow

---

[1] "Natural passion' (as used by Hume) means passion which is inherent in our nature or an evolved trait.

selfishness" (*Treatise,* 3.2.2.6). It thus tends to exclude individuals outside the family, as Hume (*Treatise,* 3.2.2.6) explains:

> For while each person loves himself better than any other single person, and in his love to others bears the greatest affection to his relations and acquaintance, this must necessarily produce an opposition of passions and a consequent opposition of actions; which cannot but be dangerous to the new established union.

Hume (*Treatise*, 3.2.2.8) refers to this love of family to the exclusion of others as "partiality" and "unequal affection." In Deleuze's exposition, Hume asserts that "human beings are more partial than they are egoistic" (Deleuze 1991, 38), and it is this partiality that generates what Deleuze calls "the paradox of sympathy" (Deleuze 1991, 35). The paradox comes about because sympathy tends to be partial, or what Hume calls "limited generosity," whereas society is possible only by the integration of sympathies, in which case 'sympathy' will have to "transcend its contradiction and natural partiality" (Deleuze 1991, 39-40). So, "what we see in nature without exception, are families" (Deleuze 1991, 39), which are mutually exclusive and bound by the passion of sympathy. The challenge of establishing a wider society consists in extending these sympathies across family boundaries. The natural partiality of "our natural uncultivated ideas of morality" (*Treatise,* 3.2.2.8) is ill-equipped to do this, because, as Hume explains, virtue and vice conform to what is inherent in our nature, and as partiality of affection is inherent in our nature, it "must not only have an influence on our behaviour and conduct in society, but even on our ideas of vice and virtue."

Thus “an invention is needed,” as Cohon (2010, 28) puts it. Exactly how this happens is what Hume sets out to show next. He shows how man overcomes the contrary force of partial affection (the immiscibility of the sympathies) and manages to establish and maintain society. This stems from man coming to appreciate the infinite advantages of society through his early education and later in his enjoyment of company and conversation. The chief advantage of society is that it greatly improves man’s ability to acquire the resources necessary for life. On the other hand, “the instability in their possession, together with their scarcity” is also a chief impediment to the maintenance of society. In fact Hobbes (1651, 56) argues that in the state of nature, in which this problem of the insecurity of property is real, “there is no place for Industry; because the fruit thereof is uncertain.” Hume notes that the chief source of interpersonal violence is the limited supply of the necessary resources for life or what is needed to meet man’s desires. So when men discover the advantages of society and at the same time observe that society is threatened by the insecurity of private possessions, they take steps to ensure the security of such goods.

This is a confluence from which key historic figures in moral and political philosophy diverge. At this hypothetical juncture in man’s social history, we have to our rear “man in the state of nature,” and ahead in time is political society. The lack of government and laws in the former, as well as its smaller size, distinguishes it from the latter. There are two issues of contention regarding man in the state of nature. In the first place, there is discordance among philosophers as to the conditions of man in the state of nature. Did man have a social and moral fabric? Secondly, was man better off in the state of nature than he is under government?

For Locke, the natural state of humans is “a state of perfect freedom to order their actions;” it is a state also of equality, in which “power and jurisdiction are reciprocal, no one having more than another.” (*Second Treatise of Government* (henceforth, STG), 4). Locke cautions however, that we should not misconstrue the state of liberty for a state of licence, for the state of nature has a law of nature to govern it, which obliges every one. That law, according to Locke, is reason, which teaches mankind, that being all equal and independent, “no one ought to harm another in his life, health, liberty, or possessions” (STG, 4).

For Hobbes, on the contrary, the “miserable condition of War” is the dominant feature of the state of nature, in which the natural passions of men are at complete liberty with “no visible Power to keep them in awe.” War is inevitable in a state of nature, given that man is by nature competitive, diffident, and has an affinity for glory. As a consequence, men would attack each other for gain, safety and reputation.

Rousseau aligns with Locke against Hobbes, suggesting that we cannot describe as misery the condition of man in the state of nature, in which he is a free being whose heart enjoys perfect peace and body perfect health (DOI, 27). He suggests a civil life is harder to support than a natural life. He does not think savage man complained so much about life as civil man does today. The savage, as he writes (DOI, 28), “had in his instinct alone everything requisite to live in a state of nature; in his cultivated reason he has barely what is necessary to live in a state of society.” Rousseau (DOI, 50) cites Locke, whom he admires, saying: “For according to the axiom of the wise Locke, Where there is no property, there can be no injury.” Consequently, Rousseau concludes, “Let us therefore judge with less pride on which side real misery is to be placed” (DOI, 27).

The key insight which some of these scholars share with Hume and which is most relevant to my project is the idea that living in social groups is not inherent to human nature in the same way that it is with the social insects, for example. Humans form (political) society against their natural inclinations. The main obstacle to this, according to Hume, is the insecurity of external goods; and the best way to ensure the security of these external goods, Hume suggests, is “by a convention entered into by all the members of the society to bestow stability on the possession of those external goods” (*Treatise*, 3.2.2.9). This establishes a general sense of common interest among the members of the society whereby they respect the rule to abstain from the possessions of one another.

It is on this convention to abstain from each other’s possessions that justice and injustice arise, as well as property, right and obligation (*Treatise*, 3.2.2.11). According to Hume, if nature had supplied all of man’s survival needs in abundance, and if men were of unlimited generosity and benevolence, justice would be rendered useless. He notes (*Treatise*, 3.2.2.17) that “married people in particular mutually lose their property, and are unacquainted with the mine and thine, which are so necessary, and yet cause such disturbance in human society."

The relationship between a man and his property is not a natural one, but a moral one, which is founded on justice. Thus, property and justice both originate from the same artifice. Rousseau (DOI, 54) also held this view of the artificiality of property and justice, and supports this position by citing Grotius as suggesting that "the distribution of lands produced a new kind of right; that is, the right of property different from that which results from the law of nature.[1]" For Hume therefore, "the convention for the distinction of property and for the stability of possession is of all circumstances the most necessary to the establishment of human society" (*Treatise*, 3.2.2.12). In line with this, Rousseau (DOI, 41) contends that "The first man, who, after enclosing a piece of ground, took it into his head to say, "This is mine," and found people simple enough to believe him, was the true founder of civil society." The link between property, justice and society is thus clear.

Scarcely anything else is needed, beyond the protection of property rights, for the maintenance of a perfect and harmonious society according to Hume. All the other passions are either easily restrained or pose no serious

---

[1] If by this statement Grotius implies there is some other justice prior to the rise of property, I do not wish to be side-tracked into a debate about that here.

threat to the functioning of society. It is the drive to acquire goods that is "insatiable, perpetual, universal and directly destructive of society" (*Treatise*, 3.2.2.12). No other affection of the human mind according to Hume is capable of counteracting the "love of gain." They are either too weak or merely "inflame the avidity" for possession. Incidentally, it is the very affection for possession that controls itself, and it does so by altering its direction. This self-restraint is motivated by its effect in preserving society, which in turn provides better opportunities for acquiring those external goods. Thus described is Hume's theory: Humans appreciate the advantages of society. But their natural selfishness and the partiality of sympathy oppose the formation of society. In order to maintain society, therefore, humans invent artificial moral rules, chiefly, justice. Thus, the passions, as Deleuze writes, "must either be satisfied artificially and obliquely, or be snubbed out by violence" (Deleuze 1991, 43). It is the violence that follows the absence of social order, which Hobbes graphically portrays in his *Leviathan*.

VII. Harmonization

Hume's account of the evolution of moral and social behaviour is the most systematic and most coherent yet. It is based largely on moral psychology and traces the progression to political society from the family through tribal society. According to Hume, man comes to appreciate the value of society through education from parents, and also through experience of the benefits he receives from family and close associates (*Treatise*, 3.2.2.9). So he restrains his counter-social passions of selfishness and partial sympathy in order to ensure the maintenance of society. In his talk of the “origin of society” (*Treatise*, 3.2.2), he emphasises the basis of society, the forces that need to be overcome for society to form and the forces that hold society together. His analysis tells us how our moral sentiments arise from our constitution as given by nature and their role in the sustainability of society, meaning thereby to refer to the artificial or social virtues, which are motivated by the experience of the benefits of society.

We have to remember that Hume’s references to ‘society,’ mean political society or nation states to the exclusion of families and small tribal groups, which he refers as man in his “wild and uncultivated state.” So as Cohon (2010, 28) explains, “It occurs to people to form a society as a consequence of their experience with the small family groups into which they are born, groups united initially by sexual attraction and familial love, but in time demonstrating the many practical advantages of working together with others.” Thus, the experience that motivates social cooperation is that of family or the small tribal group, both of which are not included in Hume’s notion of society.

In Deleuze’s (1991, 39) rendition, society is a collection of families which tend to exclude one another, due to partiality. Hume shows what holds such collections of families together. It is justice, chiefly, and the other artificial virtues. An evolutionary account needs to show how any two or more families coalesce into a society, or how cross-family sympathies or cross-family bonds could be established between individuals from different family units. Providing the factors and mechanisms by which such associations are sustained, as proposed by Hume, is essential to establishing political society. Thus, in establishing the conditions that make the formation of society possible, Hume provides the proper foundation that would allow the evolutionary biologist, who, as we shall see in the next chapter, strives to present a systematic model of social evolution, but builds on a mistaken foundation, namely the view of the social virtues as innate (genetic traits), when they are actually contrivances of humans.

Hume discusses three mechanisms by which society is maintained: inculcated habit, reciprocity and social sanctions. On the first mechanism, Hume (*Treatise*, 3.2.2.9) suggests that humans come to appreciate the importance of the social virtues through experience and education, and therefore endeavour to uphold them. That would be the “habit” explanation. He also suggests (*Treatise*, 3.2.2.10) that “it will be for my interest to leave another in possession of his goods, provided he will act in the same manner with regard to me.” This suggests the artificial virtues are also practiced on the expectation of reciprocity. In other words, we ‘trust’ the recipients of our just and benevolent behaviour to reciprocate them.

Hume places the greatest emphasis on the third mechanism for maintaining (political) society, i.e., the artifice of justice, which he discusses at length. In his discussion of government he suggests that no force in nature can overcome our inherent selfishness and partiality in their opposition to the social virtues. This suggests that the first two mechanisms are insufficient, for which reason humans invent government and magistrates, akin to Hobbes' Leviathan, to enforce the rules of justice. On this third explanation, Hume strikes a chord with Rousseau (DOI, 28), who suggests that man's instinct provided everything requisite to live in a state of nature, and that "in his cultivated reason he has barely what is necessary to live in a state of society." In other words, left to his nature alone, man would fail as a social being, and reason, according to Rousseau, does not sufficiently compensate for that (another challenge to the moral rationalist). Thus, as Hume explains it, because empathy by nature is strongest for our immediate family and weakens with diminishing relatedness until it is virtually non-existent for strangers, some artificial mechanisms need to be put in place in order to sustain society. All three mechanisms do feature in social behaviour, but they need to be properly combined and adjusted.

Another issue on which Hume's analysis stands out is his distinction between tribal societies and political or civil societies, which as we shall see in the next chapter, contemporary Darwinians fail to distinguish, and from which arises part of their theoretical difficulty with social evolution. Hume, like many philosophers of the seventeenth and eighteenth centuries, sees two kinds of human — either uncultivated savages, whom Hume does not believe to have or need government (*Treatise*, 3.2.3.1), or people who live in societies such as his "civilized" cosmopolitan eighteenth century British society, what political philosophers refer to as *political society*.

As Hume has suggested, and as I elaborate in the next chapter, the two kinds of society, tribal and civil, have drastically different properties, and are therefore sustained by different conditions and mechanisms. The “state of nature” is a term used to describe the conditions of people with no civil society or government. Many scholars, including Hobbes, Hume and Darwin have also used the term “savages” to refer to early or primitive human social groups that had neither civil society nor government and which give us the closest approximation to the hypothetical notion of “humans in the state of nature.” The term “savage,” suggests that such humans would have been brutish, quarrelsome or amoral. Hobbes in fact, even perceived humans in the state of nature to be solitary. In reality, primitive social groups, if they indeed give us a glimpse into the state of nature, were actually harmonious cooperative social groups, which were no more prone to quarrels than the different civil societies of today. Darwin (2012 [1871],[1] 2319) noted that savages that were the subject of contemporary anthropological study were deeply loyal to theirs groups. It has also been suggested that “hunter-gatherers engaged in extensive cooperation within residential bands” (Henrich et al 2006, 1286). This sentiment has also been expressed by Kaplan, et al. (2000), Hill (2002) and Fehr and Fischbacher (2003).

Even though Hobbes is best known for his denigrating characterizations of man in the state of nature, he never suggest that ‘man to man’ violence is

---

[1] Some of the unusually high page numbers are eBook locations.

exclusive to man in the state of nature. He clarifies this in the dedication of his *De Cive*, as he writes: “To speak impartially, both sayings are very true; That Man to Man is a kind of God; and that Man to Man is an errant wolf. The first is true, if we compare Citizens amongst themselves; and the second, if we compare Cities.” Hobbes here is basically pointing to the harmony within groups and the disharmony between groups.

As a matter of fact, the small tribal groups are the subject of intense study in modern anthropology. There is a rich anthropological literature on the social dynamics of such so called “primitive cultures.” Strictly speaking, we could never say humans ever evolved sociality, since there was never a time in the evolutionary history of humans that they were not social. The social evolutionary history of humans dates back to our pre-human ancestors. All hominid species leading up to modern humans are thought to have lived in social groups. So even though humans did not have political government in the state of nature, they were no less social than the humans of the political societies of today. The difference is only in group size.

In fact, some contemporary Darwinian explanations are plausible only for such small tribal groups, in which one-on-one interpersonal associations and trust and the expectation of reciprocity sustain the observance of the rules of possession. Thus, even though Hume alludes to expanding sympathy and trust beyond the family as a way to make society possible, he is clear that that is not sufficient for the maintenance of larger civil societies, which necessarily bring total strangers into social relations, and among whom the trust necessary for reciprocal respect for the rules of possession is not established. Within such large social groups, each individual will have a small sphere of family, friends and affiliations, with whom he will have a stronger social bond and greater empathy. In such local social pockets, reciprocity and the natural virtues may still be important.

Conditions in the (larger) political societies, on the other hand, are different from those in the family or the tribal group. Empathy, trust and the sense of group identity are all minimal in such large, often cosmopolitan societies. Even though people in such societies interact mostly with family and friends, their physical space is interfused with strangers, against whom they need to take all kinds of precautions to protect themselves and their property. To this effect, Hobbes (1651, 57) asks of man, “what opinion he has of his fellow subjects, when he rides armed; of his fellow Citizens, when he locks his doors; and of his children, and servants, when he locks his chests”? Consequently, as we see in political societies, justice is sustained largely by the force of law rather than the “moral character” of man. Civil societies therefore, are not sustained by natural instincts or inclinations. Civil societies are indeed Hobbes’ (1651) commonwealth, in which man’s compliance with the social covenants are guaranteed largely by the threat of the pain of punishment by some authority in whom man reposes the power to enforce the civil laws necessary to keep the artificial social group together. Thus, as an artifice of humans, civil society is not a thing that can be explained by a theory of adaptive evolution such as natural selection. This crucial distinction between tribal and civil societies should pave the way for a more coherent explanation of social evolution to be taken up in the next chapter.

VIII. Contemporary Accounts of Social Evolution

In this chapter, I have discussed the relevant aspects of Hume's moral theory, which I will be using in the next chapter to resolve the conundrum of a modern evolutionary explanation of social behaviour. As Hume proposed his ideas on morality nearly three hundred years ago, this may raise questions as to how relevant those ideas are today. In this section, I respond to such concerns by discussing some relevant empirical data from current neuropsychology in relation to Hume's key claims in social/moral psychology. I will also discuss how the recent directions in theoretical thought align with Hume's insight.

Recall that whereas modern Darwinism seeks to explain how a heritable genetic factor for morality, sociality or altruism might evolve, Hume (as well as Darwin) envisions no such single factor for moral or social behaviour. Instead, Hume found sociality to consist of several different components, both natural and learned. This discussion will show which of the two views the empirical data from modern neuroscience support.

Social psychologists, with empirical evidence from neuroscience, have now settled on the fact, first proposed by Wilhelm Wundt (1907) and more recently by Zanjoc (1980), that the human mind consists of an emotional part that is ancient and faster, and a more recent higher rational brain that is much slower

(Haidt, 2007). Hume made this observation in the middle of the eighteenth century. He noted (EHU V.II) that the most basic and routine operations of the mind are “so essential to the subsistence of all human creatures, it is not probable, that it could be trusted to the fallacious deductions of our reason, which is slow in its operations; appears not, in any degree, during the first years of infancy; and at best is, in every age and period of human life, extremely liable to error and mistake.” This idea is one of the most useful tools of psychological explanation today. In a bestselling book, *Thinking Fast and Thinking Slow*, Daniel Kahneman uses this principle to explain the way we make everyday decisions.

Hume stated that the primary mission of his moral enterprise was to debunk the notion of a battle between reason and emotion, in which we give preference to reason over emotion. Hume suggested that the two do not oppose each other, and that reason is actually the slave of the passions. Hume’s (and Spinoza’s (*Ethics* IVp4c) before him) seemingly counterintuitive view remained unheeded until recent decades. Wundt, reckoned to be one of the founding fathers of modern psychology, contended that “the clear apperception of ideas in acts of cognition and recognition is always preceded by feelings” (see Zanjoc, 1980). According to Zanjoc, however, this “affective primacy” view was lost in modern psychology. Until the last two decades, theories of moral psychology emphasized reasoning and higher cognition (Kohlberg, 1969). The present fMRI data show both cognitive and emotional processes to play crucial roles in moral judgment (Greene et al., 2001, 2004). In fact, a region in the human brain, the ventromedial prefrontal cortex (VMPC), has now been empirically identified as responsible for the emotional aspects of moral behaviour (Koniegs et al, 2007), whereas the higher cognitive centres do the rationalizing.

Actually, however, the “affective primacy” view of morality did not start with Wundt. As I have indicated above, the main objective of Hume’s theories of morality was to debunk the then prevalent view that morality is based on reasoning. At the beginning of his discussion of morals he indicates that he will be dealing with the question as to “whether it is by means of our *ideas* or *impressions* we distinguish between vice and virtue, and pronounce an action blamable or praise-worthy” (Treatise, 3.1.1.3). Hume poses this question in the context of the debate on the role of reason and emotion in moral judgments.

He notes that some people claim that “virtue is nothing but a conformity to reason” (Treatise, 3.1.1.4), and he questions “whether it be possible, from reason alone we distinguish between moral good and evil, or whether there must occur some other principle to enable us to make that distinction” (Treatise, 3.1.1.4). He concludes that morality goes beyond “the calm and indolent judgement of the understanding” (Treatise, 3.1.1.4) and that “it is in vain, to pretend that morality is discovered only by a deduction of reason” (Treatise, 3.1.1.7). Hume’s insight here, that morality involves both reason and emotion, is affirmed by recent empirical evidence from neuropsychology that suggests the involvement of both the emotional and the higher cognitive centres of the brain. If we ask today: where does morality come from, the answer, according to Yale University psychologist Paul Bloom (2010), is that “the modern consensus on this question lies close to the position laid out by the eighteenth-century Scottish philosopher David Hume. He thought moral reason to be ‘the slave of the passions’”. Bloom (2010) adds further that “Hume’s view is supported by studies that suggest that our judgements of good and evil are influenced by emotional reactions such as empathy and disgust.” I think Bloom’s summary delivers the verdict on my invocation of Hume’s insight as the beacon for a contemporary analysis of moral and social evolution.

The technical details of the neuro-psychological experiments that corroborate this resurgent Humean view are beyond the scope of my thesis. But the main rationale is that damage to certain brain regions that results in the loss of emotions, without any effect on reasoning, also results in the impairment of moral judgment. The rationale is stated quite concisely in the following excerpt, with supporting references, from Luo et al (2006):

> Thus, patients with damage to the ventromedial frontal cortex show no impairment for many aspects of reasoning yet are impaired in their emotional responses (e.g., Damasio et al 1990), their moral emotions (Eslinger et al., 1992; Eslinger and Damasio, 1985) and their moral behavior (Anderson et al., 1999; Blair and Cipolotti, 2000; Damasio, 1999; Eslinger and Damasio, 1985; Grafman et al., 1996).814– 834, 2001).

The other insight of Hume’s on moral evolution, which differs from the neo-Darwinian account but which is gaining currency in contemporary psychological accounts of morality, is his view that the seed of social morality is sympathy, which by nature dwells within the family but can be extended to other individuals given the right circumstances. Darwin expanded on this by suggesting that "parental and filial affections lie at the base of the social instincts" (page). He adds (1871, 2330-2331) that "as man advances in civilisation, and small tribes are united into larger communities, the simplest reason would tell each individual that he ought to extend his social instincts and sympathies to all the members of the same nation, though personally unknown to him." (1871, 2331-2332).

Emile Durkheim (1912) notes in his study of indigenous tribes of Australia and the Americas that each of the tribes grew from a single ancestor. This suggests that the sympathy that bonds a tribe actually started off with a small nuclear family. Freud follows in that tradition when he writes that "civilization is a process … to combine single human individuals, and after that families, then races, and nations, into one great unity, the unity of mankind" (1930, 111). This vision of a progressively widening circle of moral commitment has been the message of recent books such as *The Expanding Circle* (Singer, 2011) and *The Better Angels of Our Nature* (Pinker 2011). The difference between Hume and Darwin on this is the scope of the sympathy. As I pointed out earlier, Hume does not see sympathy extending appreciably beyond family and close associates.

Some other scholars have used this partiality of sympathy as a basis for further hypotheses. Asma (2012, 38), for example, affirms the suggestion that "presumably blood nepotism evolved first, and this chemically based behavior developed into wider (non-blood) networks of social cohesion." As I have cautioned, however, Hume is more tempered in his treatment of the outward radiation of empathy from the nuclear family. In Hume's account, empathy exists to no appreciable extent beyond family and friends. This property of empathy is crucial for explaining social evolution in the next chapter.

Hume's claim that justice and the respect of property rights are necessary and sufficient for the establishment and maintenance of society is also reflected in contemporary thought. In Freud's view, the first act of civilization was the attempt to regulate the mutual relations of individuals, without which it would have simply been the rule of the strong (1930, 74). It is only when a majority come together to yield a force which is stronger than that of any individual that social life becomes possible, according to Freud (1930, 71). The force of the group thus becomes "right," rather than that of the individual, which is considered to be "brute force." As he writes, "the replacement of the power of the individual by the power of a community constitutes the decisive step of civilization" (1930, 71). In Pinker's (2011) account, this collective power of the community constitutes a single central force – The State (or what Hobbes called the *Leviathan*) – monopolizes violence, and becomes one of the potent drivers of modernity.

In a similar analysis, Boehm (2000) contends that "moral communities arose out of group efforts to reduce levels of internecine conflict, as well as to avoid undue competition, domination, and victimization" (85). In a view similar to Freud's so-called primordial hordes, Boehm (2000) also suggests a collective rebellion of subordinates that topples an oppressive ruler and egalitarian rules thereafter instituted to forestall such repressive behaviours. He writes, "if resentful subordinates manage to collectivize and institutionalize their rebellion, you have a human type of politically egalitarian society, in which there is a major tension between the group and its more rivalrous individuals" (84). He adds, "in effect, a large, *ad hoc*, community-wide political coalition serves as watchdog over individual behaviors that could lead to victimization of others, or to conflict within the group" (80). Even though I would not necessarily concur with the detailed mechanisms suggested by some of these scholars, I cite them to highlight the principle involved in the evolution of (civil) society – the constitution of a powerful authority to ensure compliance with binding rules. It is an artificial human enterprise, which is not amenable to explanation by natural selection, a theory designed for adaptive evolution.

IX. Summary and Conclusion

Hume's theory of morality can be summed up thus: nature has not adequately equipped man to be able to survive in his environment. Man depends on society to compensate for his deficiencies. However, man is by nature selfish and of "limited generosity." This, in conjunction with the scarcity of resources and unlimited wants, as well as the "easy change" of goods (absence of the concept of property), is not conducive to the maintenance of society. In order to ensure the maintenance of society, man needs to secure "external goods" (i.e., establish property rights) from the violence of man, which he does by inventing the artifice of justice – the central pillar of social morality and the glue of society. This statement may appear contentious to some, but most of the questions arising from it may be answered if we unpack the following from Hume (*Treatise*, 3. 2. 2. 23): “Thus self-interest is the original motive to the establishment of justice: but a sympathy with public interest is the source of the moral approbation, which attends that virtue.” Even though justice starts of as a project of self-interest, humans ultimately come to respect it (as a virtue) for the sake of society, which they value. Hume’s theory of morality is descriptive and naturalistic. It shows how social virtues, the key being justice, which are taken to be foundational by some of the prescriptive moral theories, are in fact not foundational at all. In the next chapter, we shall see how that insight, in particular, is key to resolving the problem of social evolution under Darwinian theory.

Chapter Four

Modeling Modern Society

I. Introduction

In chapter one I explained why scientific explanation is prone to error and how contemporary evolutionary explanation is not immune from that truism. Then in chapter two I described the problem of social evolution, and identified two fundamental modeling assumptions that I think are in error and are the reasons why a compete and satisfactory resolution to the paradox of social evolution has proven to be so elusive. The first is the severely narrowed conception of social behaviour as *altruism* (defined as other-regarding behaviour). The second is the assumption that social behaviour is a single locus trait. The discussion of Hume's social/moral theory in chapter three revealed a third major flaw in contemporary evolutionary explanatory models. Hume argued that the social bond that sustains the early simpler human societies is completely different from the mechanism that sustains modern political society. Under contemporary evolutionary modeling, this distinction is generally overlooked. In this fourth and final chapter, I present critical details of the models of social evolution given by Darwin and later Darwinians under the *Modern Synthesis*. We will see that the two differ in important ways. I will subsequently present Hume's account, which suggests some of the ways contemporary evolutionary modeling assumptions must change if we are to have a complete, coherent and robust genetical explanation of social evolution.

II. The Darwinian Explanation

The basic logic of natural selection, as explained by Darwin (1859, 91), is that, “as more individuals are produced than can possibly survive,” a “struggle for existence” inevitably ensues in which those individual differences and variations that are favourable are preserved and those that are injurious are eliminated (Darwin 1859, 108). This dual process of preservation of some and elimination of others, results from the differential rates of survival and reproduction amongst individuals according to their differential fitness for the given conditions of life. Such a scheme would work only if the traits that confer the reproductive advantage to individuals are transmitted to their offspring. That is why Darwin (1859, 168), suggests that natural selection relies on “the strong principle of inheritance,” emphasizing that “any variation which is not inherited is unimportant for us” (31). Thus, any trait that is not heritable is not amenable to explanation by natural selection.

Darwin’s theory of natural selection was formulated prior to the advent of modern genetics. Consequently, in spite of the cardinal importance of inheritance in the theory, Darwin lacked a robust theory of heredity. In Darwin’s day, as we saw in chapter one, *blending inheritance* was the reigning theory of heredity, to which he subscribed (Fisher, 1929, 1). This disadvantage, however, appears to have served Darwin well. For, unencumbered by any such theoretical constraints as the *phenotypic gambit* of the Modern Synthesis, Darwin was able to tackle the problem of social evolution with greater latitude. Consequently, as we shall see here, his explanation of social evolution was much more coherent than today’s explanations under the Modern Synthesis, which try to construct models based on binary phenotypes and genotypes.

The curious question would be how subsequent supposed improvement upon a theory would lead from a more coherent explanation to a less coherent one. This is a fallout from the adoption of Mendelian genetics as the hereditary theory of natural selection. Recall from chapter two that two watershed events in the history of Darwinian theory have been profoundly instrumental in shaping contemporary evolutionary explanation. The first was during the 1930s and 1940s when the modern synthesis took shape, and the second occurred in the 1960s and 1970s, during which time the synthesis, for better or worse, took on a much narrower explanatory focus, with the emergence and ascendency of the "gene's-eye view" of Darwinian evolution and its systematic rejection of the other explanatory traditions within the synthesis.

With the adoption of Mendelian genetics, under the modern synthesis, a binary conception of genotypes as well as phenotypes (Van Veelen et al. 2012, 68) emerged as its fundamental heuristic. Thus, according to Grafen (1984, 64), in modeling the evolution of behaviour (altruism for example), the behavioural ecologist takes a leap of faith, and supposes that there is an allele for one phenotype and a contrasting allele for the other. As a grand illustration of this, Hamilton’s (1964) inclusive fitness hypothesis is largely considered the best attempt since Darwin to deal with the problem of social evolution; and he is explicit that his theory of social behaviour assumes a trait that is determined exclusively by genotype (1964, 16). Trivers (1971), similarly, subscribes to this genetic view in his reciprocal altruism theory.

As an illustration of the neo-Darwinian approach to social evolution, consider the following scenario from Bshary and Bergmuller (2007). After two friends have had dinner, one of them offers to pick up the tab. Bshary & Bergmuller (2007) suggest there are four perspectives from which that behaviour can be explained. The first three they suggest are biological explanations, including Hamilton’s inclusive fitness and game theory, the fourth perspective being the cultural, moral or psychological explanation that the social scientist would give. With the biological explanations, the approach of the evolutionary biologist is to impose, first and foremost, the standard binary template of social individuals, i.e. the individual who would pick up the tab is designated ‘altruistic,’ and those who would not are selfish (Maynard Smith 1998, 639). Alternative binary labels, particularly in the case of game theory, include: cooperators versus defectors, doves versus hawks, suckers versus cheaters, etc. After designating the two phenotypic traits, we then proceed, according to Grafen (1984, 63), “as if there were a haploid locus at which each distinct strategy was represented by a distinct allele.” Alger and Weibull (2012, 42) see this as the usual evolutionary modeling assumption, and describe it as “when a trait is an action always to be taken,” as opposed to the situation in which the trait expresses one behaviour or its alternative contingent upon some exogenous factor(s).

The ascendency of the “gene’s-eye view” or “genic” movement (see chapter two) meant that all legitimate Darwinian evolutionary explanations were to be sifted through the mesh of population genetics. As population genetics is “today still considered the theoretical-mathematical backbone of evolutionary biology” (Pigliucci and Muller 2010, 6), the conception of heredity in population genetics is, *ipso fact*o, the conception of heredity in contemporary Darwinian explanation. In this way, as the primary focus of population genetics consists in calculating the changes in the gene frequencies of populations, we come to conceive of evolution under modern Darwinism simply as a change in gene frequencies.

However, this gene frequency calculation was not the standard approach to modeling social evolution during the 1930s and 1940s when the Modern Synthesis was being crystallized. It appears that scholars were for the most part content with the explanations Darwin had given for the evolution of eusociality (1859) and human sociality/morality (1871), until 1964, when W. D. Hamilton reported a problem. It appeared that putting the evolution of social behaviour through the “gene’s-eye view” of the Modern Synthesis presented a theoretical anomaly. As Ratnieks and Helantera (2009, 3169) report, “The recognition that altruism is an evolutionary puzzle, and the solution was to wait 100 years for William Hamilton.”

I should point out, Ratnieks and Helantera’s estimation that the problem was undiscovered for a hundred years is not quite accurate. In fact, Darwin

(1871) had noted that human sociality presented a problem for natural selection. As he wrote: “it hardly seems probable, that the number of men gifted with such virtues [altruism], or that the standard of their excellence, could be increased through natural selection, that is, by the survival of the fittest” (Kindle Location 2705).[1] Darwin did not stop at that. He offered a meticulous explanation as to how human sociality/morality becomes possible.

Darwin’s explanation, as I have indicated, was for the most part considered satisfactory. And social behavior was for a time not considered a serious threat to Darwinian explanation until Hamilton in 1964; and as I have explained above, the resurgence of altruism (social behavior) as an evolutionary puzzle is a direct consequence of the Modern Synthesis, which shaped up in the 1930s and 1940s. The Modern Synthesis had modified Darwinian explanation considerably with its adoption of Mendelian genetics. So when W. D. Hamilton raised the issue of altruism in 1964, it was specifically with respect to population genetics — the Mendelian based, abstract mathematical modeling of natural selection under the Modern Synthesis. He writes (1964, 1): “With very few exceptions, the only parts of the theory of natural selection which have been supported by mathematical models admit no possibility of the evolution of any characters which are on average to the disadvantage of the individuals

---

[1] Other eBook citations herein will tend to have these unusually high page numbers.

possessing them" (1964, 1). Of course, social behaviour, which is generally reduced to altruism under the Modern Synthesis, is held to be disadvantageous. Thus, Hamilton continues, "If natural selection followed the classical models exclusively, species would not show any behaviour more positively social than the coming together of the sexes and parental care" (1964, 1). The "classical models" Hamilton refers to are the population genetic models, not models prior to the Modern Synthesis. Consequently, the problem Hamilton was raising was only about a couple of decades old, rather than a century old, as Ratnieks and Helantera supposed.

The issue Hamilton was raising pertained to the extension of population genetics, the modeling tool of the Modern Synthesis, to explain social behavior. This means supposing a gene or allele for social behavior (altruism) and showing how it would evolve by natural selection in a population of non-altruists. As we shall see shortly, that is exactly the program of sociobiology. However, as Wilson (1975, 3) states, "the central problem of sociobiology [is] how can altruism, which by definition reduces personal fitness, possibly evolve by natural selection?" This was the problem Hamilton noted in 1964, which would have been only two or three decades old in the wake of the Modern Synthesis.

In fact, explaining the evolution of altruism has dominated theoretical evolutionary biology ever since Hamilton. It has been a very contentious field of research, all emanating from the drive to explain social behavior as a single genetic trait. However, the evolutionary totality of humanity encompasses biology and sociology, and the question is, how much of human social behaviour is attributable to innate biological traits. This question is central to the highly contentious and very intractable “nature versus nurture” debate. As some, such as Sidgwick (1876, 56) contend, “we cannot use any of the physical reasonings commonly used to establish the doctrine of Evolution to prove any ethical position.”

On the other hand, there exists a vigorous theoretical project of contemporary evolutionary biology, which E. O. Wilson has christened “sociobiology” in his (1975) book of the same title. Wilson defines sociobiology “as the systematic study of the biological basis of all social behavior," which he suggests, “should be compounded of roughly equal parts of invertebrate zoology, vertebrate zoology, and population biology.” Elucidating this proposal from Wilson’s other works, Kary (1984, 163) writes:

> What Wilson foresees as the biologicization of ethics …, is the grounding and direction sociobiology can give to ethical justification. Ethicists can offer satisfying, normative recommendations by showing how one ethical theory is a better fit to the genetic constraints of Human Nature than others, and hence why that ethical theory will be the most fruitful theory to adopt.

This “biologicised” or “biologically informed theory of ethics” was rejected by most philosophers, according to Ruse (1984), even though, he himself was sold on it and he then undertook to persuade other philosophers. He writes (1984, 168): I have come to the decision that he [Wilson] is right. Only by setting normative beliefs and behaviour against an evolutionary background can we hope to achieve a full understanding of morality. This is not a position I enjoy taking. … But, I believe it is true.”

The fundamental modeling approach in this sociobiological project is what Allan Grafen (1984) describes as the “Phenotypic Gambit,” whereby we represent social behaviour as a single locus haploid allele or gene.

With human social evolution, however, contemporary evolutionary researchers are coming to terms with the reality that not all human social behaviour is genetically determined, even though they would still like to conceive of all social behaviour as having evolved. As a matter of fact, some professed “enthusiastic Darwinians” known to be in the forefront of the Wilsonian “biologicization of ethics” project, have had to concede the impracticability of the sociobiological research program. Here is Dawkins (1976 [1989], 191): “The argument I shall advance, surprising as it may seem from the author of the earlier chapters, is that, for an understanding of the evolution of modern man, we must begin by throwing out the gene as the sole basis of our ideas on evolution.” This entails renouncing what he had earlier argued to be the fundamental basis of evolution by natural selection, i.e. the gene. Notice that Dawkins is not suggesting we throw out the gene entirely, but only “as the sole basis.” Thus the contention still remains as to the manner and extent of the gene’s involvement in human ethical behaviour.

Some contemporary theorists see heritable human social behaviour to consist of two parts – that transmitted through genes and that transmitted through copying (Nowak, 2006; ref). Other scholars have suggested a gene-culture coevolution (Boyd and Richerson 2009), whereby initial small groups are formed by genetic evolution followed by a culturally driven evolution of larger societies. Under this scheme, small societies with the fittest cultures spread and supplanted groups with the less fit cultures.

This state of affairs lends credence to the perspective I espouse regarding contemporary evolutionary explanation, particularly as it relates to the problem of social evolution. As I explained in chapter one, the Modern Synthesis was motivated by the quest for a hereditary theory to complement Darwinian natural selection. However, the adoption of Mendelian genetics as the hereditary theory of natural selection imposed some conditions that are simply inconsistent with the nature of social behaviour. Consequently, human morality remains an unresolved problem for contemporary Darwinian evolutionary explanation (Boehm 2012; Boyd & Richerson, 2009; Johns et al 2009, 17452; Bshary & Bergmuller 2007;Bernhard et al 2006, 914; Wilson 2005, 159), as there is no account of its evolution that is completely satisfactory.

Evolutionarily, altruism is defined to fail (Nesse 2000,_228; Wilson and Wilson 2007, 329). We explain its prevalence notwithstanding, by shifting away from that foredooming definition in our models. As Trivers (1971, 35) puts it, most of the models are "designed to take the altruism out of altruism." It is generally acknowledged by scholars of evolutionary theory that altruism, as defined, cannot evolve unless certain other factors are at play. (Allen et al. 2012; Nowak 2012;

Taylor and Nowak 2007, 2282). Those factors, as variously proposed, tend to be inconsistent with the genetic view we have of altruism under the modern synthesis. In other words, the conditions or mechanisms we propose to allow for the evolutions of altruism actually contradict or undermine the descriptions and modeling assumptions that make the evolution of altruism impossible.

Hamilton's inclusive fitness hypothesis, for example, explains altruism by stipulating that the recipients of altruism be related to the altruist by descent, in which case they will have a high probability of carrying the altruistic gene. In fact, Hamilton (1963) stresses that it is not enough that the recipients of altruism are generally related to the altruist. For the theory to be successful, the altruist and the recipient of altruism must share the hypothesized specific gene, G, for altruism. This boils down to a requirement that the altruistic gene be present in both the altruist and the recipient. This, in the case of the social hymenoptera, for example, means the selfish queen carrying the altruistic allele, since the queen is the recipient of most of the altruism. This goes against Hamilton's modeling assumption that the altruistic and selfish behaviours are determined strictly by genotype. Hamilton (1964) also suggests that inclusive fitness would be enhanced if it is accompanied by some capacity to recognize genetic relatives and to ensure that they, rather than non-relative receive the benefits of altruism. If so, then, the altruists would be behaving selfishly towards non-relatives. This would then suggest that the altruists must have the capacity to express both the altruistic and selfish traits. This again goes against the “strictly by genotype” assumption.

In Trivers’ (1971, 35) *reciprocity model,* “each individual human is seen as possessing altruistic and cheating tendencies, the expression of which is sensitive to developmental variables.” Trivers suggests moral virtues such as “friendship, dislike, moralistic aggression, gratitude, sympathy, trust, suspicion, trustworthiness, aspects of guilt, and some forms of dishonesty and hypocrisy, are important psychological adaptations that regulate the altruistic system” (1971, 35). Again, it is not realistic that all these traits would be represented by a single mutation, which is the modeling assumption.

In another example of how the solutions we propose for the problem of altruism undermine our basic modeling assumption, Alexander (2009, 35) writes that “population members typically enforce social behaviors and rules having normative force by sanctions placed on those failing to comply with the relevant norm; and the presence of sanctions, if suitably strong, explains the persistence of the norm.” There is a lot glossed over in a statement like that. If this is the way (some) social groups, and most certainly human societies, are maintained, do our standard genetic models or any special models reflect this state of affairs? This is problematic for the binary conception of phenotypes and genotypes under current genetic modeling. First, where does this norm enforcing behaviour fall in our binary model? It is a social trait that needs to be accounted for. Is it an altruistic behaviour or a selfish behaviour? More pointedly, in a social population consisting of altruists and selfish members, which of them execute these sanctions that clearly constitute harm?

Sober and Wilson (1998) suggest that “within-group” selection is suppressed by the homogenization of social behaviour through the imposition of social norms. Is that alone not sufficient to maintain social behaviour in an all-selfish population? In other words, if there are such things as policing, social norms, strong reciprocity, etc., that encourage or enforce altruistic behaviour in others, regardless of their genetic status with respect to altruism, does that not undermine the idea of distinct genetically altruistic and selfish individuals in the populations? In other words, if every individual in the group, regardless of genotype, can behave altruistically under some environmental circumstance, — coercion in this case — is that not the sort of trait we would normally describe as non-genetic? In fact, some scholars are quite explicit on this. Fehr and Fischbacher (2003, 787) assert that “it is not possible to infer the absence of altruistic individuals from a situation in which we observe little cooperation. If strong reciprocators believe that no one else will cooperate, they will also not cooperate.”All that is needed to maintain social behaviour in such cases, then, is not an altruistic gene, but an *enforcer* gene, or rather some enforcement mechanism. If you behave altruistically because you want to avoid the pain of sanctions, then your altruism does not come from your nature, but is rather a consequence of some other nature of you or combination of natures of you.

Rationality will certainly be one. That is why, as products of reason, Hume suggests such behaviours are artificial.

Hume's analysis also gives us cause for pause in the view of social evolution as a battle between altruism and selfishness, or the view of social populations as consisting of genetically selfish individuals and genetically altruistic individuals. In Yakubu (2013, 110), I demonstrate a clear pattern (using key examples of altruism in nature) "that an individual will respond altruistically only when certain environmental circumstances are present, and would respond selfishly if those environmental cues were lacking. It is no different from the arctic fox expressing white fur in the winter and brown in the summer." In that set of examples (Yakubu (2013, 112), it was remarkably consistent that the altruistic and selfish phenotypes were determined by environmental circumstances rather than genotype. Boyd and Richerson (2009, 3283) also note that "If cheaters are despised by others in their group … and, as a consequence, suffer social costs … then they may be motivated to cooperate, even though prosocial motivations are entirely absent from their psychology." This again undermines the modeling assumption of distinct altruistic and selfish individuals in social populations.

III. Darwin's Account of Human Sociality

The eighteenth century British philosopher Thomas Reid made the astute and succinctly rendered observation that: "Of the various powers and faculties we possess, there are some that nature seems to have planted and reared, so as to have left nothing to human industry” (Reid 1765, 5). Such powers Reid thinks we share with the brutes, and are necessary for the preservation of the individual and the kind. There are other powers, according to Reid, “of which nature has only planted the seeds in our minds, but has left the rearing of them to human culture” (Reid 1765, 5). These will include morals and intellectual capacity. As we shall see in the ensuing discussion, this is the critical distinction the neo-Darwinian moral theorists miss in their genetical models of altruism and moral/social evolution. In their models, they treat traits that are nurtured as if they are planted and reared by nature.

That explains Darwin’s greater success in explaining the evolution of social/moral behaviour than the later explanations of the Modern Synthesis. Unlike the latter, Darwin did not construe social behaviour as a single on-or-off heritable trait (such as the “altruism” of contemporary modeling) that evolves by natural selection. The social instincts that give “impulse to some of our best actions,” according to Darwin, are rooted in our savage or even primate ancestors. The vast majority of people’s actions however, according to Darwin, are “determined by the expressed wishes and judgement of their fellow men and … their own strong selfish desires” (2012 [1871], 2283). Consider for example, a person who, upon the sight of someone drowning, jumps into the water to save that drowning person. Often, Darwin (1871,) explains, “such actions are performed too instantaneously for reflection, or for pleasure or pain to be felt at that time," which suggests that they are purely instinctive acts.

With the above illustration, Darwin (1871) notes, consistently with Reid's first category of instincts, that even though people have no special instincts to tell them how to aid their fellows, we still have the impulse, or instinctive love and sympathy retained from an extremely remote period. Also consistently with Reid, Darwin (1871 …) asserts that people, "with their improved intellectual faculties would naturally be much guided in this respect by reason and experience." According to Darwin, on rare occasions, our primitive social instinct will compel people to act impulsively. On most other occasions however, human social behaviour is reasoned and influenced by circumstances.

Darwin's account of the evolution of social/moral behaviour is based on what is today called *group selection*. I must caution, however, that Darwin's group selection account differs in important ways from today's modeling of it under the Modern Synthesis. Darwin notes that if a social group has more (members that are) "courageous, sympathetic and faithful members, who were always ready to warn each other of danger, to aid and defend each other," such a group would be more successful than other groups whose members lack these moral qualities. Consequently, he writes (2012 [1871], 2691): "A tribe rich in the above qualities would spread and be victorious over other tribes …. Thus the social and moral qualities would tend slowly to advance and be diffused throughout the world." Simply, a socially cooperative and cohesive group will be evolutionarily fitter than a comparable group whose members are uncooperative; and as groups compete for the same resources, the fitter cooperative groups will outcompete the non-cooperative ones and eventually supplant them.

The above, however, is only half of the explanation. The other half, and in fact, the harder part, Darwin poses as follows (2012 [1871], 2691): "But it may be asked, how within the limits of the same tribe did a large number of members first become endowed with these social and moral qualities, and how was the standard of excellence raised?" This is not a question germane to group selection theory alone. It is a fundamental question every evolutionary model must aim to answer. It is what E. O. Wilson (1075, 3) called "the fundamental question of sociobiology;" and it is the very question, if we recall, that Ratnieks and Helantera (2009) claimed no one had pondered over prior to W. D Hamilton in 1964.

Darwin (1871, 163) notes that: "It is extremely doubtful whether the children of such [altruistic] individuals would be reared in greater number than the children of selfish and treacherous members of the same tribe." He elaborates (2012 [1871], 2700):

> He who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature. The bravest men, who were always willing to come to the front in war, and who freely risked their lives for others, would on an average perish in larger numbers than other men.

Thus, even though a group with altruistic members fairs very well against other groups, the feature that gives the group that advantage appears not to be sustainable under natural selection. This is what is today called the *within-group* *selection* problem. So, how are the numbers of the “good social individuals” raised and maintained in such groups?

Darwin proposes a number of paths, or what he calls “some of the probable steps,” through which such virtues might evolve and persist in a population. The first path to social evolution he describes is as follows (1871, 163): “as the reasoning powers and foresight of members improve, each man would soon learn from experience that if he aided his fellow men, he would commonly receive aid in return.” There are two points to note here. First, this first path hinges on what is today called “reciprocal altruism,” and credited to Robert Trivers (1971). Notice, secondly, that Darwin does not propose a heritable trait for helping or cooperating with others, as contemporary population genetic models try to do. Instead, it is *reason*, for Darwin, which determines whether one group member helps another or not. He unfolds the rest of this particular roadmap to a moral society, saying that (2012, 2711):

> From this low motive he [man] might acquire the habit of aiding his fellows; and the habit of performing benevolent actions certainly strengthens the feeling of sympathy which gives the first impulse to benevolent actions. Habits, moreover, followed during many generations probably tend to be inherited.

This final step in this particular model is problematic, and I shall return to it.

The penultimate step, however, is quite plausible. Sympathy is an evolved and innate trait, which is an undisputable basis for benevolence. If we make a habit of regularly performing benevolent acts, even for other (“low”) motives, for example, as in this case, the hope that the benevolence will be reciprocated, our empathy will tend to expand. This is because, a feeling of association and affiliation develops between us and other individuals with whom we have reciprocal benevolence, and we tend to have some empathy for any one with whom we feel any kind of affiliation.

Darwin proposes a second path to moral behaviour in a community, which is not based on the expectation of direct reciprocity. Consider for example, virtuous acts that involve self-sacrifice, where one does not live to benefit from any reciprocation. Darwin proposes *praise and blame* as another feature of human nature that is an even “more powerful stimulus for the development of the social virtues.” Darwin suggests that humans by nature love to be praised, and dislike being blamed. Darwin asserts in several places (2013 [1871], 2336, 2716, 2368) that this follows from sympathy, but he does not show how, and I will propose an explanation here. It is that not only do we have sympathy for certain individuals, we also crave sympathy from those for whom we have sympathy. This is because we are benevolent towards people for whom we have sympathy and therefore we expect benevolence from people who have sympathy for us. We also know however that people who earn our praise also earn our sympathy, and those whom we blame also lose our sympathy.

To continue with the steps toward moral behaviour in a population, the love of praise and blame inspires courage in a population and it is spread through emulation rather than by genetic inheritance. As Darwin explains, if a person who, even though lacking the innate sacrificial instinct, is motivated by glory to sacrifice his life for his group, the glory he gets for his courage would inspire others to his courage. In this way, according to Darwin, such a person will do far more good for his tribe than begetting children who will inherit his trait, as he will influence far more people this way (2013 [1871], 2723).

In contemporary explanations, this idea of non-genetic transmission of some behaviour has been the broad strategy adopted by Boyd and Richerson in their numerous works on what they call “gene-culture co-evolution,” which they adopt as a strategy to explain the evolution of large human social groups of unrelated humans. Their premise is that “Cultural adaptation is much more rapid than genetic adaptation. … Thus a shift from genetic adaptation to cultural adaptation should greatly increase the heritable behavioural variation among groups” (Boyd and Richerson 2009 p. 3284).

However, they pitch “culture,” specifically artifacts and superior technology, as the non-heritable component of social behaviour that spreads rapidly in one group through copying and enabling that group to conquer other groups. The problem with this model is that the rapidly spreading cultural phenomenon is some advantageous technique or technology rather than altruistic behaviour. The models do not suggest that altruistic behaviour is the thing that is copied, or demonstrated any motivation for altruistic behaviour to be copied in the group. Under this scheme, while the number of altruists may be increasing as a result of the increasing total population of its home group, there is no mechanism to stem the falling frequency of altruists within the group. The superior technology which has rapidly spread within the population is not in itself a social glue and does nothing to compensate for the fitness disadvantage of the altruists within the population. So their frequency will continue to fall, which may lead to the eventual collapse of the society.

So for Boyd and Richerson (2009, 3284-5), “once rapid cultural adaptation in human societies gave rise to stable, between-group differences, the stage was set for a variety of selective processes to generate adaptations at the group level.” Thus, the stage is set for group selection between cultures. The problem is that culture spreads largely by assimilation of groups rather than supplanting of groups by groups. This incidentally is what Boyd and Richerson assert with their Nuer - Dinka example as well as the studies they cited from New Guinea. In fact, they go on to assert very clearly that: “ancient imperial systems often expanded militarily but the durable ones, such as Rome, succeeded by assimilating conquered peoples” (Boyd and Richerson 2009, 3285). Cultural assimilation simply boils down to culture spreading from group to group. That is clearly not social evolution, which Boyd and Richerson purport to be explaining. In fact, it is even not cultural evolution, which they sometimes call their model. If the Roman Empire was perceived as a single large multi-ethnic community, it was because all the diverse communities were under a single Roman government and system of laws – Hume’s “human artifacts.” Culturally, however, they had nothing in common.

The case of the European invasion of North America is a rare exception that comes close to total supplanting rather than assimilation. But even that does not explain the evolution and maintenance of large cosmopolitan societies, which we consider our task. One group successfully exterminating another group and occupying their geographic location does not do away with the barriers to the establishment of large groups of unrelated people that Hume identified. Boyd and Richerson's "cultural evolution explanation" is therefore not satisfactory,

Darwin's model provides a more robust mechanism for maintaining non-inherited altruism in a group within which genetically inherited altruism would be declining. However, we still see Darwin's limited understanding of heredity blemish his otherwise thoughtful explanation, as he takes a striking Lamarckian turn in talking about habits that we form from reasoning. He suggests, as we saw in his first model discussed above, that "[such] habits, moreover followed through several generations probably tend to be inherited" (1871, 164). That a learned behaviour will eventually become heritable if it is transmitted through enough generations culturally is manifestly inconsistent with the understanding of heredity today. In fact, it is the very essence of Lamarckism, which modern Darwinians uncompromisingly reject.

Elsewhere Darwin writes about moral behaviour in heritable terms: "In regard to the moral qualities, some elimination of the worst dispositions is always in progress even in the most civilized nations. Malefactors are executed, or imprisoned for long periods, so that they cannot freely transmit their bad qualities"

(Darwin, 1871, 172). Transmit their bad qualities? Is this through heredity or through cultural influence? I think Darwin here means heredity; and the elimination he talks about is by nothing other than natural selection. To be charitable to Darwin, we may take it that the "bad qualities" the malefactors transmit are not straight antisocial behaviour, but rather low reasoning capacity that leads to antisocial behaviour. This is because he explained the positive social behaviours not as inherent instincts but deriving from the interplay of reason and some other factors. So we assume that he would have a similar modeling conception of the anti-social traits.

So here is the problem with Darwin's account of social evolution. His account seems to avoid the problematic assumptions of population genetics, where social/moral behaviour is reduced to a monolith called "altruism," which is expressed by a single allele. Instead, Darwin shows that most of the behaviours we would call "altruistic" are not directly heritable instincts, but rather products of the interplay between reason and other instincts that are not in themselves "altruism," such as sympathy and sensitivity to the approbation of fellow group members. He gives a plausible step-wise account of the process. However, Darwin did not leave it at that. Instead, he goes on to tarnish that beautiful piece of work by suggesting that the contingent moral behaviours that are determined by rational judgement tend to be inherited if they are made into habits and followed over many generations. But even if we forgive or ignore this problematic view of hereditary by Darwin, his account of moral evolution is still incomplete.

This is because most social interactions today are driven neither by sympathy nor sensitivity to the judgments of other members of society. I take that up next.

IV. The Community - Society Dichotomy

Let us consider again Darwin's account of social evolution as encapsulated in the following passage (Darwin 2012 [1871], 2283): "But as love, sympathy and self-command become strengthened by habit, and as the power of reasoning becomes clearer, so that man can value justly the judgments of his fellows, he will feel himself impelled, apart from any transitory pleasure or pain, to certain lines of conduct."

This model, as I have suggested, may explain the evolution and maintenance of cooperation in certain societies, but not all. Today, the most essential and prevalent social groups are nation states such as Canada, USA, Germany, China, etc. In these societies, empathy and the judgment of fellow individuals play little role in most social interactions. Ironically, that is the kind of society Darwin thought he was explaining, as he frequently drew parallels and contrasts with what he called "savages" or "rude people." But actually, he was not; and Hume's account shows that quite clearly. As Hume characterizes it: "so barbarous and uninstructed are all societies on their first formation that many years must elapse before these can increase to such a degree, as to disturb men in the enjoyment of peace and concord." Hume suggests that these primordial social communities must grow in size and wealth, which will then serve as triggers for artificial mechanisms to be devised to maintain them as harmonious social organizations. That is why contemporary modeling could benefit from such insights from Hume.

Contemporary evolutionary biologists recognize two broad types of animal society. The first is *eusocial* society, which is distinguished by its characteristic reproductive division of labour, where a few members are dedicated to reproducing new members for the community while the rest of the population do not reproduce but work to maintain the community. The other is *non-eusocial* society, in which reproductive division of labour, if present, is not nearly as extreme. Largely, each individual sees to its own reproduction, but cooperates in some way with the other members of the group for their mutual benefit. Within the non-eusocial groups, in which human societies fall, evolutionary biologists recognize no further categories that are relevant for modeling purposes. As we saw in chapter two, and have been frequently reminded in this chapter, evolutionary biologists model social behaviour simply as *altruism*, which supposedly evolves all at once from non-social (or selfish) individuals.

The other feature of contemporary evolutionary modeling that makes social evolution problematic is the representation of every adaptive trait, however complex, as a single genetic mutation. This collapses the evolution of a complex trait such as social behaviour into one giant evolutionary step that is triggered by a single mutation and then spreads in the population. As Grafen (1984, 64) puts it, we proceed ‘‘as if enough mutation occurred to allow each strategy to invade.’’ Examples of such strategies are “always cooperate,” “always defect,” “tit for tat,” etc. This is purported to be Darwinian, even though Darwin emphasized that the evolution of complex traits (which clearly must include social behaviour) must proceed gradually over time. Darwin called the sadden appearance of a drastic trait in an individual a “monstrosity” and said they almost always fail to establish in the population. Thus, as we have seen above, Darwin gives a stepwise evolution of social behaviour. We must note, however, that even though Darwin gives the impression that the moral society whose evolution he is modeling is one that has progressed beyond that of “savage” groups, the essential features of his model are endemic to the most rudimentary social groups and have little relevance in more complex social organizations.

This brings us to the outstanding problem of social evolution today. The most widely accepted explanations of social evolution today, which date back to Darwin, are those based on kinship and reciprocity. The problem lies in the fact that these models work only for small communities in which most of the members are kin and/or there is repeated interaction between the same individuals. Such conditions, however, exist primarily in the societies Herbert Spencer (1967 [1886], 9) described as “small wandering hordes,” which of course are not the societies we are familiar with today. As Spencer (1967 [1886], 9) writes, “by integrations, direct and indirect, there have in course of time been produced social aggregates a million times in size the aggregates which alone existed in the remote past.” These dramatic increases in size are accompanied by correspondingly dramatic increases in complexity. We should therefore not expect the same principles and mechanisms to hold such drastically different types of society together.

As a matter of fact, in the social sciences, this distinction between small simple and large complex societies is prominently relevant in the explanations of most social phenomena, unlike in evolutionary biology. Political scientists and political philosophers identify two distinct conditions in human social experience: humans in the state of nature, i.e. the imagined state of society without government, and humans in political society. Some sociologists perceive a continuum of evolutionary stages between the two. Anthropologists and sociologists generally recognize three broad categories of human societies on the basis of culture, which tends more narrowly to be based on the sophistication of production technology. Thus we have *hunter gatherers*, *agricultural society* and *industrial society*. As we saw in Hume's analysis (chapter three), human societies vary according to size and complexity. The smaller, simple societies, often approximated to humans in the so-call state of nature, are held together by evolved instincts, particularly, empathy. Larger more complex (political) societies are maintained through a number of human artifices, notably justice and government. It is this insight to which I would like to call the evolutionary biologists' attention, so that it may help bring their abstract models closer to reality.

As we have seen above, the social models under evolutionary biology, if successful, only explain the establishment of the most primitive human social groups. Interestingly, the social scientists' explanation of social evolution begins with the primitive groups in which the account of the evolutionary biologist culminates. Their preoccupation is the transformation from "primitive" groups to more "civilized" groups. Lewis Henry Morgan (1877), for example, postulates social evolution to proceed from savagery through barbarism to civilization. Others have suggested that civilization proceeds from hunting to pottery to writing. The broad consensus among anthropologists is "that we moved from hunting, through agriculture to urban civilization," Fox notes (2000, xxx). Other equivalent characterizations of the stages are: "hunting and gathering," followed by "domestication" or "horticulture" and then "urban society" (Fox, 2000, xxvi). All these accounts point to the fact that today's nation-states, as well as the empires and kingdoms of antiquity, are very different from the primordial social groups that evolved by natural selection. Modern societies represent the further development of the primitive social groups. Therefore, a mechanism for their evolution needs to be sought. Is it a different mechanism, or is it the same natural selection applied to the primitive groups? The latter is what Boyd and Richerson (2009) propose, which, as we shall see in Hume's account, is highly unlikely.

For unlike the family and the tribe, which emanate from natural processes or by natural selection, the nation-state (like the kingdoms and empires of antiquity) is an artificial group formed by the amalgamation or assimilation of natural groups either through trade or conquest (see Boyd and Richerson 2009, 3285; Spencer (1967 [1886], 9). Experience suggests that two distinct human populations that occupy contiguous territories are more likely to be at war than at peace (Darwin 2012 [1871], 2766), unless their leaders work out some agreement for peace. But this peaceful arrangement will occur only if the two groups are closely matched in power, otherwise the stronger group is often inclined to overrun and absorb the weaker group (which is how nations and empires generally come about).

V. Levels of Sociality

For the purposes of explaining human sociality, contemporary evolutionary modeling has generally recognized only two kinds of social group: the family, which is generally small and consisting often of genetically related individuals, and then the wider society, which consists of both related and unrelated individuals. The binary classification ignores important differences between the two distinct types of society outside the family, which are critical for the evolutionary explanation of social behaviour. It is generally acknowledged that Hamilton's (1964) model, which has aptly been dubbed "kin selection," accounts for social cooperation among family members, but is inapplicable to non-kin cooperation. Trivers' (1971) reciprocal altruism model is widely reputed as setting the basis for explaining non-kin altruism but as Trivers (1971, 37) suggests, his theory is applicable only to non-kin altruism that fit a specified set of conditions: "(1) when there are many such altruistic situations in the lifetime of the altruists, [and] (2) when a given altruist repeatedly interacts with the same small set of individuals," He adds as a third condition that the costs and benefits to the two interacting individuals must be symmetrical. The first and second conditions are most relevant because they circumscribe the type of social group for which reciprocity is sustainable. It is what game-theoretic models refer to as an “iterated prisoner’s dilemma.” It is not the family, but it is also not the modern cosmopolitan society in which most members are strangers to each other. It is the primordial small hunter-gatherer or agrarian groups that were not entirely family members. For about ninety-nine percent of human evolutionary history, humans lived in those groups (Boehm 2012; Pinker 2006), some of which later transformed into today’s cosmopolitan political societies.

These two contrasting societies have been described variously as *organic versus mechanical* societies (Durkheim 1912), *simple versus compound* societies (Spencer (1967 [1886]), *status society versus contract society* (Maine 1861), *Gemeinschaft versus Gesellschaft*, commonly translated as *community versus society* (Tonnies, 1887).

Under contemporary evolutionary modeling, modern society matches what has been characterised as the *non-iterated prisoner’s dilemma* in *game-theoretic models*. They share the evolutionary modeling characteristic whereby the two potential interactants are strangers who will be interacting for the first and most probably the last time. Theoretical models predict that in such cases, “the only strategy that can be called a solution to the game is to defect always despite the seemingly paradoxical outcome that both players do worse than they could have had they cooperated” (Axelrod and Hamilton 1981, 1391). Thus, according to

Axelrod and Hamilton 1981, the breakdown or impossibility of cooperation is the solution both in game theory and biological evolution. In fact, Hume (T 3.2.7) discovers this very paradox when he considers our natural inclinations and their opposition to the "social virtues." Hume (T 3.2.7.3) observes that when humans fail to cooperate (as in the "strangers' dilemma" here), they "act in contradiction to their known interest," as they seem to prefer a trivial advantage to the maintenance of society. Thus, according to Axelrod and Hamilton (1981, 1392), in populations in which interactions between pairs of individuals are random and not repeated, "a mixture of heritable strategies evolves to a state where all individuals are defectors." Defection or selfishness as a strategy is therefore stable, as no mutant strategy, Axelrod and Hamilton assert, can do better in a population with such a strategy.

What this conclusion from evolutionary theory tells us is that it is impossible prima facie, for a stable cooperation to exist in a large group of genetically unrelated strangers, such as modern cities and nations. This is the kind of social group that would be characterized by the so-called "single shot prisoner's dilemma," as opposed to the one in which the groups are small, and the members, even though they may be unrelated, interact repeatedly and are familiar with each other's *cooperative attitude* or *reputation*.

From Hume's analysis of morality, we get a psychological explanation of how this happens. In small (often tribal) communities such potential actors have either met before and will certainly meet again or they know each other's reputation. These two factors are absent in the *two strangers* situation; and in such cases, Hume (*Treatise,* 3.2.7.3) notes that they make their choices on their understanding of general human nature. That is, each thinks the other has the same “propension” as he does “in favour of what is contiguous above what is remote.” It would be folly, each would think, “if I alone should impose on myself a severe restraint amidst the licentiousness of others” (*Treatise,* 3.2.7 3). This is the basis of the robustness of Hume’s account of human morality/sociality. It is deep, and rooted ultimately in the most fundamental human psychological dispositions.

The best overall strategy for the iterated prisoner’s dilemma is TIT FOR TAT (hence forth TFT), which, as Axelrod and Hamilton (1981) explain, “is simply one of cooperating on the first move and then doing whatever the other player did on the preceding move.” It is thus a strategy of cooperation based on reciprocity. Axelrod and Hamilton (1981) also demonstrated that once a TFT strategy becomes fixed, it is resistant to invasion by any other strategy as long as the system remains as “iterated Prisoner's Dilemma.”

However, the problem of modern evolutionary explanation is not limited to accounting for the evolution and maintenance of today’s cosmopolitan societies, which consist of unrelated strangers. Even with small communities, for which TFT may be a solution, there is still the problem of how a TFT strategy can invade and establish in an ALL D population,[1] which happens to be a stable strategy and which we generally assume in evolutionary modeling to be the pre-social condition. A TFT strategy works in a situation in which there are other TFT players, so that they can cooperate to their mutual benefit. If evolution by natural selection starts by a single mutation, how does that single TFT individual succeed when there is no other TFT FOR TAT to cooperate with? Axelrod and Hamilton (1981) propose two possible mechanisms. The first is that there can be kinship between mutant strategies, so that the mutual cooperation between genetically related individuals can help the TFT strategy gain a foothold. This looks like having a gene starting off discriminating on the basis of kinship and then at some point transforming into one that discriminates on the basis of reciprocity. Of course this is a fantasy as there is no credible basis for thinking such a thing could happen. The second possible mechanism for the invasion of an ALL D, by a TFT strategy is for a bunch of TFT strategists to arrive in a cluster. Once again, there is no credible basis for supposing a thing like that. In any case, such an explanation is just kicking the can down the road, for we would still need to explain how the TFT strategy evolved in the invading group. All this explanatory difficulty stems from the failure to entertain a non-Darwinian solution to the problem of maintaining a cosmopolitan society of non-genetically related

---

[1] An "ALL D" player defects all the time.

individuals, or even a failure to recognize the distinction between the two kinds of society.

This conflation of tribal and cosmopolitan societies is evident once again in Cohon's treatise on Hume. Cohon (2010, 2) identifies as one of the meta-ethical controversies "the dilemma of understanding the ethical life either as the 'ancients' do, in terms of virtues and vices of character, or as the 'moderns' do, primarily in terms of principles of duty or natural law." She observes that Hume explicitly favors an ethic of character along "ancient" lines, but that he also insists on a role for rules of duty in what he calls the artificial virtues. As I see it, however, the artificial virtues pertain to political societies, while the "natural virtues" have to do with tribal societies. As Cohon (2010) comments, the artificial virtues make "*im*personal cooperation" possible, whereas the natural sentiments are too partial to support cooperation between strangers. In fact, in Cohon's (2010, 25) commentary, Hume describes people who "cooperated only within small familial groups" as "people who belonged to no society." Hume (*Treatise* 3.3.1) also discusses the distinction between the natural and artificial virtues, and suggests that unlike the natural virtues, the artificial virtues are not inherent in us, but are acquired through education and experience.

VI. Hume's 6-Step Model of Social Evolution

Hume's account of social/moral evolution may be explained in the following six steps:

Step One. Parenting

Hume roots social evolution ultimately in parenting. He suggests that the attraction between the sexes brings individuals together. Of course, the attraction between the sexes does not necessarily bring about social cooperation, but it presents an opportunity for it. In fact in very few species does the attraction between the sexes lead to any lasting cooperative bond beyond copulation. The attraction between the sexes exists in nearly every animal species, yet not all have evolved sociality. This suggest that sexual attraction alone is not sufficient for sociality and that some additional factors are required. Nevertheless, individual organisms cannot cooperate unless something brings them together in the first place. The attraction between the sexes definitely performs that function. Hume suggests that the attraction between the sexes results in common offspring, which further strengthens the bond of love between parents and between parents and offspring. Again, we have to point out that invariably, living organisms produce offspring, but only in the minority of cases has that led to the evolution of sociality. The rudiment of sociality is parenting, which many species have not evolved.

In fact, there are hardly any cases in which animals establish durable mating pairs that do not involve parenting. The root of sociality, therefore, more plausibly lies in parenting, which started off with predominantly females as solitary parents. Parenting is already quasi-social, the difference being that it is temporary and benefits go almost exclusively from parent to offspring. The paradox of altruism does not arise here, or we can say that the sustainability of altruism at this level is explained by kin selection.

Step Two: The Family Group

Out of parenting might develop love/empathy between the parents and between the parents and their children. Note that at this stage, the love or empathy exists only between the family members. So Hume calls it “limited generosity” or “partiality.” From this mutual love and cooperation within the family, the members will experience, learn and appreciated the benefits of fellowship. This is the stage in the evolution of sociality that Hamilton’s (1964) *inclusive fitness hypothesis* may explain, if we grant that it is successful. Hamilton’s hypothesis suggests that altruism may evolve if the benefits of the altruistic behaviour fall predominantly on genetic relatives of the altruist.

Step Three: Socializing Outside the Family

Having learned the benefits of cooperation, such individuals may then try to extend the social cooperation to neighbours outside the family. Hume recognizes that “sympathy” is at the root of the group instinct or social behaviour. But he also notices that it is naturally confined to the family, and requires special circumstances to extend to individuals outside the family. The factors that facilitate this extension of sympathy to non-kin are a class of virtues Hume calls the “social virtues,” and says they are artificial, in the sense that they are not natural instincts or inherent in us. They are cultivated or learned. Indeed, Darwin’s model is broadly consistent with Hume’s on this point. He used natural selection to explain all kinds of physical traits. However, when he was confronted with the altruism associated with the neuter condition in eusocial insects he altered natural selection to fit the circumstance by suggesting family (group) selection instead. Similarly, he appealed to group selection when he faced human sociality/morality. In this case however, group selection was not sufficient because altruism in these situations frequently extends beyond genetic relatives. He therefore proposed this extra-familial altruism not to be an innate instinct, but instead a product of certain more basic instincts and factors, such as empathy, reciprocity, group affiliation and social approbation and disapprobation.

This is where Hamilton's theory loses its efficacy, as the interacting individuals will be non-kin. Instead, Trivers' (1971) *reciprocal altruism hypothesis* comes into play here, together with *game theoretic models*, which derive from the reciprocal altruism hypothesis. Since at this stage, the social interaction will mostly be between a small cluster of individuals, the *iterated prisoner's dilemma* will be applicable and efficacious.

Step Four:

Of course, as we have seen above, these initial social communities are not yet the cosmopolitan societies we are familiar with today. They would be the pre-industrial primitive societies that anthropologists describe as *hunter-gatherer* and *agricultural*. The difficulty for evolutionary biologists is making their models for social behaviour robust enough to encompass social relationships in today's mega-cosmopolitan societies, in which cooperators are predominantly strangers.

This is what has been modeled as the *non-iterated prisoners' dilemma* for which there is no solution under the two dominant theories of modern evolutionary biology — inclusive fitness and reciprocal altruism. It is these that Fehr and Fischbacher (2003, 785) have described as "current gene-based evolutionary theories" and suggest they "cannot explain important patterns of human altruism."

Hume notes that generosity towards others "proceeds merely from empathy," which by nature, he explains, is confined to the family. In fact, as Darwin (1871) points out, even "species which are not social, such as [lions?] and tigers, no doubt feel sympathy for the suffering of their own young, but not for that of any other animal" (Kindle Locations 2035-2036). That kind of empathic behaviour is akin to what happens in the eusocial insects, and is readily accounted for by kin selection (Hamilton 1964) or multilevel selection (Wilson and Wilson 2007). As Boehm (2012, 11) puts it, the real puzzle of social behaviour is "the genetically 'reckless' generosity of humans whereby generosity extends beyond nepotism to non-kin."

Hume's analysis of this issue is most persuasive because he identifies the root psychological causes of the problem, and shows how they are subdued in order to allow for such societies to endure. But problems arise due to the following forces: (i) the inherent selfishness of humans; (ii) the partiality or limited generosity of humans; (iii) and the fact that goods are forever limited in availability and are easily transferable. In other words, there is no natural barrier to one person's property becoming another person's.

To elaborate on (ii), humans are by nature selfish; but also have, as part of their natural constitution, empathy, which puts certain other individuals within each person's sphere of self-interest. Empathy, as the Scottish philosopher Alexander Bain (1864, 481) explains, "identifies us with the pleasures and pains of [other individuals], and supplies the motive to work for those to some extent as if they were our own." Thus, according to the principle of empathy, our selfishness is in many cases, broader than strict love for the self alone. Empathy extends self-love to include family and friends. Sometimes this love for others can even be greater than the love for the self. As Hume (*Treatise*, 3.2.2.5) observes: "there are few that do not bestow the largest part of their fortunes on the pleasures of their wives, and the education of their children, reserving the smallest portion for their own proper use and entertainment."

We also observe that empathy by nature is restricted to family and other individuals with whom we are associated, such as relatives and friends. It is this partiality of empathy that makes it more an impediment to the formation of societies than selfishness is. As Hume (*Treatise*, 3.2.2.6) notes, "though this generosity must be acknowledged to the honour of human nature, we may at the same time remark, that so noble an affection, instead of fitting men for large societies, is almost as contrary to them, as the most narrow selfishness."

More critically, Hume notes, there is no remedy in "uncultivated nature" for this web of opposition to sociality. So an invention has to be sought. And that invention is justice; as Hume (*Treatise*, 3.2.2.18) writes: "it is only from the selfishness and confined generosity of men, along with the scanty provision nature has made for his wants, that justice derives its origin.” This suggests that scholars modeling social evolution need to move away from the vain search for an evolved human trait that remedies the problem. Modern society is maintained by artifice rather than instinct. The failure to realize this has been the bane of the evolutionary biologist.

Step Five:

So we need justice to protect ownership. Natural social groups are those that are held together by our natural constitution alone, or solely by our natural passions. Thus, they are amenable to explanation by natural selection as adaptive evolution. Artificial social groups are those that are put together by human convention. The cooperation between members in the former is driven directly by natural passion and the principle of empathy, whereas the cooperation between individuals in the latter is maintained by some external superior force. This is the distinction Hume has emphasized – natural groups and artificial groups, which require different explanatory models.

Hume (*Treatise*, 3.2.1.17) takes justice as one example of a social virtue and argues that it “is not derived from nature, but arises artificially, though necessarily from education, and human conventions.” This is because, he explains, “we have naturally no real or universal motive for observing the laws of equity.

Step Six:

But we fail to behave justly because we naturally prefer the present or near good to the future and distant good. Thus, even though we know that justice if followed will be good for all members of the society, our nature does not recommend it to us most of the time.

Therefore we need government and magistrates especially when possessions increase considerably. This point of Hume's is more or less a better explained and a better applied version of contemporary evolutionary explanations than that involving "policing" or "sanctions" against violators of norms. Hume recognizes that although justice is a great and important duty, it is a distant interest, and consequently we are often seduced away from it by the allurement of the present, which is often a rather frivolous temptation. It is a great weakness, according to Hume, that is incurable in human nature. In his treatise *On Government* (hence forth *OG*), Hume (OG, 16088) suggests that due to the frailty of human nature, "it is impossible to keep men faithfully and unerringly in the paths of justice." Consequently, men "establish political society in order to administer justice, without which there can be no peace among them, nor safety, nor mutual intercourse" (OG, 16089). (This is elaborated more in *Treatise* 3.2.7). Here, Hume recognizes that we cannot hope for or rely on men acquiring the moral virtues through experience and education and becoming upright moral citizens and making society possible.

VII. The Insight for the Modern Darwinian

The beauty of Hume's stepwise approach to human social evolution is that it clearly identifies what in human nature or human ingenuity is the basis of the bond between the individuals at the two distinct types of social groups — primitive groups and political societies. Hume's analysis of social and moral evolution is largely descriptive and some of his key claims have been confirmed by modern empirical neuropsychology. So what are modern evolutionary biologists to do with this insight? They may choose to ignore it, but they will have no empirical grounds to reject it. On the other hand, supposing they accept it, how should the approach to evolutionary modeling change?

Hume does not lay out a mechanism for social evolution; rather, he provides the principles and conditions that allow for the evolution and maintenance of society. He suggests that there is a natural instinct for the individual to love and sacrifice for the family. This together with self-interest are the two principles of human nature that make social behaviour possible. Thus, the family as a cohesive social unit, for Hume, is a given. What he thought needed explanation is the existence of "political society," which these two principles of human nature oppose.

Surprisingly, Hume has some modern Darwinians as allies in his view that cooperation within families requires no special explanation. Incidentally, a theory of kin based altruism is what made W. D. Hamilton the most famous personality in modern evolutionary biology. However, many scholars after Hamilton, including evolutionary biologist Robert Trivers of reciprocal altruism fame, and particularly social scientists (Boehm 2012; Fehr and Fischbacher 2003) working on that subject, define altruism to exclude behaviour that benefits genetic relatives.

On the basis of Hume's insight, therefore, social evolutionary explanation ought to be about extra-familial social groups, thus rendering Hamilton's kin based inclusive theory irrelevant. Another way contemporary evolutionary modeling of social evolution must change, if we were to adopt Hume's description of social behaviour as a basis, is that we could no longer construe the entirety of social behaviour simply as altruism that is directly implanted in human nature by means of natural selection. Hume (*Treatise*, 3.2.1.12) asserts that there is no natural instinct for kindness "to men independent of their merit and every other circumstance." In other words, altruism is not an evolved trait, but rather something that is generated by circumstances. That is why he says that "there is no such passion in human minds as 'the love for mankind'." Hume wants us to see how problematic the supposition of a natural love for all other persons might be by asking the following: "what if he be my enemy and has given me just cause to hate him? What if he be a vicious man and deserves the hatred of all mankind? What if I am in necessity and have urgent motives to acquire something for my family?" And so on. These are circumstances under which it would make no sense, and would in fact be contrary to the principles of justice, to extend kindness or altruism to another person. Loving or hating is situational, i.e. provoked by circumstances.

In contemporary evolutionary modeling of sociality, we imagine that a genetic mutation confers upon an individual an innate altruistic trait, and then we propound hypotheses as to how such a trait might spread within the population by supplanting a supposedly instinctive selfishness. However, research data from social anthropology suggest that most social behaviour is learned and infant social animals that are denied contact with the social group are unable to interact socially when subsequently introduce to the social group (Jolly 1966). Instead of discussing altruism as a thing in our nature, Hume, like Darwin, suggests that such behaviours are generated by other factors, such as empathy, reciprocity, selfishness and the approbation of our fellows. The inclusion of selfishness on this list and, atop that, the complicity of Darwin should confound the modern Darwinian.

Is the modern evolutionary biologist prepared to square off against Hume and Darwin on this claim? The data coming from experiments in game theory, evolutionary anthropology and behavioural ecology are heavily in favour of Hume and Darwin. Fehr and Fischbacher (2003) discuss a number of them in a review. The studies suggest, generally, that in non-kin interactions, individuals act altruistically in order to avoid sanctions or retaliation. They observe that as a result of this and “depending on the environment, a minority of altruists can force a majority of selfish individuals to cooperate” (785). Also, recall Boyd and Richerson’s (2009, 3283) observation that “if cheaters are despised by others in their group, and, as a consequence, suffer social costs — lose status, mating opportunities, the benefits of mutual aid when ill or injured — then they may be motivated to cooperate, even though prosocial motivations are entirely absent from their psychology.” This makes self-interest the underlying motive for those instances of altruistic behaviour. If that is the case, then a seismic shift is required in the contemporary Darwinian explanation of social evolution. If selfishness is indeed one of the underlying causes of altruism rather than a force opposing it, then explaining social evolution by modeling how an altruistic trait might supplant selfishness is deeply flawed.

So what are modern Darwinians to do under these circumstances? They may concede that social behaviour does not fit the fundamental modeling assumptions of modern Darwinism, which would mean the end of sociobiology, the modern Darwinian discipline dedicated to explaining the genetic evolution of social behaviour. Alternatively, they could try to model the genetic evolution of the more fundamental principles of human nature that underlie altruism and social behaviour. The problem with the latter response would be how to determine what contrasting traits are to be supplanted by each of the fundamental principles — selfishness, empathy, rationality, imitation, concern for the approbation of others etc.

Gene-culture co-evolution has been proposed as a strategy to explain the evolution of large human societies of unrelated individuals (Fehr and Fischbacher 2003; Boyd and Richerson 2009). It is based on multi-level selection that combines kin selection, reciprocity and group selection. The basic claim of this strategy is that once small groups are formed through kin selection and reciprocal altruism, group selection then ensues among the small groups, whereby the groups with superior culture take over groups with inferior culture, thereby becoming large societies.

There are a number of problems with this model. In the first place, a superior culture may spread among several groups or societies, but that does not make them all a single big cosmopolitan society. “Western culture,” for example, is rapidly spreading around the world, but that does not make all the communities in which it is established a single social group in the sense of the discussion of social evolution. Secondly, a mere overrun of several weaker groups by a superior group does not guarantee a stable mega-society until civic institutions are set up to administer justice and mediate relations among individuals. Such institutions or human artifice, as Hume observes, are the reason for the existence of large human societies. What gives one group the edge to conquer another is really not a social puzzle.

VIII. Other Virtues of Hume’s Account

This insight of Hume’s also provides a resolution to the protracted debate as to whether humans are by nature good and are subsequently corrupted by social circumstances, or whether our natural tendency is towards evil, which then has to be corrected by socialization and education. Pinker (2006, 1) lays out three competing views of human nature in contemporary thought. The first, associated with Locke, is the idea that humans are born with an empty mind (blank slate) to be filled with life's experiences, good or bad. The second, associated with Rousseau, is the so-called noble savage idea, which suggests that humans are good natured and peaceful in their primitive uncivilized state. Then the third point of view, associated with Hobbes, is that humans in their natural state are war-like and brutish and require some social force to keep them in check. Pinker (2006, 4) concludes that, "when it comes to life in a state of nature, Hobbes was right; Rousseau was wrong."

Hume's insight shows that human nature does not fall squarely into any of these three competing viewpoints. Humans according to Hume are neither naturally vicious nor naturally virtuous. Each individual has the capacity to express either behaviour depending on the circumstances. The first option too is out because, according to Hume, there exist some natural passions such as love and hatred that are excited by circumstances to produce behaviour that is either virtuous or vicious. In other words, people are not emotional blank slates. According to Hume (*Treatise,* 3.2.1.12) there may be variations in temper between individuals, but in the main, "man in general, or human nature, is nothing but the object both of love and hatred, and requires some other cause [that] may excite these passions."

Hume (*Treatise*, 2.1.3.4) points out that "in all nations and ages, the same objects still give rise to pride and humility," for example. He asks whether "we can imagine it possible, that while human nature remains the same, men could ever become entirely indifferent to their power, riches, beauty and personal merit"

(*Treatise*, 2.1.3.4). In other words, people's circumstances determine their behaviour, not some innate trait. This view gives pre-eminence to "the situational determinants of behavior," which has gained currency in contemporary social psychology under the term "Situationism" (Bowers 1973; Allport 1966; Harman 1999 etc.). At the same time, Bowers (1973, 307) reports, "the influence of trait theory as a viable model of man has dwindled." Trait theory assumes distinctive behaviour to be due to an agent's distinctive character traits. It is a presumption many psychologists now view as erroneous and refer to it as the "Fundamental Attribution Error" (Ross 1977; Harman 1999). The ascendency of the situationist view is spurred by a number of famous experiments in social psychology, including Milgram (1963); Zimbardo (1970) and Darley and Batson (1973).

Some scholars argue, according to Rousseau, that since man in the state of nature did not live in societies, there could be no kind of moral relations or duties between men in that state. Consequently, "they could not be either good or bad, and had neither vices nor virtues" (DOI, 28). Rousseau tries to turn this argument in favour of his "noble savage" view that contradicts Hobbes' bellicose savage view. He argues that "the state of nature, being that where the care for our own preservation interferes least with the preservation of others, was the most favourable to peace and most suitable to mankind" (DOI, 29). This is another example where Hume perspective stands above the fray on a contentious issue among philosophers. On this issue, Rousseau accepts the premise of those who argue that man could be neither good nor evil in his original savage state because the opportunity did not exist for either, as there were no social interactions between individuals. Hume, on the contrary, believed savage man lived in social groups, only that such groups were not political societies. For Hume, human nature, whether in savages or civilized people, has the capacity to produce both good and evil.

In Cohon's (2010) rendition of Hume, we are actually both selfish and humane. We are actually greedy, but also have "limited generosity" — dispositions to kindness and liberality which are more powerfully directed toward kin and friends and less aroused by strangers. If this is indeed the nature of humans, then modeling human social/moral nature as a single heritable trait that some individuals in society have and others do not, as modern Darwinians do, is inconsistent with human nature.

Hume does not share Hobbes' pessimism that there will be "all against all" brutality among people in the absence of organized society. According to Cohon (2010), Hume suggests there would still be caring and cooperation between friends, and of course, families, due to the inherent empathy between such individuals. The inherent partiality of empathy, however, would make wider cooperation impossible. So there will always be social cooperation of some sort within human populations. That is why the so-called savage societies, as we see with Durkheim's *totemic clans*, were not solitary humans. They lived in tribal groups. Darwin (2012 [1871], 2319) also describes such savages as living in social groups to which they are deeply loyal.

This again gives us an insight into the level of socialization to which Hume is applying his moral principles. Hume's conception of 'society' does not include the tribe, or the early hunter-gatherer or agrarian groups, in which nearly every member knows every other member. Even though he recognizes that empathy and cooperation existed in such groups, they were not societies. By "society," in his analysis, Hume has in mind something like the cosmopolitan eighteenth century British society in which he lived, and which I have been calling political society. For Hume, without government, such simple groups are not political societies.

We know, according to Henrich et al (2006, 1286), that "hunter-gatherers engaged in extensive cooperation within residential bands." Also, as we saw from Darwin earlier, members of those primitive groups were deeply loyal to their groups. All these considerations point to what most scholars today will consider a social group. For Hume however, these are not societies as long as the social relations are driven by innate instinct rather than artifice. In this view, Hume has an ally in Ferdinand Tonnies (1887) who distinguishes between *Gemeinschaft* (community) and *Gesellschaft* (society). In a treatise dealing solely with that subject, Tonnies presents detailed descriptions of the two contrasting human social groups.

In a *Gemeinschaft*, according to Tonnies, individuals have simple and direct inter-personal relations with each other that are driven by what he calls *Wesenwille (natural will), which is characterized by* natural and spontaneous emotions and expressions of sentiments. On the other hand, the *Gesellschaft*, which is typified by modern, cosmopolitan societies is an invention of *Kürwille (rational will).* In the *Gesellschaft*, human relations tend to be impersonal and indirect. These distinctions are similar to Hume's distinctions between groups that are maintained by the natural virtues and certain principles of human nature on the one hand, and societies that are maintained by the artificial virtues and other human artifices on the other. Also, similar to Hume, Tonnies suggests that members of a community (Gemeinschaft) are bound together by ties of kinship, fellowship and custom, whereas in a society (Gesellschaft), individuals are free-standing and their interactions are regulated largely by "the external constraints of formally enacted laws."

This distinction became pronounced following the industrialization of Europe that began in the eighteenth century and the concomitant rapid urbanization. As a consequence, many scholars tend to talk of the distinction as that between pre-industrial and post-industrial societies. However, I think that is just one more difference between the two types of society. The key point pertinent to the explanation of human sociality, however, is that what Hume calls (political) society or what Tonnies calls *Gesellschaft* is a purely human invention that will always present a problem for theories that seek some inherent qualities of human nature to account for human social cooperation generally.

IX. Summary

In this final chapter, I discussed three different approaches to explaining social/moral behaviour. They are the approach of Darwin, the modern evolutionary biologist and the philosopher David Hume. The approach of the modern evolutionary biologists, which is based on abstract mathematical modeling, relies fundamentally on certain simplifying assumptions that are plainly inconsistent with the empirical data on human behaviour from social psychology and behavioural ecology. We see that the other two accounts, Darwin's and Hume's, are descriptive and naturalistic and ultimately more consistent with the field data on social behaviour.

However, of the three, it is only Hume's account that truly explains contemporary modern society. In Darwin's (1871) and Boyd and Richerson's (2009) group selection accounts, groups that are better organized take over other groups. We suppose that when that happens they become bigger groups, but the changes that need to take place in order to maintain harmony in the new, bigger groups are not addressed in those models. Those models assume that the specific "cultural" innovation that gives one group the military superiority to gobble up other groups will also provide the administrative mechanism that will sustain the bigger society. The gist of Hume's social theory is that a modern mega-cosmopolitan political society is sustained by the artifices of justice and government, rather than by any direct innate human instinct. This appears to be the most satisfactory account of the maintenance of modern cosmopolitan society. No other model is nearly as satisfactory.

Some Key Points of the Thesis

It is widely acknowledged that social behaviour remains a puzzle under the modern Darwinian evolutionary paradigm. This thesis provided an in-depth analysis of the nature and basis of the puzzle and proposed how the Darwinian evolutionary philosophy and modeling assumptions will have to change to resolve the difficulty. The most fundamental principle of Darwinian evolutionary explanation is that evolutionary change occurs by one inherent trait replacing another in an organism. This thesis suggests that this fundamental principle of Darwinian evolutionary explanation may not be applicable to social behaviour, since the evolution of social behaviour, at least in humans, is not marked or enabled by the replacement of one innate character by another.

The second major argument in the thesis is that the modern Darwinian reduction of social behaviour to altruism — defined as other-regarding behaviour — is problematic. Even within the Darwinian camp, some evolutionary models, in game theory particularly, do not cast altruists exclusively as benefiting others. Altruists harm some individuals when they sanction or punish defectors. The thesis also criticized the modeling assumption known as the phenotypic gambit, under which sociality, as a single-locus trait called altruism, competes with selfishness for the same gene locus. Even the staunchest proponents of the phenotypic gambit admit that it is most likely false.

Another point argued in the thesis is that the most widely accepted theories under the modern Darwinian paradigm — kin selection and reciprocal altruism — may only explain sociality in small primordial human groups such as hunter-gatherer societies in which all group members are either relatives or acquaintances. However, the human societies we know today are mostly large and consist mostly of strangers. They are non-Darwinian and are maintained by human artifice. Under current Darwinian explanation, a hypothesis known as gene-culture co-evolution has been proposed to deal with this problem of large human societies. That theory suggests group selection on the basis of cultural differences between groups as the mode of evolution and maintenance of the mega-societies. Such an explanation, I think, is still unsatisfactory. I argue that a superior culture may spread widely, but it is not necessarily what forms or maintains such societies.

I dedicated the first chapter to giving a context to the Darwinian problem of social evolution. Using the works of Karl Popper and Thomas Kuhn, I argued that puzzles are frequent in scientific theories as well as paradigms. Using the puzzle of retrograde planetary motion in the Aristotelian cosmological paradigm, I argued that scientific puzzles tend to endure when a deeply held fundamental assumption is erroneous. I then showed that the puzzle of social evolution is enduring because a number of fundamental modeling assumptions for social behaviour, which I have mentioned above, are erroneous.

The thesis identified some insights from Hume's analysis of human morality and sociality that suggest how those Darwinian modeling errors may be corrected. We see from Hume's analysis that nothing inherent in human nature needs to change in order to move beyond parenting to sociality. Perhaps we should except rationality, arising, as it does from the complexification of the brain, but certainly nothing that directly codes for sociality.

I discussed Hume's account of the evolution of society in six steps. In a nutshell, Hume identifies two principles in human nature — selfishness and empathy — that are the ultimate basis of human sociality. Those two principles suffice to form small, primitive human societies. However, they oppose the formation of large, cosmopolitan societies and there is nothing innate in human nature that could remedy this difficulty. So such large societies are formed and maintained only through human inventions such as governments and justice. Thus, in Hume's analysis, altruism (which arises from empathy) works with selfishness to form and maintain small social groups and to oppose the formation of large social groups. We should note that the partiality of empathy is key to the cogency of Hume's social analysis. The modern Darwinian view, on the contrary, has selfishness solely opposing sociality (particularly altruism) at all levels. Thus there is a clash in fundamental philosophies between Hume and the modern Darwinian. As I suggest in the thesis, the empirical research clearly favours Hume. So what is the modern Darwinian to do?

It appears to me that it may be easier to give up on efforts to extend Darwinian evolutionary explanation to social behaviour than to renounce the fundamental Darwinian philosophy from which this explanatory approach to social evolution arises.

Bibliography

Agrawal, A. A. (2001) Phenotypic Plasticity in the Interactions and Evolution of Species. Science 294: 321-326.

Alcock, J. (2001). The triumph of sociobiology. Oxford University Press.

Alexander, J. M. (2009) Evolutionary Game Theory. The Stanford Encyclopedia of Philosophy (Fall 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/game-evolutionary/>.

Alexander, R. D. (1974) The Evolution of Social Behavior Annual Review of Ecology and Systematics, 5, 325-383

Allen, N.B. and Badcock, P.B. (2006) Genes for Susceptibility to Mental Disorder are not Mental Disorder: Clarifying the Target of Evolutionary Analysis and the role of the Environment. Behavioural and Brain Sciences. 29 (3): 405-406.

Allport, G. W. (1966) Traits Revisited. American Psychologist, 21, 1-10.

Asma, S. T. (2012) Against Fairness. University of Chicago Press Chicago and London.

Axelrod, R. and Hamilton, W.D. (1981) Evolution of Cooperation. Science 211, 1390 1397.

Bain, A. (1864) The Emotions and the Will. Longmans Green and co. (1865).

Bernhard, H.; Fischbacher, U and Fehr, E. (2006) Parochial altruism in humans. *Nature.* Vol 442 (24) 912-915.

Bloom, P. (2010) How do morals change? Nature, Vol 464. p. 490.

Boehm, C. (2012) Moral Origins. Basic Books, New York,

Boehm, C. (2000) Conflict and the evolution of social control. Journal of Consciousness Studies, 7, No. 1–2, pp. 79–101.

Bourke, A. F. G. (2014) Hamilton's rule and the causes of social evolution. Phil. Trans. R. Soc. B 369: 2-10. http://dx.doi.org/10.1098/rstb.2013.0362

.Bowers, K. S. (1973) Situationism in Psychology: An Analysis and a Critique. Psychological Review, Vol 80 (5), 307-336.

Bowles, S. (2006) Group Selection, Reproductive Levelling, and the Evolution of Human Altruism. Science 314: 1569-1572.

Boyd, R. and Richerson, P. J. (2009) Culture and the evolution of human cooperation. Philosophical Transactions of the Royal Society B (2009) 364, 3281–3288.

Bshary, R. and Bergmuller, R. (2007) Distinguishing four fundamental approaches to the evolution of helping. Journal of Evolutionary Biology 21, 405–420.

Buss. D. (2004) Evolutionary psychology: The new science of the mind (2nd ed.). Boston: Allyn & Bacon

Buss, D. M. (2009). How Can Evolutionary Psychology Successfully Explain Personality and Individual Differences? Perspectives on Psychological Science 4 (4): 359. Doi: 10.1111/j.1745-6924.2009.01138.x.

Campbell, D. (1978). On the Genetics of Altruism and the Counter hedonic

Components of Human Nature. In L. Wispe (Ed.). Altruism, sympathy, and helping (pp. 39–58). New York: Academic Press.

Castro L., Serrano J., and Toro M. (1998) Conceptual Capacity to Categorize and the Evolution of Altruism. Journal of Theoretical Biology. 192(4) 561-565.

Carroll, S. B. (2005) Endless Forms Most Beautiful (Norton, New York).

Cavalli-Sforza, L. L., and Feldman, M. W. (1978). Darwinian selection and "altruism." Theoretical Population Biology 14: 263-281.

Cohon, R. (2010) Hume's Moral Philosophy, The Stanford Encyclopedia of Philosophy (Fall 2010 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2010/entries/hume-moral/>.

Copernicus, N. (1543) On the Revolutions if the Heavenly Spheres. *Running Press Book Publishers*, Philadelphia.

Copernicus, N. (1543) De Revolutionibus http://www.webexhibits.org/calendars/year- ext-Copernicus.html

Copernicus, N. Three Copernican treatises: the Commentariolus of Copernicus, the Letter against Werner, the Narratio prima of Rheticus. *Edward Rosen Transl.) Octagon Books (1971)*.

Darley, J. M. and Batson, C. D. (1973) From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior, *Journal of Personality and Social Psychology*. Vol. 27, No. 1, 100-108.

Darwin, C. (1871) The Origin of Species and the Descent of Man. *Modern*

Library, New York.

Darwin, C. (1859) The Origin of Species. *Modern Library, New York* (1998).

Dawkins, R. (2004) Extended Phenotype – But Not *Too* Extended. A Reply to Laland, Turner and Jablonka. *Biology and Philosophy* 19: 377–396.

Dawkins, R. (1982) Replicators and Vehicles. In: Currents Problems in Sociobiology. Eds. King's College Sociobiology Group. *Cambridge University Press*.

Dawkins, R. (1976), The Selfish Gene. *Oxford University Press* (1989).

Depew, D. J. (2013) The Rhetoric of Evolutionary Theory. *Biological Theory* 7:380–389

Doebeli, M., Hauert, C. Killingback, T. (2004) The Evolutionary Origin of Cooperators and Defectors. *Science* Vol. 306 no. 5697 pp. 859-862 DOI: 10.1126/science.1101456

Doris, J., Stich, S. (2012) Moral Psychology: Empirical Approaches. *Stanford Encyclopedia of Philosophy* (winter 2012 Edition).

Durkheim E. The Elementary Forms of Religious Life. *The Free Press (1995).*

Durkheim E. (1902) The Division of Labor in Society. (Translated by George Simpson). New York: *The Free Press, 1947*.

Durkheim E. (1902) The Division of Labor in Society. (Translated by W. D. Halls). New York: *The Free Press, 1984*.

Fayerabend, P. (1975) Against Method. *Verso, London, UK* (2008).

Fehr, E. & Fischbacher, U. (2003) The nature of human altruism. . *Nature* **425**, 785-791 (23 October 2003).

Fisher, R. A. (1930) The Genetical Theory of Natural Selection. *Dover Publications (1958)*.

Fisher, R. A. (1918) The Correlation between Relatives on the Supposition of Mendelian Inheritance" *Philosophical Transactions of the Royal Society of Edinburgh*. 52, 399–433.

Fletcher, J.A. & Doebeli, M. (2006) How altruism evolves: assortment and synergy. *Journal Evolutionary Biology*. 19: 1389–1393.

Fletcher, J.A. & Doebeli, M. (2009) A simple and general explanation for the evolution of altruism. *Proceedings of the Royal Society, London. (B)* 276: 13–19.

Fletcher, J.A. & Zwick, M. (2006) Unifying the theories of inclusive fitness and reciprocal altruism. *American Naturalist* 168: 252–262.

Forster, K. R. (2009) A Defense of Sociobiology. *Cold Spring Harbor Symposia on Quantitative Biology* 2009 74: 403-418 originally published online April 7, 2010.

Foster, K. R., Shaulsky, G., Strassmann, J. E., Queller, D., C., Thompson, C. R. L., (2004) Pleiotropy as a Mechanism to Stabilize Cooperation. *Nature*. Vol. 431, 693-696. Oct. 2004.

Freud, S. (1913) Totem and Taboo. *Trans*. James Strachey. *W. W. Norton and Company, New York*. Futuyma D (2006) Evolutionary Biology. Sunderland, MA: Sinauer.

Gardner, A., West, S.A. & Barton, N.H. 2007. The relation between multilocus population genetics and social evolution theory. *American Naturalist*. 169: 207-226.

Gardner, A, West, S. A., Wild, G (2011) The genetical theory of kin selection. *Journal Evolutionary Biology* 24:1020–1043.

Gert, B. (2004) Common Morality: Deciding What to Do. Oxford Scholarship Online: February 2006.

Gill, M (2000) Hume’s Progressive View of Human Nature. *Hume Studies.* Vol XXVI, (1) 87-108.

Gintis, H., Bowles, S., Boyd, R., and Fehr, E. (2003). Explaining altruistic behavior in humans, *Evolution and Human Behavior, 24*, 153-172.

Godfrey-Smith, Peter (2009) Darwinian Populations and Natural Selection (*Oxford University Press*).

Gould, S. J. (2002) The Structure of Evolutionary Theory. *Belknap – Harvard* (2002).

Gould, S. J. (1996) The Mismeasure of Man. *W. W. Norton Company*, New York.

Grafen, A. (2006) Optimization of inclusive fitness. *Journal of Theoretical Biology* 238 (2006) 541–563

Grafen, A. (2007a) Detecting Kin Selection at Work Using Inclusive Fitness. *Proceedings: Biological Sciences,* Vol. 274, No. 1610 (Mar. 7, 2007), pp. 713-719.

Grafen, A. (2007b) An inclusive fitness analysis of altruism on a cyclical network. *Journal of Evolutionary Biology*. 20. 2278–2283

Grafen A (1984) Natural selection, kin selection and group selection. In: Krebs J.. R, Davies N. B (eds) Behavioural ecology: an evolutionary approach, 2nd ed. *Blackwell, Oxford*, pp 62–84.

Grotius, Hugo (2013). The Rights of War and Peace: including the Law of Nature and of Nations (annotated & illustrated) *Professional Etext version Professional Kindle Editions*.

Grotius, Hugo (1901) The Rights Of War And Peace Including The Law Of Nature And Of Nations. A. C. Campbell (Trans). *M. Walter Dunne, Publisher, Washington & London*

Haldane, J. B. S. (1932) The Causes of Evolution. *Longmans, Green and Co. London. New York. Toronto*.

Haidt, J. (2007) The New Synthesis in Moral Psychology. *Science* **316**, 998-1002.

Halliburton R (2004) Introduction to population genetics. *Pearson Prentice Hall, Upper Saddle River*.

Hames, R. (2001) Human Behavioral Ecology. *International Encyclopedia of the Social & Behavioral Sciences*.

Hamilton, W.D. (1970) Selfish and spiteful behaviour in an evolutionary model. *Nature* 228: 1218–1220.

Harman, G. (1999) Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society, New Series*, Vol. 99 (315-331).

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. and Ziker, J. (2006) Costly Punishment across Human Societies. *Science, New Series*, Vol. 312, No. 5781 (1767-1770).

Hewstone, M., Rubin, M and Willis, H (2002) *Inter-Group Bias.* Annual Review of Psychology 53, 575-604.

Hill, K. (2002) Altruistic Cooperation During Foraging By The Ache, And The Evolved Human Predisposition To Cooperate. *Human Nature* 13 (1), 105–128.

Hobbes, Thomas (2013). The Collected Political Works. e-artnow. Kindle Edition.

Hobbes, Thomas (1651) Leviathan*. Dutton, London: New York: Dent*; 1973.

Hume, David (1740) A Treaties on Human Nature. *Oxford University Press* (1978).

Hume, David (2013) On Government. In: The David Hume Collection: 17 Classic Works. *Waxkeep Publishing. Kindle Edition*.

Jablonka, E. (2004) From Replicators to Heritably Varying Phenotypic Traits: The Extended Phenotype Revisited. *Biology and Philosophy* 19: 353–375.

Johns, P. M., K. J. Howard, N. L. Breisch, A. Rivera, and B. L. Thorne (2009) Nonrelatives inherit colony resources in a primitive termite. *Proceedings of the National Academy of Science* vol. 106, no 41.

Jolly, A. (1966) Lemur Social Behavior and Primate Intelligence. Science, New Series, Vol. 153, No. 3735 (Jul. 29, 1966), pp. 501-506

Kahneman, D, (2011) Thinking Fast and Thinking Slow. *Doubleday, Canada*.

Kaplan, H., Hill, J., Lancaster, J. and Hurtado, A. M. (2000) A Theory of Human Life History Evolution: Diet, Intelligence, and Longevity. *Evolutionary Anthropology*. 9, (156–185).

Kary, C. E. (1984) Sociobiology and The Redemption Of Normative Ethics *The Monist*, Vol. 67, No. 2, (161-166).

Krebs, D. (2011) The origin of the moral senses. In The Origins of Morality: An Evolutionary Account. *Oxford University Press*.

Kuhn, T. (1962) The Structure of Scientific Revolutions. *The University of Chicago Press* (1996). Lakatos, Imre (1978) The Methodology of Scientific Research Programmes. *Cambridge University Press* (1978)..

Lehmann, L. & Keller, L. 2006a. The evolution of cooperation and altruism: a general framework and classification of models. J. Evol. Biol. 19: 1365–1376.

Lewontin, R.C. (1980) Theoretical Population Genetics in the Evolutionary Synthesis. In:The Evolutionary Synthesis, E. Mayr and W.B. Provine (eds.), *Cambridge MA, Harvard University Press*, 58–68.

Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., James R. and Blair, R. (2006) The neural basis of implicit moral attitude—An IAT study using event-related fMRI. *NeuroImage* 30 (2006) 1449 – 1457.

Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences USA* 104: 8597-8604

Maine, H. S. (1861) Ancient Law. John Murray, London.

Marshall, M. (2010) Sparks Fly Over Origin of Altruism. *New Scientist. Issue 2780,* September 29, 2010.

Marx, K. Capital. *J.M. Dent & Sons. London (1930)*.

Maynard-Smith, J. (1964) Group Selection and Kin Selection. *Nature* 201: 1145 1147.

Maynard-Smith, J. (1972) On Evolution. *Edinburgh University Press*.

Maynard Smith, J. (1982). Evolution and the Theory of Games. *Cambridge, Cambridge University Press*.

Maynard-Smith, J. (1998) The Origin of Altruism. *Nature* 393: 639-640.

Mayr, E. (1993) What was the evolutionary synthesis? Trends Ecol. Evol. 8:31-33

Michod, R. E. 1982 The theory of kin selection. A. Rev. Ecol. Syst. 13, 23-55.

Milgram, S. (1963). "Behavioral study of obedience." *Journal of Abnormal and Social Psychology* **67.**

Miller, Jon. (2014) "Hugo Grotius", *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.) URL = <http://plato.stanford.edu/archives/spr2014/entries/grotius/>.

Millstein. R. L. (2012) Exploring the Status of Population Genetics: The Role of Ecology. Biological Theory 7: 346-357.

Morgan, L. H. (1974) Ancient Society. *P. Smith., Gloucester, Mass. (1974)*.

Nesse, R. M. (2000). How selfish genes shape moral passions. In L. D. Katz (Ed.), Evolutionary origins of morality: Cross-disciplinary perspectives (pp. 227–231). *horverton, UK: Imprint Academic*

Nowak, M. A. (2012) Evolving Cooperation. *Journal of Theoretical Biology* 299 (2012) 1–8.

Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010) The Evolution of Eusociality. *Nature*. Vol. 466: 1057–1062 (August 2010).

Okasha, S. (2002) Genetic Relatedness and the Evolution of Altruism. *Philosophy of Science* 69: 138-149.

Okasha, Samir, (2008) "Population Genetics", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.). URL = <http://plato.stanford.edu/archives/fall2008/entries/population-genetics/>.

Okasha, S. (2009) Biological Altruism. *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.), URL=

http://plato.stanford.edu/archives/win2009/entries/altruism-biological/>.

Okasha, S. (2010) Altruism Researchers Must Cooperate. *Nature* VoL 467, 6 5 3-655. Okasha, S. (2012) Population Genetics. *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2012/entries/population-genetics/>.

Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature* 441: 502 505.

Packer, C. (1977) Reciprocal Altruism in Olive Baboons. *Nature* 265: 441-443.

Peirce, C. S., The Collected Papers of Charles Sanders Peirce. Vols. I-VI ed. Charles Hartshorne and Paul Weiss (*Cambridge, MA: Harvard University Press*, 1931 1935), Vols. VII-VIII ed. Arthur W. Burks (same publisher, 1958).

Pigliucci M (2008) Is Evolvability Evolvable? *Nature Reviews Genetics* 9:75–82.

Pigliucci, M., and Muller, G. B. (2010) Evolution: The extended Synthesis. *MIT Press*.

Pinker, Steven (2011) The Better Angels of Our Nature: Why Violence Has Declined. *Penguin Books*.

Pinker, Steven (2006) The Blank Slate. *The General Psychologist*. 41 (1).

Popper, K. R., (1963) Conjectures and Refutations: The Growth of Scientific Knowledge. *Harper and Row (1968)*.

Popper, K R., (1974) Darwinism as a metaphysical research programme. In *The*

Philosophy of Karl Popper. Editor P A Schilpp, 133-43. Vol. 1. LaSalle, Ill.: Open Court.

Prinz, Jesse (2012) Beyond Human Nature. *Penguin Books*.

Provine, W. B. (1971) The Origins of Theoretical Population Genetics. *University of Chicago Press*.

Queller, D. C. (1992) Quantitative Genetics, Inclusive Fitness, and Group Selection. The *American Naturalist*, Vol. 139, No. 3 (540-558).

Queller, D.C. & Strassmann, J.E. (2006) Models of cooperation. *Journal Evolutionary Biology* 19: 1410–1412.

Queller, D. C. and Strassmann, J. E. (1998) *Bioscience* 48 (3) 165 – 175.

Ratnieks, F. L. W. and Helantera, H. (2009) The evolution of extreme altruism and inequality in insect societies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 **(1533)** 3169-3179.

Reed, A. and Aquino, K. F. (2003) Moral Identity and the Expanding Circle of Moral Regard Toward Out-Groups. *Journal of Personality and Social Psychology.* Vol. 84, No. 6, 1270–1286.

Reid, Thomas (1765) An inquiry into the human mind, on the principles of common sense.

Rifkin, J (2010) The Empathic Civilization: The Race to Global Consciousness in a World in Crisis. Penguin, New York

Ross, L. (1977) The intuitive psychologist and his shortcomings. In: Advances in

Experimental Social Psychology, Vol 10 L. Berkowitz (ed.*) Academic Press*, New York.

Rousseau, Jean Jacques (1762/2010). The Social Contract. *Kindle Edition*.

Rousseau, Jean-Jacques. The Social Contract; and, Discourses. Published 1973.

Rousseau, Jean-Jacques, *Discourse on the Origin of Inequality* (Dover Thrift Editions) *Dover Publications. Kindle Edition* (2012).

Rousset, F. and Roze, D. (2007) Constraints on the Origin and Maintenance of Genetic Kin Recognition. *Evolution* 61-10: 2320–2330.

Ruse, M (2009) Creationism. *The Stanford Encyclopedia of Philosophy*. Spring 2009 Edition. http://plato.stanford.edu/archives/spr2009/entries/creationism/

Ruse, M (1984) The Morality of the Gene. *The Monist*, Vol. 67, No. 2, Sociobiology and Philosophy (APRIL, 1984), pp. 167-199.

Sidgwick, Henry (1876) The Theory of Evolution in its Application to Practice. *Mind*, Vol. 1, No. 1 (Jan., 1876), 52-67.

Singer, Peter. (2011) The Expanding Circle Ethics, Evolution, and Moral Progress. Princeton: *Princeton University Press*.

Sober, E. (1984) The Nature of Selection. *Dover Books, Princeton University Press*.

Sober, E. & Wilson, D. S. (1998) Unto Others: The Evolution and Psychology of Unselfish Behaviour. *Harvard University Press*. Cambridge.

Spinoza, B. Ethics. (ed.) John Fabian. (trans. R. H. M. Ewes, 1883). *Commodius Vicus*

Spinoza, B. Ethics: Treatise on the Emendation of the Intellect and Selected Letter. Hackett *Publishing Company, Indianapolis, Cambridge*.

Taylor, C and Nowak, M. A. (2007) Transforming the Dilemma. *Evolution* 61-10: 2281- 2292.

Tonnies, F. (1887) Gemeinschaft und Gesellschaft . See Jose Harris (*Translation*) Community and Civil Society. *Cambridge University Press (2001)*.

Trivers, R. L. (1971) The Evolution of Reciprocal Altruism. *Quarterly Review of Biology* 46: 35-57.

Traulsen, A. (2010) Mathematics of kin- and group-selection: formally equivalent? Evolution 64: 316–323.

Traulsen, A., Shoresh, N. & Nowak, M.A. (2008) Analytical results for individual and group selection of any intensity. Bull. Math. Biol. 70: 1410–1424.

vanVeelen, M., Garcia, J., Sabelis, M. W., Egas, M. (2012) Group selection and inclusive fitness are not equivalent; the Price equation vs. models and statistics.*Journal of Theoretical Biology* 299 (64–80).

van Veelen, M., Garcia J., Sabelis M. W., Egas M (2010) Call for a return to rigour in models. Nature 467:661.

Vogt, Katja (2015) "Seneca", *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/spr2015/entries/seneca/>.

Wade, M. (2005 [2008]) Evolutionary Genetics. *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed.), URL =<http://plato.stanford.edu/archives/fall2008/entries/evolutionary-genetics/>.

Wade, M. J. (1977) An experimental study of group selection. *Evolution* 31:134 153.

Wenseleers, T. (2006) Modelling social evolution: the relative merits and limitations of a Hamilton’s rule-based approach. *Journal Evolutionary Biology* 19: 1419–1422.

West, S. A., Griffin, A. S. and Gardner, A., (2007) Evolutionary Explanations for Cooperation - Review. *Current Biology* 17, R661–R672.

West, S.A, Gardner, A. (2013) Adaptation and inclusive fitness. *Current Biology* 23, R577–R584.

West, S. A. and Gardner, A. (2010) Altruism, Spite, and Greenbeards. *Science* 327,- 134i.

West-Eberhard, M. J. (2003) Developmental plasticity and evolution. *Oxford University Press*,

Wheeler, D. E. (1986) Developmental and Physiological Determinants of Caste in Social Hymenoptera: Evolutionary Implications. Am. Nat. 128 (1), 13-34.

Whewell, William (1847) The Philosophy of the inductive Sciences. *Frank Cass & Co, (1967)*.

Williams, C. G. and Williams D. C. (1957) Natural Selection of Individually Harmful Social Adaptations Among Sibs With Special Reference to Social Insects. *Evolution*, Vol. 11, No. 1 (Mar., 1957), pp. 32-39.

Wilson, E. O. (2005) Kin selection as the Key to Altruism: Its Rise and Fall*. Social Research.* Vol. 72 (1): 159-166.

Wilson, E. O. (1975) Sociobiology: The New Synthesis. *Belknap Press – Harvard University Press*.

Wilson, D. S. (1989) Levels of Selection: An Alternative to Individualism in Biology and the Human Sciences*. Social Networks* 11: 257-272.

Wilson, D. S. and Wilson, E. O. (2007) Rethinking the Theoretical Foundation of Sociobiology. *The Quarterly Review of Biology* vo. 82 (4) 327-348.

Wright, S. (1945). Tempo and Mode in Evolution: A Critical Review. Ecology 26:415- 419.

Wright, S. (1932). Evolution in Mendelian populations. Genetics 16:97-159.

Wundt, W. (1907) Outlines of psychology*. Leipzig: Wilhelm Englemann*, (1907).

Yakubu Y (2013a) The Altruism Paradox: A Consequence of Mistaken Genetic

Modeling. *Biological Theory* 8(1): 103-113. DOI 10.1007/s13752-013 0120-4.

Yakubu, Y. (2013b) Altruism: Analysis of a Paradox. *Lambert Academic Publishing* 2013) (ISBN 978-3-659-48069-0).

Yakubu, Y. (2012) Altruism: Analysis of a Paradox. *Open Access Dissertations and Theses.* Paper 6374. digitalcommons@mcmaster,ca.

Zajonc, R. B. (1980) Feeling And Thinking Preferences Need No Inferences. *American Psychologist*, Vol. 35, No. 2, 151-175.

Zimbardo, P. (1971) The pathology of imprisonment. *Society, 9*, 4-8. p.4.

Zwick, M. and Fletcher, J. A. (2014) Levels of Altruism. *Biological Theory* 9:100 107.