9

The Sociology of the Genome

> In the evolution of life. . . there has been a conflict between selection at several levels. . . individuality at the higher level has required that the disruptive effects of selection at the lower level be suppressed.
>
> John Maynard Smith

> A Mendelian population has a common gene pool, which is its collective or corporate genotype.
>
> Theodosius Dobzhansky

> **C:** What do you think of the book *One hundred authors against Einstein*?
>
> **AE:** Why 100 Authors? If I were wrong, then one would have been enough.
>
> Choreographer Interview

Many animals interact in groups for at least a part of their lifecyle. Such groups may be called flocks, schools, nests, troupes, herds, packs, prides, tribes, and so on, depending on the species. There appears not to be a common term for these groups, so I will call them *animal societies*, or simply *societies*. Animal societies have at least rudimentary social structures governing the typical interactions among group members. Even animals that live solitary lives often have mating practices involving signaling and ritualistic interactions (Noe and Hammerstein 1994; Fiske et al. 1998).

Sociobiology is the study of the social structure of such species. Edward O. Wilson introduced the term in his pathbreaking book (Wilson 1975). Wilson is an expert on social insects, not humans, but the concluding chapter of his book addressed human sociobiology. At the time, the idea that biology had anything useful to say about human society had few proponents. Virtually all social scientists at the time believed that the only thing biologically distinctive about humans was *hypercognition* (see Chapter 2), and that human behavior was completely determined by social and cultural institutions (Cosmides et al. 1992). Biology, it was thought, simply had nothing to add.

Wilson's book, not surprisingly, generated some years of heated and indeed venomous criticism (Segerstrale 2001). However, science eventually won out over tradition. We are now all sociobiologists. Indeed, the pendulum has perhaps swung too far in the other direction: all sorts of human behaviors are currently attributed, without much foundation, to our evolved dispositions (Gould and Lewontin 1979; Boyd and Richerson 2005).

Animal societies exist because living in a society enhances the fitness of its members. In economics this is called *increasing returns to scale*, the term applying perfectly to the aggregation of individuals in an animal society. While there may be some contingent and variable aspects of animal societies, with the exception of humans, the social structure of agiven species is quite uniform across time and space. The social structure of animal societies is thus likely optimized, or close to optimized, for contributing to the fitness of members of the species, within the bounds set by the gene pool of the species. The same cannot be said of human society, given of the massive effects of cumulative culture and technology (see Chapter 1),

The general social equilibrium model developed in Section 6.3 applies nicely to animal societies. There are social roles and social actors that fill these roles, the goal of social theory being to describe how actors are recruited to fill roles, and how roles interact to attain some degree of social efficiency. Sociobiology is part of sociology.

A basic principle of sociobiology is that behavior is conditioned by genes. In most species, age, sex, and caste condition the individual to assume a particular role. In highly social species, differential nurturing can create castes, such as worker vs. soldier vs. reproductive in eusocial bees and ants. In humans, of course, culture and socialization influence the allocation of individuals to social roles.

Basic evolutionary theory asserts that a gene for a particular behavior can persist in the population only if the behavior leads the gene's carrier (the individual) to contribute a sufficient number of copies of the gene to the next generation. The most straightforward way for this to occur is if the behavior enhances the fitness of the individual himself. Cooperation among social actors in this case is called *mutualistic* (Milinski 1996; Dugatkin 1997). Mutualistic interaction is particularly important in humans, and is called *collaboration* (Tomasello 2014). Genes for mutualism induce individuals to seek cooperative rather than solitary solutions to problems, and provide them with skills for effective collaboration.

Mutualism, however, is not enough to capture increasing returns to scale in social life. Often cooperation demands that participating individuals incur personal fitness costs. This is called *altruism*, and genes that code for altruistic behavior are called *altruistic genes*. Except in humans, this sort of biological altruism has no connection with moral sentiments, of course. Clearly, altruistic genes can spread only if the fitnesses of the beneficiaries of the altruistic act carrying the altruistic gene increase sufficiently to offset the sacrifice of the altruist. William Hamilton (1964a) was the first fully to develop this idea, culminating in *Hamilton's rule*. This rule says that if the altruist incurs fitness cost c, confers fitness benefit b on another individual with relatedness r to the altruist, the altruistic gene will spread if $br > c$. The reason is that br is the expected number of copies of the altruism gene gained in the recipient and c is the number of copies lost in the donor. Calling $br - c$ the *inclusive fitness* of the altruist, the implications of Hamilton's rule are called *inclusive fitness* theory.

My aim in this chapter is to clarify the position of inclusive fitness theory in sociobiology, drawing on Gintis (2014). The issue is highly contentious. Edward O. Wilson, for instance, who strongly supported Hamilton's analysis in the years immediately following its appearance, has become a serious critic. He writes in his recent book, *The Social Conquest of Earth* (2012):

> The foundations of the general theory of inclusive fitness based on the assumptions of kin selection have crumbled, while evidence for it has grown equivocal at best.... Inclusive fitness theory is both mathematically and biologically incorrect.

To supporters of inclusive fitness theory, this statement is outrageous, striking a blow at population genetics itself. As Stuart West et al. (2007a) explain:

> The importance of Hamilton's work cannot be overstated—it is one of the few truly fundamental advances since Darwin in our understanding of natural selection.

Richard Dawkins' (2012) review of *The Social Conquest of Earth*, exclaims:

> To borrow from Dorothy Parker, this is not a book to be tossed lightly aside. It should be thrown with great force.

Edward O. Wilson's critique culminated in a powerful paper, with coauthors Corina Tarnita and Martin Nowak (Nowak et al. 2010), that appeared in the high-profile journal *Nature*. The authors argue:

> considering its position for four decades as the dominant paradigm in the theoretical study of eusociality, the pro-

> duction of inclusive fitness theory must be considered meagre. . . inclusive fitness theory. . . has evolved into an abstract enterprise largely on its own.

This paper drew the ire of a host of population biologists. *Nature* subsequently published several "brief communications" vigorously rejecting the claims of Nowak, Tarnita, and Wilson. One of these was signed by no fewer than 137 well known biologists and animal behaviorists (Abbot 2011; Boomsma 2011; Strassmann 2011). In a leading biology journal article, Rousset and Lion (2011) accuse Nowak, Tarnita, and Wilson of saying nothing new and of using "rhetorical devices." They then attack the journal *Nature* itself, arguing that

> the publication of this article illustrates more general concerns about the publishing process.. . . *Nature*'s extravagant editorial characterization of the paper as "the first mathematical analysis of inclusive fitness theory" recklessly tramples on nearly 50 years of accumulated knowledge.

This controversy, a veritable clash of the titans (Gintis 2012a), has been avidly followed in the popular science literature, which has characterized the disagreement as to whether societies can be best model using concepts of *group selection* (with Nowak, Tarnita and Wilson) or *individual selection* (with Dawkins and the signers of protest letters to *Nature*), who argue that the notion that genes maximize inclusive fitness lies at the very *core* of evolutionary theory. For instance, West et al. (2011, p. 233) assert:

> Since Darwin, the only fundamental change in our understanding of adaptation has been Hamilton's development of inclusive fitness theory.. . . The idea [is] that organisms can be viewed as maximizing agents.

By contrast, opponents claim that *higher-level social organization is the driving force of evolutionary change*, and gene flows react by conforming to and promoting such higher-level social forms. For instance, Nowak et al. (2010) argue that eusocial species are successful because they develop social systems that suppress kin favoritism and promote generalized loyalty to the hive. Organisms that maximized inclusive fitness surely would not behave this way.

Prominent popular writers with solid academic backgrounds have strongly supported the inclusive fitness maximization position of Dawkins et al., yet do not seriously address the issues raised by Nowak, Wilson, and others (Pinker 2012; Coyne 2012).

I argue in this chapter that inclusive fitness theory is analytically valid, and is very important. However, it does not imply that individuals maximize inclusive fitness, and it fails to elucidate central driving forces in animal society formation and evolution. Nowak and Wilson correctly note the limitations of inclusive fitness theory, but they err in questioning its validity and in understating its contribution to sociobiology. Their critics correctly defend inclusive fitness theory, but they err in claiming that organisms in a social species maximize their inclusive fitness and that that inclusive fitness theory explains social structure.

The conditions under which evolutionary dynamics leads to inclusive fitness maximization have been careful studied by Alan Grafen and his associates, who have shown that Darwinian population dynamics entail inclusive fitness maximization at the individual and gene levels, but only assuming that fitness effects are *additive* (Grafen 1999, 2006; Gardner et al. 2011; Gardner and Welsh 2011). But if fitness effects were additive in general, then there would be no increasing returns to scale, and animal societies would not exist. Because societies are complex adaptive nonlinear systems, inclusive fitness is only *one tool* in the explanation of the social structure of animal societies.

Another way of expressing this point is that inclusive fitness theory applies to *single* gene in the organism's genome, or to several *non-interacting genes*. But the evolutionary success of an organism depends on the way the various genes *interact synergistically*. Claiming that inclusive fitness theory explains societies is like claiming that the analysis of word frequency in a book is sufficient to comprehend the book's meaning.

9.1 The Core Genome

Social relations in non-human societies are coded in the genes of its members. The characteristic rules of cooperation and conflict, as well as the meaning of signals passed among individuals, are *shared by all members of an animal society*. We call this communality of genes the species' *core genome*. The core genome is the complex of genes that are broadly shared by all members of a species (Dobzhansky 1953). Section 9.11 develops this notion in greater detail. The core genome is like the computer code for a software program in an agent-based computer model. The core genome sets up the rules for social interaction and the conditions for individual social success, creates a heterogeneous set of agents, each of whom incorporates both the core genome plus an idiosyncratic *variant genome* that defines its individuality. These agents interact according to the rules coded by the core genome, which rewards the more successful agents with more copies of itself in the future. In the case of human societies, additional rules and meanings are culturally specified, and as we explained in Chapter 1, human culture and the human core genome *coevolve*.

The core genome of a social species endows individuals with incentives to aggregate into social groups—packs, flocks, tribes, hives and the like. The size and social structure of these groups coevolve with the genetic constitution of its members, as reflected in the evolution of the core genome over time. Group selection is not selection *among* groups, but rather *for* groups with a fitness-enhancing size and social structure. Selection for group characteristics requires individual selection because the social rules are inscribed in individuals who both instantiate the rules and are evolutionarily successful given these rules.

Societies are complex dynamical systems with emergent properties—properties that we cannot deduce from the DNA of the core genome, any more than we can deduce consciousness and mind from the chemical composition of the brain (Deacon 1998; Morowitz 2002).

Yet societies are effective because of the behaviors of its members, these behaviors are determined by the core genome, and an individual gene can evolve only if it *directly* enhances the fitness of its carriers, or it promotes *interactions among its carriers* that enhance its *inclusive fitness*—the sum of the increases in fitnesses of all carriers of the gene influenced by the behavior. In particular, a gene that leads its carrier to sacrifice its inclusive fitness certainly cannot evolve, except possibly in very small societies where random luck can temporarily outweigh systematic selective forces.

Although the concept of the core genome is somewhat new, I cannot conceive of there being any serious objection to the above paragraphs. Indeed, the danger is more that they are uncomfortably close to tautologies.

Why then this conflict between group and individual selection proponents? The participants themselves agree that whether one does the accounting on the level of the group, the individual, or the single gene, the answer must come out the same (Dugatkin and Reeve 1994). What then can account for Richard Dawkins' venom in attacking Edward O. Wilson (Dawkins 2012), or David Sloan Wilson's sense of triumph in observing that group selection has been resurrected from its status as an outcast of biological theory (Wilson 2008)? Must it not be simply a matter of personal preference and modeling ease which perspective one chooses in any particular situation?

I suspect the answer is that inclusive fitness theorizing leads researchers to think *atomistically*, while group selection theorizing leads researchers to think *structurally*. Inclusive fitness theory leads one to the beautiful Margaret Thatcher headquote of Chapter 2: "There is no such thing as society. There are only individual men and women, and there are families." Group selection theorizing, by contrast, leads researchers to the Martin Luther King headquote in that chapter: "We are caught in an inescapable network of mutuality, tied in a single garment of destiny." Of course, I am not suggesting that sociobiologists are embroiled in the ideologies of Left and Right, or any other political ideology. Nor are they closely connected to any particular set of moral or ethical principles. Rather, they are personal preferences—highly contrasting yet equally useful ways of thinking about society. The correct way of thinking is to embrace both atomistic and structural approaches and analyze the corresponding interplay of forces. This is the approach defended in this chapter.

There is, however, a certain asymmetry in the mutual criticism of the two schools of thought. Few supporters of group selection deny the importance of inclusive fitness theory, while virtually all its opponents regularly deny the importance of group selection theory. For instance, Steven Pinker writes, quite disingenuously, in *The False Allure of Group Selection* (2012):

> Human beings live in groups, are affected by the fortunes of their groups, and sometimes make sacrifices that benefit their groups. Does this mean that the human brain has been shaped by natural selection to promote the welfare of the group in competition with other groups, even when it damages the welfare of the person and his or her kin?

The first problem with this description is that group selection does not require "competition with other groups" any more than individual selection requires "competition with other individuals." For instance, a mutant rabbit may be evolutionarily successful because it is more adept at escaping the fox, not because it wins conflicts with other rabbits. Similarly, a society may be evolutionarily successful because it better exploits its prey or contains its predators, not because it vanquishes other societies in head-to-head competition.

The more important problem with Pinker's critique is the notion that group selection theory suggests that the group's success depends on behaviors that damage "the welfare of the person and his or her kin." This is of course simply impossible. If the inclusive fitness of the gene for some behavior is less than unity, that gene must in the long run disappear from the population. No one disagrees with this.

Here is another rather randomly drawn, equally disingenuous, critique from a prominent biologist (Coyne 2012):

> The idea that adaptations in organisms result from "group selection"...rather than from selection among genes themselves...[is] in stark contrast to the views of most evolutionary biologists.

Of course, no group selection proponent sees group-level adaptations as an *alternative* to selection among genes. Rather, they think of group selection models as explanations of why particular gene are successful and others are not.

In the first half of the twentieth century, most naturalists believed that animal societies were effective because natural selection favors *altruism*, in the form of individuals who sacrifice for the good of the species (Kropotkin 1989[1903]; Simpson 1941; Lorenz 1963). For instance, in times of food scarcity, many believed that individuals would voluntarily restrict their reproductive activity (Wynne-Edwards 1962). This phenomenon was termed *group selection* because the argument was that the altruist may have fewer offspring, but its contribution to the success of the group would allow more of these offspring to survive and reproduce. However, John Maynard Smith (1964), George Williams (1966), David Lack (1966) and others showed that virtually all apparent examples of animals sacrificing for the group could plausibly be explained by standard individual fitness maximization. Williams (1966) used the *principle of parsimony* to counsel that group selection be used only when the simpler principle of individual selection is incapable of explaining animal behavior. At that time no important examples of sacrifice for the good of the group were found.

As it stands today, there are *two* mechanisms of group selection. The first is the *evolutionary success of more effective collaboration* (Parsons 1964; Boyd and Richerson 1990; Bowles and Gintis 2011; Tomasello 2014).

That is, social structures that effectively promote cooperation and punish antisocial behavior will tend to evolve. This mechanism works by an individual genetic mutation fostering a social structure mutation, the new social structure enhancing the fitness of social members, some of whom carry the mutant gene, which then is more frequently represented in the next generation. In this case it is the *social structure* that is favored by natural selection, and the genes that induce the behaviors given by the social structure are the beneficiaries of natural selection on the level of social structure.

Two forms of social organization are especially favored by this evolutionary process: *eusociality* and *extensive parental care*. In a eusocial species, one or very few individuals reproduce, and the remaining social members are sterile workers, soldiers, and foragers (Wilson 1975). Therefore a mutation in a reproductive will be inherited by a large fraction of her offspring, who will synergistically follow the principles of coordination, signaling, and task allocation indicated by the mutation. Not surprisingly, the eusocial insects have evolved into extremely complex and sophisticated societies—for instance the waggle dance in honeybees (Riley et al. 2005). A similar argument holds for animals that care for their young. Because there are at only one or two individuals involved in mating and in nurturing offspring, a mutation in a male or female leading to a new social structure of mating can easily spread. Darwin called this *sexual selection*, an evolutionary process that has engendered sophisticated signaling and collaboration in many species (West-Eberhard 1983).

The second mechanism of group selection is exactly the altruistic behavior that had be discredited by Williams, Maynard Smith, Lack, and others, although now better understood in terms of game-theoretic models of social cooperation. Often the effectiveness of social cooperation is strongly enhanced when individuals are willing to incur personal costs to further collective goals. For instance, when a group of human hunters venture into the forest, they usually fan out in such way that they are not visible to one another. Because the prey is share irrespective of who killed the animal (Kaplan et al. 1984), and since the process of searching for prey is highly strenuous, each hunter has an incentive to shirk. Altruists do not. Successful groups foster altruism, which complements mutualistic collaboration in promoting efficient cooperation. Note that for species in general, this notion of biological altruism has nothing to do with either morality or psychology.

This sort of altruism was recognized by Darwin himself (Darwin 1871):

> An advancement in the standard of morality will certainly give an immense advantage to one tribe over another. A tribe including many members, who from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection.

The existence of altruism, the importance of which is not now widely disputed, nevertheless presents a serious problem for evolutionary theory: How can genes that promote altruistic behavior spread, since they disadvantage their carriers? Inclusive fitness theory provides the answer.

9.2 Inclusive Fitness and Hamilton's Rule

Classical genetics does not model cases in which individuals sacrifice on behalf of non-offspring, such as sterile workers in an insect colony (Wheeler 1928), cooperative breeding in birds (Skutch 1961), and altruistic behavior in humans (Darwin 1871). This problem was addressed by William Hamilton (1963, 1964ab, 1970), who noticed that if a gene favorable to helping others is likely to be present in the recipient of an altruistic act, then the gene could evolve even if it reduces the fitness of the donor. Hamilton called this *inclusive fitness* theory.

Hamilton developed a simple inequality, operating at the level of a single locus that gives the conditions for the evolutionary success of an allele. Thus rule says that if an allele in individual A, I will call it the *focal allele*, increases the fitness of individual B whose degree of relatedness to A is r, and if the cost to A is c, while the fitness benefit to B is b, then the allele will evolve (grow in frequency in the reproductive population) if

$$br > c. \tag{9.1}$$

We call $br - c$ the *inclusive fitness* of the focal allele. Subsequent research supported some of Hamilton's major predictions (Maynard Smith and Ridpath 1972; Brown, 1974; West-Eberard 1975; Krakauer, 2005).

A critical appreciation of Hamilton's rule requires understanding when and why it is true. The rigorous derivations of Hamilton's rule (Hamilton 1964a; Grafen 1985; Queller 1992; Frank 1998) are mathematically sophisticated and difficult to interpret. For this reason, it is easy to assert implications of inclusive fitness that cannot be evaluated by a non-expert. I suspect that this accounts for the fact that non-experts have tended to support one or another in this debate without really understanding the technical issues involved. My goal in this chapter is to lay these issues bare, so that they can be appreciated by anyone willing to endure a bit of elementary algebra.

The usual popular argument (for instance, Bourke 2011) assumes that an altruistic helping behavior ($b, c > 0$) is governed by an allele at a single locus, and r is the probability that the recipient of the help has a copy of the helpful allele. The net fitness increment to carriers of the helpful allele is then $br - c$, so the allele increases in frequency if this expression is positive.[1]

However attractive, the popular argument has key weaknesses that render it unacceptable. First, the intuition behind Hamilton's rule is that r is the probability that the recipient has a copy of the helping gene, so br is the expected gain to the helping gene in the recipient, which must be offset by the loss c to the helper if the helping behavior is to spread. This argument, however, is clearly specious. Many have pointed this out, but perhaps none more elegantly than Washburn (1978, p. 415), who writes:

> All members of a species share more than 99% of their genes,
> so why shouldn't selection favour universal altruism?

Dawkins (1979) considers Washburn's argument the fifth of his "Twelve Misunderstandings of Kin Selection." Dawkins draws on Maynard Smith (1974) to show that Washburn's conclusion in favor of universal altruism is faulty. But he fails to explain what is wrong with Washburn's argument, except to say (correctly): "This misconception arises not from Hamilton's own mathematical formulation but from oversimplified secondary sources to which Washburn refers." (p. 191) One might, with Dawkins (1979), claim that r represents the probability of the identity of the helping gene in the two parties by *descent from a common ancestor*, but why should it matter whether it is identity by descent or otherwise? Descent is clearly beside the point. A copy is acopy, whatever its provenance.

In fact, part of the problem with the popular argument is rather subtle: it considers the conditions for an increase in the *absolute number* of copies

[1] Hamilton's rule extends directly to behavior that is governed by alleles at multiple loci, provided that the interactions among the loci are frequency independent, or equivalently, that the effects at distinct loci contribute additively to the phenotypic behavior. Grafen (1984) calls a such a phenotype a *p-score*. In this chapter I will use the term "single locus" even in places where the p-score generalization applies.

of the helping allele in the population, but says nothing about its *relative frequency*, which is the quantity relevant to the evolutionary success of the helping allele. Indeed, $br - c$, the net increase in the number of copies of the helpful allele, is less than $b - c$, which is the net increase in the number of copies of all alleles at the locus, so the frequency of the helping allele in the next period will be *lower* than $br - c$, and *prima facie* may even decrease.

A second problem with the popular argument is that it makes sense if the relatedness r is a *probability*, so that br can be interpreted as the expected gain to the helping allele in the beneficiary. But in this case r must be *nonnegative*. By contrast, in a valid derivation of Hamilton's rule, r can be positive *or* negative. In the case $c > 0$ and $b, r < 0$, but with $br > c$, we call this *spite* (Hamilton 1970; Gardner et al. 2004). In fact, as we shall see, there is no simple relationship between the r in Hamilton's rule and genealogical coefficients of relatedness. The appropriate value of r in Hamilton's rule is necessarily less than unity, but it is generally a function of the social structure of the species in question, and can be positive or negative.

To address these deficiencies, we begin our study of inclusive fitness theory with a careful derivation of Hamilton's rule assuming, with Hamilton, that all interactions are dyadic. For simplicity, I will assume the species is haploid but sexual. That is, each new individual inherits a single gene from one of its two parents at each locus of the genome. A more general diploid treatment (individuals have two alleles at each genetic locus) is presented in Appendix A1, where we also drop the requirement that all interactions must be dyadic.

Our derivation of Hamilton's rule makes numerous simplifying assumptions. However, the argument can be extended to deal with heterogeneous relatedness, dominance, coordinated cooperation, local resource competition, inbreeding, and other complications (Uyenoyama and Feldman 1980; Michod and Hamilton 1980; Queller 1992; Wilson et al. 1992; Taylor 1992; Rousset and Billard 2007), with an equation closely resembling (9.1) continuing to hold. In general, however, the frequency q of the focal allele will appear in (9.1), and b and c may be functions of q as well, so the interpretation of r as relatedness becomes accordingly more complex (Michod and Hamilton 1980).

In general b and c will also depend on the frequency of alleles at other loci of the genome, and since the change in frequency q of the focal allele in the population will affect the relative fitnesses of alleles at other loci, inducing changes in frequency at these loci, which in turn will affect the values of b, c, and even r. For this reason, Hamilton's rule presupposes *weak selection*, in the sense that population gene frequencies do not change appreciably in a single reproduction period. Therefore Hamilton's rule does not imply that a successful allele will move to fixation in the genome. Moreover, alleles at other loci that are enhanced in inclusive fitness by the focal allele's expansion may undergo mutations that enhance the inclusive fitness of the focal allele, while alleles at other loci that are harmed by the expansion of the focal allele may develop mutations that suppress the focal allele. Such mutations can be evolutionarily successful and even move to fixation in the core genome.

Now to our derivation. Suppose there is an allele at a locus of the genome of a reproductive population that induces carrier A (called the *donor*) to incur a fitness change c that leads to a fitness change b in individual B (called the *recipient*). We will represent B as an individual, but in fact, the fitness change b can be spread over any number of individuals. If $b > 0$, A bestows a *gain* upon B, and if $c > 0$, A experiences a fitness *loss*. However, in general we make no presumption concerning the signs or magnitudes of b and c, except that selection is weak in the sense that b and c do not change, and the population does not become extinct, over the course of a single reproduction period. This assumption, which is extremely plausible, will be made throughout this chapter.

Suppose the frequency of the focal allele in the population is q, where $0 < q < 1$, and the probability that B has a copy of the allele is p. Then if the size of the population is n, there are qn individuals with the focal allele, they change the number of members of the population from n to $n + qn(b - c)$, and they change the number of focal alleles from qn to $qn + qn(pb - c)$. Thus the frequency of the allele from one period to the next will increase if

$$\Delta q = \frac{qn + qn(pb - c)}{n + qn(b - c)} - q = \frac{q(1-q)}{1 + q(b-c)} \left(b\frac{p-q}{1-q} - c \right) > 0. \quad (9.2)$$

The condition for an increase in the focal allele thus is

$$b\left(\frac{p-q}{1-q}\right) > c. \quad (9.3)$$

To derive Hamilton's rule from (9.3), we must have

$$r = \frac{p - q}{1 - q}, \tag{9.4}$$

which can be rewritten as

$$p = r + (1 - r)q. \tag{9.5}$$

Equation (9.5) makes intuitive sense using the concept of *identity by descent* (Malécot 1948; Crow 1954), where r is the probability that both donor A and recipient B have inherited the same focal allele from a common ancestor. For instance, if A and B are full siblings, then $r = 1/2$ because this is the probability that both have inherited the focal allele from the same parent. Moreover, if the siblings have inherited the focal allele from different parents, then they will still be the same allele with a probability equal to the mean frequency q of the focal allele in the population, assuming no assortative mating. In general, r will then be the expected degree of identity by descent of recipients. This logic is developed in full by Michod and Hamilton (1980).

However, this cannot be the general argument because there is no reason for p to be greater than q; i.e., the recipient need not be more likely than average to carry the helping gene. But if $p < q$, then equation 9.4 shows that $r < 0$, so r cannot be interpreted as a genealogical relatedness coefficient. Population biologists have generally responded to this problem by defining r as a beta coefficient in a least squares linear regression of the donor genotype on the recipient phenotype (Hamilton 1972; Queller 1992). This is an elegant approach, but rather mystifying. Why linear regression? Why least squares estimation? Why is it not just an approximation, as with standard linear regressions? Why is it a good approximation, given the strong nonlinear interactions of loci in the genome? It is comforting that the approach gives a reasonable result in many cases, but the conceptual foundations are quite shaky. Moreover, for an elementary exposition, like the present, where the reader should be able to follow perfectly what is going on, it is like the magician pulling a rabbit out of a hat.

There is another way to explain negative relatedness while sticking to a rigorously correct logic. Each potential recipient B has a certain relatedness to the donor A. Therefore we can partition the population of potential recipients into groups $j = 1, \ldots, k$ such that all individuals in group j have the same genealogical relatedness r_j to the donor A. Let q_j be the mean frequency of the helping allele in group j, and let π_j be the probability that the donor encounters a recipient from group j, so $\sum_j \pi_j = 1$. Then the probability that a recipient in group j has a copy of the helping allele is, using the same reasoning as led to equation (9.5),

$$p_j = r_j + (1 - r_j)q_j. \tag{9.6}$$

Moreover, we have $p = \sum_j \pi_j p_j$, and if we define $r^* = \sum_j \pi_j r_j$ and $q^* = \sum_j \pi_j q_j$, we then have

$$\begin{aligned} p &= \sum_j \pi_j (r_j + (1 - r_j)q_j) \\ &= r^* + q^* - \sum_j \pi_j r_j q_j \\ &= r^* + (1 - r^*)q^* - \left(\sum_j \pi_j r_j q_j - r^* q^* \right), \\ &= r^* + (1 - r^*)q^* - \mathrm{cov}_\pi(r_j, q_j), \end{aligned} \tag{9.7}$$

from the definition of the covariance of two variables. If we recast this result in terms of the standard equation (9.5), we get

$$r = \frac{r^* + (1 - r^*)q^* - q - \mathrm{cov}_\pi(r_j, q_j)}{1 - q}.$$

Note that this reduces to the identity $r = r^*$ when $q = q^*$ and the covariance term is zero.

Two points are notable in equation (9.7). First, p can now be smaller than q, so $r < 0$ is possible in (9.5). Indeed, this is more likely the smaller is q^*, the average frequency of the helping allele in the donor's potential beneficiaries and the larger the covariance between relatedness and mean frequency of the helping allele. The latter effect enters because p will be higher if low-relatedness beneficiaries tend to have high average q_j because low r_j means the random allele will be chosen with high frequency. For example, if there is only one group ($k = 1$), the covariance term in 9.7 drops out and we can write

$$p - q = r^* + (1 - r^*)q^* - q = r(1 - q) + (1 - r)(q^* - q). \tag{9.8}$$

The first term on the right hand side of (9.8) is positive but the second is negative for $q^* < q$, and the second term dominates when r is small; i.e., when the behavior attacks non-relatives that do not share the focal allele.

It is reasonable to call the array $\{\pi_j, r_j, q_j\}$ the *social structure* of the population with respect to the behavior induced by the helping allele. This array in general is not defined at the level of the helping locus, but at the social level, coded by the core genome. The core genome determines particular mating patterns, particular rituals and signals, certain patterns of offspring care and social collaboration. *Inclusive fitness thus presupposes a general type of social structure* and does not elucidate this social structure.

While the simple inequality $br > c$ at first sight appears to connect genealogical relatedness, costs, and benefits at the level of a single locus, in fact a correct derivation of the inequality reveals a complex social structure underlying each of the three terms. This fact does not detract from the importance of Hamilton's rule. Indeed Hamilton's rule must be satisfied by any plausible social structure. But it is an accounting relationship, not an explanatory model.

9.3 Kin Selection and Inclusive Fitness

William Hamilton's early work in inclusive fitness focused on the role of genealogical kinship in promoting prosocial behavior. Hamilton speculates, in his first full presentation of inclusive fitness theory (Hamilton 1964a, p. 19):

> The social behaviour of a species evolves in such a way that in each distinct behaviour-evoking situation the individual will seem to value his neighbours' fitness against his own according to the coefficients of relationship appropriate to that situation.

Because of this close association between inclusive fitness and the social relations among genealogical relatives, John Maynard Smith (1964) called Hamilton's theory *kin selection*, by which he meant that individuals are predisposed to sacrifice on behalf of highly related family members.

A decade after Hamilton's seminal inclusive fitness papers, motivated by new empirical evidence and Price's equation (Price 1970), Hamilton (1975, p. 337) revised his views, writing:

> Kinship should be considered just one way of getting positive regression of genotype…the inclusive fitness concept is more general than kin selection.

Nevertheless the two concepts are often equated, even in the technical literature. For instance, throughout his authoritative presentation of sexual allocation theory, West (2009) identifies inclusive fitness with kin selection in several places and never distinguishes between the two terms at any point in the book. Similarly, in Bourke's (2011) ambitious introduction to sociobiology, we find:

> The basic theory underpinning social evolution [is] Hamilton's inclusive fitness theory (kin selection theory).

This curious identification of inclusive fitness theory, which models the dynamics at a single genetic locus and is equally at home with altruistic and predatory genes, as we explain below, with kin selection theory, which is a high-level behavioral theory of kin altruism, is a source of endless confusion. For most sociobiologists, kin selection remains, as conceived by Maynard Smith (1964), a social dynamic based on *close genealogical association*:

> By kin selection I mean the evolution of characteristics which favour the survival of close relatives of the affected individual.

The Wikipedia definition is similar:

> Kin selection is the evolutionary strategy that favours the reproductive success of an organism's relatives, even at a cost to the organism's own survival and reproduction.…Kin selection is an instance of inclusive fitness.

Moreover, while kin selection is a special case of inclusive fitness in the sense that Hamilton's rule applies generally, not just to situations where organisms favor their close genealogical kin, in another sense kin selection is far more general than inclusive fitness. This is because in all but the simplest organisms, kin selection does not describe the behavior at a single locus, or even at a set of independently contributing loci, but rather an inherently *high level social behavior* in which individuals recognize their close relatives through complex phenotypic associations that require significant cognitive functioning and synergistic interactions among loci. Indeed, in general these phenotypic associations arise precisely to permit cooperation among close genealogical kin.

9.3.1 Inclusive Fitness without Kin Selection

A simple example shows that Hamilton's rule in principle has no necessary relationship with genealogy or kin selection, but rather is an expression of the social structure of the reproductive population. The model is based on, but is more transparently presented than Hamilton (1975), which develops a similar model for the same purpose. For related models of positive assortment not based on kin selection see Koella (2000), Nowak (2006), Pepper (2007), Fletcher and Doebili (2009), and Smaldino et al. (2013).

Consider a population in which groups of size n form in each period. In each group individuals can cooperate by incurring a fitness cost $c > 0$ that bestows a fitness gain b that is shared equally among all group members. Individuals who do not cooperate (defectors) receive the same share of the benefit as cooperators, but do not pay the cost c and do not generate the benefit b. Let p_{cc} be the expected fraction of cooperating neighbors in a group if an individual is a cooperator, and let p_{cd} be the expected fraction of cooperating neighbors if the individual is a defector. Then the payoff to a cooperator is $\pi_c = bp_{cc} - c$, and the payoff to a defector is $\pi_d = bp_{cd}$.

The condition for the cooperative allele to spread is then $\pi_c - \pi_d = b(p_{cc} - p_{cd}) - c > 0$, or

$$b(p_{cc} - p_{cd}) > c. \tag{9.9}$$

Now p_{cc} is the probability that a cooperator will meet another cooperator in a random interaction in a group, so we can define the relatedness r between individuals, following (9.5), by

$$p_{cc} = r + (1 - r)q, \tag{9.10}$$

where q is the mean frequency of cooperation in the population. If we write $p_{dd} = 1 - p_{cd}$ for the probability that a defector meets another defector, then we similarly can write

$$p_{dd} = r + (1 - r)(1 - q), \tag{9.11}$$

since $1 - q$ is the frequency of defectors in the population. Then we have

$$\begin{aligned} p_{cc} - p_{cd} &= r + (1 - r)q - (1 - (r + (1 - r)(1 - q))) \qquad (9.12) \\ &= r. \qquad (9.13) \end{aligned}$$

Substituting in (9.9), we recover Hamilton's rule, $br > c$.

Of course, if group formation is random, then $p_{cc} = p_{cd}$ so $r = 0$ and Hamilton's Rule cannot hold. However, to illustrate the importance of social structure, suppose each group is formed by k randomly chosen individuals who then each raises a family of n/k clones of itself. We need not assume parents interact with their offspring, or that siblings interact preferentially with each other. There is no kin selection in the standard sense of Maynard Smith (1964). At maturity, the parents die and the resulting n individuals interact, but do not recognize kin. In this case a cooperator surely has $k-1$ other cooperators (his sibs) in his group, and the other $n-k$ individuals are cooperators with probability q. Thus

$$p_{cc} = \frac{k-1}{n-1} + \frac{n-k}{n-1}q = q + \frac{(k-1)(q-1)}{n-1}.$$

Similar reasoning, replacing q by $1-q$ gives

$$p_{dd} = 1-q + \frac{q(k-1)}{n-1}.$$

Then

$$r = p_{cc} - p_{cd} = p_{cc} - 1 + p_{dd} = \frac{k-1}{n-1},$$

so Hamilton's rule will hold when

$$br = b\left(\frac{k-1}{n-1}\right) > c.$$

Note that the related recipients are all clones of the donor, with relatedness unity, although the r in Hamilton's rule is $(k-1)/(n-1)$. The inclusive fitness inequality is accurate here, but kin selection as defined above is inoperative in this model: the altruistic behavior is more likely to spread when the number of families n/k in a group is small.

This model suggests that the interesting question from the point of view of sociobiology is how the core genome of the species manages to induce individuals to aggregate in groups of size n and to limit family size to n/k, so that the benefits of cooperation $(b-c)$ can accrue to the population. This is a true miracle of Nature.

9.4 A Generalized Hamilton's Rule

When we think of Hamilton's rule in the context of an animal society, we must account for the possibility that the focal allele may impose a cost β uniformly on all members of the population, We call this a *social fitness effect*. The case $\beta > 0$ may be termed a *pollution effect*. It occurs, for instance, in "tragedy of the commons" cases (Hardin 1968; Wenseleers and Ratnieks 2004), such as when the focal allele depletes a protein used in chemical processes by somatic cells in conferring the benefit b on others and incurring a cost c (Noble 2011). The case $\beta < 0$ may be called *public good effect* (West et al. 2007b). This follows the common use of the term in economic theory (Olson 1965).It occurs in a parasite when the focal allele induces its carriers to suppress an alternative allele at the focal locus that induces carriers to grow so rapidly that it kills its host prematurely (Frank 1996). Equation (9.17) below shows *the degree of pollution or public good has no bearing on whether the allele can evolve*.

It is interesting to note that Hamilton's seminal paper (1964a) explicitly includes the pollution and public goods aspect of inclusive fitness, an aspect of his analysis that later writers have ignored. Hamilton called the public good/pollution effect the *dilution effect* because it affects the *rate* but not *direction* of change in the frequency of the focal allele. Hamilton also notes that the dilution effect can lead a successful allele to *reduce* population fitness. A streamlined presentation of Hamilton's argument, which is quite opaque in the original, is presented in Gintis (2014).

We will also consider the case where the focal allele imposes a cost α on all alleles *other than* the focal allele (Keller and Ross 1998). We may call α a *thieving effect*. For example, $\alpha > 0$ can occur if A redirects brooding care from non-relative to relative larvae in an insect colony, and $\alpha < 0$ (stealing from one's kin to help others) can occur if the focal allele helps other alleles at the focal locus that benefits carriers by avoiding possibly deleterious homozygosity at the focal locus. We can clearly treat α as cost imposed on all alleles at the focal locus, plus a benefit of equal magnitude enjoyed by carriers of the focal allele. Thus if the population size is n in the current period, population size n' in the next period will include $n + qn(b + \alpha - c)$ individuals because of the behavior induced by the focal allele, but this will be reduced by $n(\alpha + \beta)q$ due to the effects on non-focal alleles. The number of relatives of the focal allele in the current period is qn, which is increased by the behavior by $qn(pr + \alpha - c)$, and decreased through lower efficiency by $qn(\alpha + \beta)q$. Thus the new population size is given by

$$n' = n(1 - (\alpha + \beta)q) + qn(b + \alpha - c), \tag{9.14}$$

and (9.2) becomes

$$\Delta q = \frac{qn(1 - (\alpha + \beta)q) + qn(pb - \alpha - c)}{n(1 - (\alpha + \beta)q) + qn(b - \alpha - c)} - q > 0, \tag{9.15}$$

which simplifies to

$$b(p - q) > (c - \alpha)(1 - q). \tag{9.16}$$

Substituting $p = r + (1 - r)q$, we get the generalized Hamilton's rule

$$br > c - \alpha. \tag{9.17}$$

The effect of an increase in the focal allele on population fitness is the sign of dn'/da, where $a = qn$ is the number of helping genes, which is given by

$$\frac{dn'}{da} = b - c - \beta. \tag{9.18}$$

Note that in the case of Hamilton's rule, which is the above with $\alpha = \beta = 0$, population fitness increases with the frequency of the focal allele in the case of altruism or cooperation, where $b > c$, and decreases in the case of spite ($b - c < 0$). In the case of the generalized Hamilton's rule, the fitness effect is indeterminate. As we explain below, Hamilton (1964a) included the $\beta \neq 0$ affect in his calculations, but he did not consider the case where the generalized fitness effects are unevenly distributed among the alleles at the focal locus ($\alpha \neq 0$).

It is useful to give descriptive names to the social interactions when α is nonzero. We may call the case $\alpha > 0$ *theft*, and the case $\alpha < 0$ as *charity*. Moreover, a thieving altruist ($b, c, \alpha > 0$) will evolve, as will a thieving cooperative allele ($b, \alpha > 0 > c$). Finally, the producer of a public good will evolve only if it gains in inclusive fitness from so doing ($br > c$).

The most critical implication of the generalized Hamilton's rule is that neither social generosity nor pollution has any bearing on whether an allele will evolve, as seen in equation (9.17), despite the fact that a socially generous allele unambiguously enhances the population fitness, and a polluting allele unambiguously has the opposite effect, as seen in equation (9.18). In addition, a thieving allele does not directly affect the mean population fitness (see equation 9.18) but it allows the generalized Hamilton's rule to be satisfied even when $br - c < 0$ (see equation 9.17).

9.5 Harmony and Disharmony Principles

A rather stunning conclusion can be drawn from our exercise in elementary algebra and gene-counting. I call it the Harmony Principle. To state this principle succinctly, we say the allele a is *helpful* if its carriers enhance the fitness of other individuals that it encounters ($b > 0$), *altruistic* if it is helpful and incurs a fitness cost ($b, c > 0$), *predatory* if it is harmful to others but helps itself ($b, c < 0$), *mutualistic* if it helps itself and others ($b > 0$, $c < 0$), *prosocial* if increases mean population fitness ($b - c > 0$), and *antisocial* if it reduces mean population fitness ($b - c < 0$). We then have:

Harmony Principle: An evolutionarily successful gene that is a helpful non-polluter is necessarily prosocial.

From equation (9.18), the allele is prosocial if $b - c - \beta > 0$. We can write $b - c - \beta = (br - c) + b(1 - r) - \beta$. Now $br - c > 0$ by Hamilton's rule, $b > 0$ by the assumption of helpfulness, since the probability p of the recipient having the helpful allele is nonnegative, $r < 1$ by equation 9.4, and $\beta <= 0$ because the allele is a non-polluter. Thus $b - c - \beta$, the net contribution per focal allele to the population, is strictly positive.

Because each individual gene is utterly selfish, the importance of this principle for sociobiology is *inestimable*, and mirrors similar assertions concerning the social value of selfishness in humans offered by Bernard Mandeville in his famous *Fable of the Bees* (1705), in which "private vices" give rise to "public virtues," and Adam Smith's (1776) equally famous dictum, "It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest." While economists have determined the precise conditions—they are far from universal—under which Mandeville and Smith are correct (Mas-Colell et al. 1995), the Harmony Principle is true under much broader conditions. While genes are utterly selfish according to inclusive fitness theory, evolutionarily successful genes that are helpful non-polluters are necessarily prosocial. Note that we have not assumed that $c > 0$, so this principle applies both to altruistic genes and mutualistic genes that help others as well as helping themselves.

However, what is the social status of genes that are *not* helpful? It is curious that this case appears never to have been treated in the literature. I cannot imagine why not. An alternative to the Harmony Principle is, indeed, *prima facie* equally possible. Suppose the focal allele is predatory.

Then Hamilton's rule becomes $(-b)r < (-c)$, which can be satisfied even though the focal allele is antisocial. Indeed, this will be the case whenever $|b|(1-r) > c-b > 0$. We have:

Disharmony Principle: A gene that is evolutionarily successful but predatory may be antisocial even if it is a non-polluter.

To see this note that $b-c-\beta = (br-c)+b(1-r)-\beta$, where $br-c > 0$ and $1-r > 0$. Thus for sufficiently large (negative) b, we must have $b-c < 0$.

Note that the Disharmony Principle is distinct from the *spite* phenomenon (Hamilton 1970; Foster et al. 2001; Gardner et al. 2004), in which $r < 0$ and $c > 0$, which is well developed in the literature. Indeed, it is a common occurrence that the interaction is costly but involves reducing the fitness of others, and (9.5) can hold with $r < 0$, while the focal allele is still altruistic (Bourke 2011). Examples are warfare in ants (Hölldobler and Wilson 1990) and humans (Bowles and Gintis 2011), as well as generally spiteful behavior in many species (Hamilton 1970; Foster et al. 2001; Gardner et al. 2004).

More generally, we have the taxonomy of Table 9.1, where if $b > 0 > c$, then the allele is *cooperative*, and since $b-c > 0$, the allele contributes unambiguously to the fitness of its carrier. A cooperative allele will always be selected, as in this case Hamilton's rule is always satisfied. The unnamed boxes in the table necessarily violate Hamilton's rule.

b	c	$r > 0$	$r = 0$	$r < 0$
> 0	> 0	Altruistic	—	
> 0	< 0	Cooperative		
< 0	> 0	—		Spiteful
< 0	< 0	Predatory		

Table 9.1. Variety of behaviors that can satisfy Hamilton's Rule

9.6 The Utterly Selfish Nature of the Gene

Hamilton's rule ensures that the gene is *selfish* in the sense described by Dawkins (1976). In particular, Hamilton's rule implies that *the conditions for the evolutionary success of a gene are distinct from the conditions under* *which the gene enhances the mean fitness of the reproductive population*. The validity of the Disharmony Principle shows that inclusive fitness does *not* explain the appearance of design in nature or, in other words, why the genome of a successful species consists of genes that predominantly *collaborate* in promoting the fitness of its members (Dawkins 1996). Indeed, Hamilton's rule equally supports the evolutionary success of prosocial altruistic genes and antisocial predatory genes, whereas the former predominate in a successful species and account for the appearance of design.

It is common for sociobiologists who, with Dawkins, adopt the "gene's-eye point of view" to overlook this fact, despite its being a simple logical implication of Hamilton's rule. Indeed, many population biologists claim that the appearance of design in nature is explained by Hamilton's rule. For instance, in the protest letter to *Nature* mentioned above, 137 professional evolutionary biologists agreed with the following statement:

> Natural selection explains the appearance of design in the living world, and inclusive fitness theory explains what this design is for. Specifically, natural selection leads organisms to become adapted as if to maximize their inclusive fitness (Abbot 2011).

In fact, as we shall see, organisms do *not* generally maximize inclusive fitness. Rather, organisms in a social species interact strategically in a complex manner involving collaboration, as well as enhancement and suppression of gene expression. Moreover, relatedness may play a *derivative* role in the dynamics of a species, especially a species that exhibits a complex division of labor involving the suppression of kin altruism.

Inclusive fitness theory, however, permits a formulation of the central problem of sociobiology in a particularly poignant form: *how do interactions among loci induce utterly selfish genes to collaborate, or to predispose their carriers to collaborate, in promoting the fitness of the organism?* Inclusive fitness theory, because it ignores interactions among loci, does not answer this question. But it does provide important insights.

Fitness-enhancing collaboration among loci in the genome of a reproductive population requires suppressing alleles that decrease, and promoting alleles that increase the fitness of its carriers. Suppression and promotion are effected by *regulatory gene networks*, each member of which is itself utterly selfish. This implies that genes, and *a fortiori* individuals in a social species, do not generally maximize inclusive fitness but rather interact strategically in complex ways. It is the task of sociobiology to model these complex interactions.

9.7 Prosocial Genes Maximize Inclusive Fitness

Egbert Leigh (1971) famously compared the genome to a *parliament of genes*:

> each acts in its own self-interest, but if its acts hurt the others,
> they will combine together to suppress it.

Leigh was concerned with the maintenance of Mendelian segregation, but the remark applies quite broadly. Certainly some such mechanism must account for the tendency of genes in the genome to cooperate. However, the mechanism does not operate through inclusive fitness maximization.

To see this, we return to the model explored in Section 9.2. Let q_a be the frequency of the focal allele in the population, and let p_a be the probability that the recipient shares a copy of this allele. Now let q_b be the frequency of some allele b at another locus of the genome, and let p_b be the probability that the recipient shares a copy of this allele with the donor. Note that the cost c imposed on the donor is imposed *equally* on the allele b assuming Medelian segregation, because both allele a and allele b have probability ½ of being passed on to each offspring. Similarly, allele b receives the same benefit b as the focal allele in all carriers of both alleles. Therefore if the size of the population is n in the current period, the size in the next period will be $n+qn(b-c)$ and the number of b alleles will be $nq_b+qn(bp_b-c)$. Thus the change in frequency of allele b is given by

$$\begin{aligned}\Delta q_b &= \frac{nq_b + qn(bp_b - c)}{n + qn(b-c)} - q_b\\ &= \frac{q_a(1-q_b)}{1+q_a(b-c)}\left(b\frac{p_b - q_b}{1-q_b} - c\right)\\ &= \frac{q_a(1-q_b)}{1+q_a(b-c)}(br-c). \qquad (9.19)\end{aligned}$$

Note that we have used the equation $r = (p_b - q_b)/(1-q_b)$, which is equation (9.4) for allele b.

Thus every allele at locus B benefits from the behavior induced by the focal allele, and hence a mutation at locus B that suppresses the focal allele will, *ceteris paribus*, be at a disadvantage as compared with the incumbent type alleles at this locus. Moreover, this is true whether the focal allele is prosocial or antisocial, so long as it satisfies Hamilton's rule. Therefore there is no intragenomic incentive for genes to evolve to suppress an antisocial allele.

This result must of course be qualified in the diploid case, both because meiotic drive can favor an allele at one locus that harms the other loci in the genome (Haig and Grafen 1991; Burt and Trivers 2006), and males and females may have distinct fitness enhancement conditions based on physiological differences (Haig 2002).

These and related situations aside, we can safely conclude that *even utterly selfish genes have common interests on the intragenomic level*. It follows that suppression of antisocial alleles must be a response to the joint reduced fitness of all alleles *at the society level*, through natural selection. It also follows that a prosocial allele that satisfies Hamilton's rule will provoke suppression responses on neither the intragenomic nor the intergenomic level. Hence prosocial genes *do* maximize inclusive fitness.

9.8 The Boundaries of Inclusive Fitness Maximization

It asserting that "natural selection leads organisms to become adapted as if to maximize their inclusive fitness," Abbot (2011) doubtless expresses a view with which at least 137 of the world's most prominent population biologists appear to agree. The main source cited in support of this statement is a series of papers written by Alan Grafen (1999, 2002, 2006). For instance in a paper devoted to exposing the "misconceptions" of others, West et al. (2011) write:

> Individuals should appear as if they have been designed to maximize their inclusive fitness. Grafen (1999, 2002, 2006a, 2007b) has formalised this link between the process and purpose of adaptation, by showing the mathematical equivalence between the dynamics of gene frequency change and the purpose represented by an optimisation program which uses an "individual as maximising agent" (IMA) analogy.

However, Grafen expressly declares in each of his papers on the subject that *additivity across loci*, or what is equivalent, *frequency independence*, is assumed. Others who have carefully studied the conditions under which a population genetics model of gene flow implies fitness maximization at the gene or individual level, including Metz et al. (2008), Gardner and Welsh

(2011), and Gardner, Welsh and Wild (2011), require the same assumption. No careful researcher has *ever* claimed analytical support for the notion that individuals maximize inclusive fitness without making the frequency independence assumption.

If a gene is prosocial, we have seen that the behavior it fosters can be modeled as the maximization of inclusive fitness. But if the genome's success is based on a pattern of cooperation, promotion, and suppression of antisocial genes across loci, which will occur, for instance, if the production of a protein, RNA sequence, or social behavior requires the collaborative activity of many genes (Noble 2011), or if there are frequency dependent social interactions among individuals in a social species (Maynard Smith 1982), then neither genes nor individuals can be characterized as maximizing inclusive fitness.

9.9 The One Mutation at a Time Principle

Because genes code for proteins or RNA with very precise chemical functions, most mutations are fitness-reducing or fitness-neutral. The rate at which fitness-enhancing mutations occur is very low. Let us say that genes at two loci are *synergistic* if their joint presence in the genome of an individual is fitness-enhancing, but each alone is fitness-reducing. Clearly the rate at which two synergistic mutations occur in an organism is generally orders of magnitude less likely that single favorable mutations. Moreover, even when two such mutations are present, unless they are tightly linked so that they are not broken up by meiosis, they will only rarely and sporadically occur together. Bodmer and Felsenstein (1967) show that synergistic double mutants can survive if $1 < (1-r)w_{14}/w_{44}$, where r is the recombination rate, w_{44} is the fitness of the wild type genome, and w_{14} is the fitness of the same genome with two relevant wild type alleles replaced with the mutants. Thus with no linkage ($r = 1/2$), the mutants would have to be twice as fit as the wild types to evolve. Except in the case of highly improbable macromutations, the linkage rate $1-r$ would have to be very close to unity for the pair of mutants to survive. Moreover, in the case of extremely high linkage, it is a good approximation to treat the two genes as one.

Therefore for most purposes we can assume that *only one favorable mutation occurs at a time*, and its success depends on the frequency distribution of alleles at other loci at the time the mutation appears. We call this the *one-gene-at-a-time principle*.

9.10 The Phenotypic Gambit

The genome of a multicellular organism includes a myriad of interdependent RNA-producing, protein-producing genes and regulatory gene networks. The dynamics of gene interaction are very poorly understood, to the point where it is practically impossible to isolate exactly how a particular gene behaves and interacts with others. Indeed, when we say that a certain allele produces or controls a certain phenotypic trait, what we really mean is that the *absence* of the allele entails the *absence* of the trait. This is, of course, quite a weaker statement, merely asserting that the allele in question contributes in an essential way to the production of the phenotypic effect.

One implication of this state of affairs is that there are few, if any, cases in which a social behavior can be attributed to the choice of an allele at a particular locus of the genome. This fact does not compromise Hamilton's rule, but without additional assumptions, it renders Hamilton's rule inapplicable to analytical models of social behavior without additional assumptions. By far the most widely used such assumption is the so-called *phenotypic gambit* (Grafen 1984). The phenotypic gambit assumes that a behavior that may be extremely complex at the genetic level can be modeled as though it were the product of the choice of allele at a single locus. In the words of Alan Grafen (1984, p. 63),

> The phenotypic gambit is to examine the evolutionary basis of a character as if the very simplest genetic system controlled it: as if there were a haploid locus at which each distinct strategy was represented by a distinct allele, as if the payoff rule gave the number of offspring for each allele, and as if enough mutation occurred to allow each strategy the chance to invade.

The haploid assumption is not necessary—there are many examples in the literature of the phenotypic gambit models behavior as controlled by a diploid locus. Moreover, the assumption of a single locus is not necessary, as there is a research tradition in which the production of a phenotypic effect is controlled by two loci, one of which modulates the effects produced at the other locus (Liberman and Feldman 2005). Two-locus models, however, are generally extremely difficult to model and yield few additional insights.

The one-gene-at-a-time principle, however, often justifies the phenotypic gambit, especially in conjunction with the core genome concept. The latter suggests that most behavior-relevant genes will be either fixed in the genome or exist in such stable form that changes in the frequency of a mutant allele will not appreciably alter the frequency of other relevant genes in the genome. In that situation, the one-gene-at-a-time principle suggests that we are not likely to go wrong by considering an evolving behavior as the effect of an allele substitution at a single locus.

9.11 The Anatomy of the Core Genome

If a gene has no social effects, that is if $b = \alpha = \beta = 0$ in the generalized Hamilton's Rule (equation 9.17), then it obviously evolves only if it is prosocial ($c < 0$), in which case its increase in the population benefits all other loci in the genome. We may call this an *asocial allele*. Moreover, if a gene that evolves is prosocial and non-polluting, it also benefits all genes both in the genome in which it is located, and in the population as a whole. These are strong harmony of interest principles that flow from inclusive fitness theory. But if a gene satisfies the generalized Hamilton's rule but is *antisocial* ($b - c - \beta < 0$) then, as we have seen in Section 9.7, it benefits all its co-resident genes, but it harms the population. Thus natural selection will favor the emergence of social forces that suppress such antisocial genes.

Enter *complexity*, the bitter enemy of classical systems theory. The gene pool of a species, consisting of many copies of long strings of DNA, interact *biochemically* to produce a metazoan organism whose cells manage to cooperate despite the evolutionary interest of each to ignore the others, and which interact *socially* through *emergent structural properties* that suppress defection and enhance cooperation sufficiently to ensure survival. We call these properties "emergent" because in our current state of knowledge, we are no more capable of explaining their provenance that we are in understanding how a sac of chemicals in the skull of a human being can give rise to consciousness.

I call the complex system of gene that gives rise to animal society the *core genome*. The core genome of a sexually reproducing species is a subset of the loci in the genome that includes all loci that have certain key properties ensuring the general phenotypic character of the species. Included in the core genome are the *fixed loci* and *synonymous loci*. The fixed loci are those in which a single allele is shared by all members of the population, except for low-frequency mutations. The synonymous loci consist of loci in which all alleles, except for low-frequency mutations, produce identical biochemical and phenotypic effects. In addition certain non-synonymous alleles may have fitness neutral, or near-neutral, phenotypic effects (e.g., tail length or eye color). The set of such *fitness neutral gene sets* are stable across generations despite their somewhat labile internal composition, and are also part of the core genome. For instance, body size may be fitness independent over some range, and many genes interact to produce a phenotypic body size that is generally in the fitness-neutral range. The frequency distribution of these genes in the core genome is determined by natural selection and unchanged by meiosis and crossover.

In addition, if a set of alleles at a particular locus have equal fitness but distinct phenotypic effects, and if this set is preserved across generations, the alleles are likely to be equally fit alternative strategies in a Nash equilibrium among loci, each being afitness enhancing best response to the probability distribution of the other loci in the genome. We call such alleles *mixed strategy gene sets*, and we include these in the core genome. For example, a population equilibrium can sustain a positive fraction of altruistic and selfish alleles, or alleles promoting aggressive *vs.* docile behavior, under certain conditions. Similarly, loci that protect carriers against frequency-dependent variations in environmental conditions, including that of bacterial and viral enemies, can be maintained in a polyallelic state as a means of species-level risk reduction. These include the *immune system gene sets* that maintain considerable heterogeneity to deal with a variety of possible infectious agents.

Another example of a mixed strategy gene set is the interaction of suppressor genes and their targets, where the fitness of the suppressor depends on a positive frequency of target genes. Leffler (2013) document such a set stabilized by balancing selection at least since the primate-hominin split. Finally, heterozygote advantage involves a pair of alleles that maintain positive frequency despite the fitness cost to homozygous carriers. We may call these *overdominance gene sets*. Additional features arise in dealing with sex-linked genes, including maternal-paternal conflict, but these also can be identified as characteristics of the species that are conserved across many generations.

In species that recognize individuals, including many birds and mammals, such recognition is based in part on the expression of alternative alleles within a core genome gene set, as well as on genes *outside* the core genome, which are shuffled and redistributed through meiosis and recombination, accounting for the heterogeneity of phenotypes.

In sum, the typical phenotypic characteristics of the species, including biochemistry, physiology, and behavioral predispositions, are conserved across generations due to the capacity of the core genome to self-replicate across generations. The non-core genes in the gene pool, largely accounting for the heterogeneity of individuals, may be called the *variant genome*—see Riley and Lizotte-Waniewski (2009) for an application to bacterial species. The core genome is subject to the laws of natural selection: replication, mutation, and selection of superior mutants. Individuals, their societies, and the social structure of these societies, are the product of the evolution of the core genome.

While the core genome is an object of selection, it is not in any sense a *unit* of selection because it is specified by the frequency distribution of genomes in the population. Moreover, the very notion of units and objects of selection, while perhaps of use for a synthetic understanding of biological evolution, do not appear to play any role in modeling the social structure and dynamics of a reproductive population. However, recognizing the core genome as an object of selection is a useful heuristic in at least two ways. First, while not in any way undermining the insights of the gene's eye view of evolution, it captures the notion that precise combinations of gene interactions are adaptive and hence favored by natural selection. Second, the core genome allows us to conceptualize phenotypic effects that are located not in individuals, but in their social interactions. In other words, the core genome strongly predisposes a social species for certain forms of social behavior, including typical mating patterns, recognized forms of territoriality, and preferred forms of social grouping. The core genome also predisposes organisms to seek out particular natural environments, although there is natural variation in such environments that serve as epigenetic sources of social dynamics and social learning (Galef and Laland 2005; Goodnight et al. 2005; Smaldino et al. 2013).

The core genome is a replicator in the sense of Lewontin (1970). First, mutations in loci of the core genome give rise to *phenotypic heterogeneity*. Second, phenotypic differences can entail *fitness differences among members* of the reproductive population. Finally, such fitness differences are *heritable*. A mutation at a fixed locus, for instance, can lead to increased fitness of carriers of the mutated allele, leading to the increase in frequency of the new allele in the population. The focal locus then drops out of the core genome, but in the long run, with high probability, the mutation will either move to fixation or extinction, restoring the focal locus to the core genome.

Richard Dawkins (1982b) is famous for rejecting the genome as an object of selection, arguing that because of meiosis and recombination, the genome dies with the body it inhabits. Dawkins concludes that the individual is but a *vehicle* for the transportation of genes across metazoan bodies, writing that a replicator must have a

> low rate of spontaneous, endogenous change, if the selective advantage of its phenotypic effects is to have any significant evolutionary effect.…too long a piece of chromosome will quantitatively disqualify itself as a potential unit of selection, since it will run too high a risk of being split by crossing over in any generation (p. 47).

Cognizant of this important observation, I have defined the core genome so as to be impervious to meiosis and crossover. This is clear for fixed and synonymous loci, where no breaking up of synergistic genome interactions occur. Moreover meiosis creates as many heterozygote as it destroys, on average, and it does not alter the frequency distribution of mixed strategy or immune system gene sets in the population.

9.12 Explaining Social Structure

While inclusive fitness theory justifies selfish gene theory, neither inclusive fitness theory, nor any other plausible theory, supports the notion that genes or individuals in asocial species maximize inclusive fitness. We have shown that the maximization characterization is plausible for prosocial nonpolluting genes that satisfy Hamilton's rule, but not otherwise.

The evolutionary process, from the first RNA molecules to advanced metazoans and complex social species, involves solving the problem of promoting cooperation among selfish genes (Maynard Smith and Szathmáry 1995). That genes generally contribute to the fitness of the individuals in which they reside is the result, not of inclusive fitness maximization, but of a complex evolutionary and intragenomic dynamic involving the suppression of antisocial and promotion of prosocial alleles (Leigh 1971; Buss 1987; Michod 1997; Frank 2003; Noble 2011).

The evolutionary forces that determine the complex interactions among loci in metazoans and among individuals in social species must be studied using, in addition to inclusive fitness theory, the phenotypic gambit (Grafen

1984), evolutionary game theory (Wilson 1977; Taylor 1992; Taylor 1996), agent-based modeling (Gintis 2009b), the physiology of suppressor and promoter genes (Leigh 1977; Noble 2011), as well as species-level systematics and ecology.

A1 Hamilton's Rule with General Social Interaction

This section presents a version of Hamilton's rule that assumes a diploid organism, and applies to sophisticated social species in which interactions are multi-adic, such as when there is a complex division of labor in hunting, defense, or rearing offspring. The resulting equations are similar to those deduced from the regression approach to Hamilton's rule (Queller 1992) but we have no need for least squares regression arguments. The most salient implication of this exercise is that Hamilton's Rule holds with very great generality, although the three terms in the equation are reflections of the social structure of the reproductive population.

Consider a reproductive population X with individuals $\{X_i \in X | i = 1, \ldots, n\}$. Suppose the genome has a diploid autosomal locus with two alleles, s (selfish) which leads to a behavior that does not affect the fitness of other individuals, and a (altruistic), which leads its carrier X_i to incur an increased fitness cost c_i over that of the selfish allele, and to bestow fitness benefit b_i distributed over a subset Y_i of recipients. Suppose in addition that the altruistic allele has asocial fitness effect β (pollution when $\beta > 0$ or a public good when $\beta < 0$) on both alleles (see Section 9.4). This cost may be intragenomic, borne by the carrier, or intergenomic, distributed over the population in some arbitrary manner.

Hamilton (1964a) assumes the social fitness effect is distributed uniformly over the genome. This is a significant limitation of his analysis because intragenomically, meiotic drive and other forms of segregation distortion, and socially, altruistic acts that are purchased in part by reducing the fitness of non-relatives, which we call *thieving effects* (see Section 9.4), are important, although the Harmony Principle suggests that natural selection will limit their observed frequency. We can represent these thieving effects as transfers of fitness $\alpha > 0$ from non-relatives to relatives, and the reverse for $\alpha < 0$.

Standard expositions of Hamilton's rule take Y_i to be an individual. This, however, is a restrictive assumption because in many social species individuals interact in groups where it is difficult to apportion the benefit b_i among the various participants. For instance, agent i may play in an n-player public goods game in which the s allele promotes defection and the a allele promotes cooperation, or agent i may defend the nest against intruders, or punish a lazy coworker. As we shall see, Hamilton's rule does not depend on the assumption that the beneficiary is an individual.

The genotypic value X_g^i of X_i at the focal locus, the frequency of the focal allele at this locus, is 0, ½, and 1 for genotypes ss, sa, and aa, respectively. The phenotypic value X_p^i of X_i is 0, h, or 1 according as X_i is ss and never confers the benefit, is sa and confers the benefit with intensity h, or is aa and confers the benefit with intensity one. Here h can have any value, positive or negative, but if the allele effects are additive, then $h = 1/2$. Because there are $2n$ alleles at the focal locus in the population, the frequency of a is $q_a = \sum_i X_g^i/n$. Let Y_g^i be the mean genotype of members of Y_i.

The fitness cost to X_i in the current period is thus $c_i X_p^i$, and the fitness gain to the recipients Y_i is $b_i X_p^i$. The population size in the next period is then

$$n(1 - \beta q_a + (b - c)x_p) \tag{9.1}$$

where $x_p = \sum_i X_p^i/n$ is the mean phenotype of the population, $b = \sum_i b_i X_p^i/x_p$ is the mean benefit, and $c = \sum_i c_i X_p^i/x_p$ is the mean cost. Note that because the thieving effect α is a within-population fitness transfer, it does not appear in (9.1). The number of donor alleles in the next period is

$$n q_a(1 - \beta q_a + \alpha(1 - q_a)) + \sum_i b_i X_p^i Y_g^i - \sum_i c_i X_p^i X_g^i.$$

The increase in the frequency of the donor allele in the next period, writing the mean genotype of recipients as $q_a^y = \sum_i Y_g^i/n$, is then given by

$$\frac{n q_a(1 - \beta q_a + \alpha(1 - q_a)) + \sum_i b_i X_p^i Y_g^i - \sum_i c_i X_p^i X_g^i}{n(1 - \beta q_a + (b - c)x_p)} - q_a =$$

$$\begin{aligned}&\frac{\left(\sum_i b_i X_p^i Y_g^i - n b x_p q_a^y\right) + n q_a \alpha (1-q_a)}{n(1-\beta q_a + (b-c)x_p)} - \\ &\qquad \frac{\left(\sum_i c_i X_p^i X_g^i - n c x_p q_a\right) + n b x_p (q_a - q_a^y)}{n(1-\beta q_a + (b-c)x_p)} = \\ &\frac{\mathrm{cov}(X_p^b, Y_g) - \mathrm{cov}(X_p^c, X_g) + \alpha \mathrm{var}(X_p) + b x_p (q_a^y - q_a)}{1-\beta q_a + (b-c)x_p}, \qquad (9.2)\end{aligned}$$

where X_p^b and X_p^c are the variables $b_i X_p^i$ and $c_i X_p^i$, respectively, and X_g is a binomial variable, so $\mathrm{var}(X_p) = n q_a (1 - q_a)$. Note that the expression (9.2) is positive, assuming weak selection, when

$$\frac{\mathrm{cov}(X_p^b, Y_g) + \alpha \mathrm{var}(X_p) + b x_p (q_a^y - q_a)}{\mathrm{cov}(X_p^c, X_g)} > 1. \qquad (9.3)$$

This inequality is the most general form of Hamilton's rule, including both social fitness and thieving effects. If we assume donors distribute benefits that are, on average, independent from the allelic composition at the focal locus, i.e., $q_a^y = q_a$ then (9.3) becomes

$$\mathrm{cov}(X_p^b, Y_g) + \alpha \mathrm{var}(X_p) > \mathrm{cov}(X_p^c, X_g). \qquad (9.4)$$

Note that in the standard treatment, where the beneficiary is an individual, the condition $q_a^y = q_a$ necessarily holds. To see this, note that

$$q_a^y = [r + (1-r)q_a]q_a + [1-(r+(1-r)(1-q_a))](1-q_a) = q_a, \quad (9.5)$$

where r is the relatedness coefficient.

If we further assume that $b_i = b$ and $c_i = c$ for all individuals $i = 1, \ldots, n$, we get the expression:

$$\frac{b\,\mathrm{cov}(X_p, Y_g) + \alpha\ \mathrm{var}(X_p)}{\mathrm{cov}(X_p, X_g)} > c. \qquad (9.6)$$

Finally, if the effect of the altruistic allele is additive, so $h = 1/2$, then (9.6) becomes

$$b\,\frac{\mathrm{cov}(X_p, Y_g)}{\mathrm{var}(X_g)} > c - \alpha. \qquad (9.7)$$

This is a standard expression for Hamilton's rule (Michod and Hamilton 1980), except we have taken into account the thieving effect α (and the pollution/public good effect β, which does not appear in Hamilton's rule). More generally, for arbitrary h, we have

$$br > cr^p - \alpha, \tag{9.8}$$

where

$$r = \frac{\mathrm{cov}(X_p, Y_g)}{\mathrm{var}(X_g)}$$

is the regression coefficient of Y_g on X_p, and r^p is the regression coefficient of X_p on X_g:

$$r^p = \frac{\mathrm{cov}(X_p, X_g)}{\mathrm{var}(X_g)}.$$

It should be clear that, while we use mathematical terminology from statistical estimation theory, no statistical estimation is in fact involved.

To illustrate the increased generality of the form (9.4) of Hamilton's rule, suppose the reproductive population is partitioned into *social castes* $\{Z^j \subset X | j = 1, \ldots, m\}$, where caste j has frequency z_j in the population, and suppose members of the same caste j have the same costs c_j and benefits b_j. Let Y^j be the weighted sum of $\{Y_i | X_i \in Z^j\}$, where each individual is weighted by the number of times the individual appears in the sum. Then we can write (9.4) as

$$\sum_{j=1}^{m} \left((b_j \,\mathrm{cov}(Z^j_p, Y^j_g) - c_j \,\mathrm{cov}(Z^j_p, Z^j_g)\right) + \alpha \,\mathrm{var}(X_p) > 0. \tag{9.9}$$

Equation (9.9) shows that in general the social structure of the population allows a caste to be *fundamentally altruistic* in the sense that its net costs of helping exceed the net benefits that the caste contributes to the population. Because the inclusive fitness of caste j is

$$b_j \,\mathrm{cov}(Z^j_p, Y^j_g) - c_j \,\mathrm{cov}(Z^j_p, Z^j_g) < 0 \tag{9.10}$$

it is then clear that caste j members would maximize their inclusive fitness by simply refusing to contribute to the social process. This shows that *in a caste social structure, individuals do not necessarily maximize their inclusive fitness*. Of course, if castes are genetically determined, then the partition $\{z_j | j = 1, \ldots, m\}$ will be variable across periods and a fundamentally altruistic caste will become extinct in the long run. However, if castes are determined by developmental conditions (e.g., feeding in eusocial insects or socialization in humans), fundamentally altruistic castes can be maintained in the long run.

A1.1 The Sociobiological Dynamics of Hamilton's Rule

The mapping $X_i \to Y_i$, which we have taken as given, reflects the *social structure of the reproductive population*. This mapping does not presume any particular set of social relations of kinship, which is why we suggest that *kin selection* is in general an inappropriate description of inclusive fitness dynamics. Note that if the frequency of the a allele in the population does not affect the fitnesses of alleles at other loci in the genome, then the a allele will move to fixation in the population if Hamilton's rule is satisfied, and will become extinct if the reverse inequality is satisfied. Ultimately, the focal locus will be heterozygous with zero probability.

With frequency dependence, when the focal allele becomes prevalent in the population, if $b - c > 0$, so the allele is beneficial to its carriers, there will be no selection at the level of the genome for genes that suppress the a allele at the focal locus, so the a allele will still move to fixation in the population. When the focal allele is prevalent and $b - c < 0$, there will be natural selection at other loci for genes that either alter the sociobiological mapping $X_i \to Y_i$ or otherwise suppress the a allele at the focal locus, so that Hamilton's rule no longer holds for the antisocial allele. This is the essence of the Inclusive Fitness Harmony Principle. Of course there may be no likely mutation that suppresses an anti-social a allele, in which case the antisociality reflected in the behavior induced by the a allele will become ubiquitous in the population. natural selection does not guarantee optimality.

This phenomenon also represents a plausible counterexample to Fisher's Fundamental Theorem (Ewens 1969; Price 1972; Frank and Slatkin 1992; Edwards 1994; Frank 1997): as an antisocial allele moves to fixation, the average fitness of population members declines. Some population biologists save Fisher's theorem by calling this a *transmission effect*, and insisting that natural selection always produces fitness-enhancing gene frequency changes (Edwards 1994; Frank 1997; Gardner et al. 2011). This interpretation of natural selection should be avoided because it is arbitrary and difficult to understand for those who are not experts in population biology.

It follows that Hamilton's rule is useful only in charting short-term genetic dynamics. Weak selection and additivity across loci are extremely powerful analytical tools, but in the long run changes in gene frequency at one locus are likely to induce compensatory and synergistic changes at other loci. Indeed, the very mapping $X_i \to Y_i$ on which Hamilton's rule is based is itself coded in the core genome of the reproductive population, and hence in the long run is modified in the course of evolutionary selection and adaptation.

A1.2 Altruism Among Relatives

A relative is a person "allied by blood... a kinsman" (Biology Online). The argument to this point has nothing to do with genealogy, and hence says nothing about altruism among family members. This is an attractive property of our exposition because in a highly social species, individuals interact frequently with non-relatives.

It remains to determine the exact relationship between the sociobiological conception (9.6) and the genealogical conception of relatedness. We follow Michod and Hamilton (1980), except that we assume the population is outbred at the focal locus. Suppose that each Y_i is an individual recipient, and all recipients have the same genealogical relationship to their donors (e.g., Y_i is a sibling of X_i). Let $\{p_{xyzw}\}$ be the joint distribution of genotypes xy for donor and zw for recipient where $x, y, z, w \in \{s, a\}$. Let p^x_{ss}, p^x_{as}, and p^x_{aa} be the marginal distribution of the genotypes ss, sa, and aa for the donor (i.e., the fraction of these genotypes in the population), and similarly for p^y_{ss}, p^y_{as}, and p^y_{aa} for the recipient.

We have

$$x_p = hp^x_{as} + p^x_{aa},$$
$$y_p = hp^y_{as} + p^y_{aa},$$

because p^x_{as} is the fraction of sa genotypes, their phenotypic value is h, and p_{aa} is the fraction of aa genotypes, which have phenotypic value one. Also,

$$p^x_{as} = 2q_n q_a \tag{9.11}$$
$$p^x_{aa} = q^2_a \tag{9.12}$$

To derive (9.11), note that either the paternal allele is s with probability $q_n = 1 - q_a$ and the second is a with probability q_a, or else the paternal allele is a with probability q_a and the second is s with probability q_n. The second equation is derived in a similar manner.

We thus have

$$x_p = 2hq_n q_a + q^2_a \tag{9.13}$$
$$y_p = 2hq_n q_a + q^2_a \tag{9.14}$$

Note that

$$\begin{aligned} x_g &= 1/2 p^x_{as} + p^x_{aa} = q_a \\ y_g &= 1/2 p^y_{as} + p^y_{aa} = q_a. \end{aligned}$$

To derive $\mathrm{cov}(X_g, X_p)$, note that

$$\begin{aligned} \sum_i X^i_p X^i_g / n &= h p^x_{as}/2 + p^x_{aa} \\ &= h q_n q_a + q_a^2 \end{aligned}$$

Given the values of p^x_{as} and p^x_{aa} from equations (9.11) and (9.12), and after algebraic simplification, we find

$$\mathrm{cov}(X_p, X_g) = q_n q_a \gamma/2, \tag{9.15}$$

where

$$\gamma = 2(h + q_a(1 - 2h)). \tag{9.16}$$

Also,

$$\mathrm{cov}(y_g x_p) = h p_{sasa}/2 + h p_{saaa} + p_{aasa}/2 + p_{aaaa} - y_g x_p.$$

Now let p_{11} be the probability X_i and Y_i share both alleles at the focal locus identically by descent, let p_{10} be the probability the share one allele at the focal locus identically by descent, and let p_{00} be the probability they share neither allele identically by descent. then we have

$$\begin{aligned} p_{asas} &= 2q_n q_a p_{11} + q_n q_a p_{10} + 4q_n^2 q_a^2 p_{00} && (9.17) \\ p_{asaa} &= q_a q_n^2 p_{10} + 2q_n q_a^3 p_{00} && (9.18) \\ p_{aaas} &= q_n q_a^2 p_{10} + 2q_n q_a^3 p_{00} && (9.19) \\ p_{aaaa} &= q_a^2 p_{11} + q_a^3 p_{10} + q_a^4 p_{00}. && (9.20) \end{aligned}$$

If we define f_{XY} as the probability that a random allele in X_i and a random allele in Y_i are identical by descent, then

$$f_{XY} = p_{11}/2 + p_{10}/4. \tag{9.21}$$

Then a little algebra shows that the r in Hamilton's rule is given by

$$r = \frac{\mathrm{cov}(X_p, Y_g)}{\mathrm{cov}(X_p, X_g)} = 2 f_{XY}. \tag{9.22}$$

Note that r is then the expected number of copies of the focal allele in the recipient.

Consider, for instance, the case of siblings. The two share the same allele from the father with probability 1/2, and similarly for the mother. therefore $p_{11} = 1/4$, $p_{10} = 1/2$, and $p_{00} = 1/4$. Substituting these values in (9.17), we get

$$r = \frac{\mathrm{cov}(Y_g, X_p)}{\mathrm{cov}(X_g, X_p)} = \frac{1}{2}. \tag{9.23}$$

Thus the sociobiological definition of relatedness and the genealogical definition coincide.