Metadata of the chapter that will be visualized online

Chapter Title	Hamilton’s Rule and Theoretical Implications	
Copyright Year	2016	
Copyright Holder	Springer International Publishing Switzerland	
Corresponding Author	Family Name	**King**
	Particle	
	Given Name	**Robert**
	Suffix	
	Division/Department	School of Applied Psychology
	Organization/University	University College Cork
	City	Cork
	Country	Ireland
	Email	r.king@ucc.ie

H

Hamilton's Rule and Theoretical Implications

Robert King
School of Applied Psychology, University
College Cork, Cork, Ireland

Definition

Hamilton's rule is that an altruistic behavior can be selected for in a population under the circumstances that

1. The behavior is heritable (variance explained by genetic difference).
2. The gene underlying it provides a benefit to those who share that gene by common descent that is higher than the cost exerted, multiplied by the coefficient of relatedness.

This is usually simplified to $r B - C > 0$. In this formulation r is the relatedness coefficient between actor and beneficiary of behavior; B is the reproductive benefit provided to the recipient; and C is the cost to the actor in terms of direct reproduction.

Introduction

The term "paradigm shift" should come with a health warning. Hamilton's rule, formalizing inclusive fitness, is one of those very few developments in science that genuinely deserves the accolade. Inclusive fitness is a central, axiomatic concept in evolutionary biology. Darwin's discoveries, concerning descent with modification from common ancestry directed by natural selection, were the result of years of painstaking observation and the synthesis of vast amounts of empirical data. Hamilton's rule is an extension of Darwin's insight, based on pure deductive reasoning as laid out in (Hamilton 1964) and further developed in papers of equal mathematical sophistication (e.g., Price 1972).

While the complexities of these original papers are rarely directly engaged with, the take-home message seems simple: Namely, that altruistic behavior can be selected for just in case that the benefit bestowed on the recipient (B) multiplied by the coefficient of relatedness between actor and recipient (r) minus the cost to actor (C) is greater than zero.

This is typically expressed as $r B - C < 0$.

Rarely has such a deceptively simple formulation had such profound consequences or provoked such large misconceptions and fights over interpretation. It has been argued that the rule as it stands is too simple to permit simple predictions on the basis of it (Frank 1998). Whether or not this is true, it has been tempting for scholars to rush to predictions based on it, perhaps in a version of physics envy. The upshot of this haste can then be that, following a supposed failure of Hamilton's rule to apply, scholars seek for other explanations

T.K. Shackelford, V.A. Weekes-Shackelford (eds.), *Encyclopedia of Evolutionary Psychological Science*,
DOI 10.1007/978-3-319-16999-6_1488-1

for the source of a social behavior. In this vein it is worth emphasizing that there are a large number of things that Hamilton's rule does not imply and does not apply to – much though it may appear to.

Why does all this matter? It is not hyperbole to say that Hamilton's rule explains the otherwise miraculous. Miracles are, strictly speaking, things that cannot be explained by appeal to natural laws. Darwin's insight explains how the world appears to be designed but without needing a designer. Hamilton's extension of Darwin's insight is no less momentous. It explains how moral behavior – which at its bedrock – requires the capacity to benefit others at a net cost to oneself (in other words, true altruism) can come into the world without a divinity to underwrite it.

Throughout history, humans have typically sought for supernatural explanations for the way that the universe contains both beauty and goodness. As Kant famously put it "Two things awe me most, the *starry sky* above me and the *moral law* within me." Darwin's discoveries showed us that no designer was needed to create functionality and in the process reminded us that not all functionality was beautiful. Hamilton's rule unites all of nature in terms of how genuine altruism – a crucial social behavior – can exist at all without supernatural interference. In the process he similarly showed us that our intuitions about what is truly good cannot be relied on. Of course, this mathematical extension of evolution by natural selection has far more implications than simply that it is the most general formulation of natural selection yet devised.

Useful Terms

Au3

Adaptation: A trait that improves fitness – defined in terms of representation of genes in the next generation. Since Darwin and the modern synthesis, the only non-supernatural force that explains the appearance of design in this fashion is natural selection, although other forces (such as drift) can explain differential representation of genes.

Altruism: A behavior that imposes a cost on the actor and gives a benefit to the recipient. The word is usually applied to behaviors, but any trait could be altruistic – such as a physical trait that acts to benefit others at a cost to the user like a honeybee's sacrificial sting.

Gene: The basic unit of selection. Whatever has the requisite properties of longevity, fecundity, and fidelity is a gene in the sense needed for evolutionary biology. Since the double-helix nature of DNA has been uncovered (which has those three required properties), this has become the focus of research. However, interesting complexities such as the various kinds of interactions between genes make judgements that rely on a single "gene for x" potentially misleading. A useful way to think of genes for evolutionary biology purposes is as a catalyst whose catalyzing reactions influence its representation in the next generation (Dawkins 1976; Haig 1997).

Green beards: A putative tightly aligned property linking genes to phenotype that would allow them to recognize one another directly. It is controversial whether any genes do this, but no one claims that they do in the case of humans. However, humans do possess adaptations allowing them to recognize kin at better-than-chance levels.

Group selection: There are a number of meanings for this term, and not all are mutually consistent. The original use of the term, by Wynne-Edwards, to refer to voluntary limitation of fitness-producing behaviors so as to benefit the group has fallen out of favor as being shown to suffer from being fatally vulnerable to selfish invaders. While other uses do persist, there is to date no additional explanatory power that has been shown to be dependent on modeling social behavior in this way, rather than in terms of existing mechanisms such as inclusive fitness and mutualism. Group selection in the way precisely defined by Price (1972) can exist but only in extreme situations that do not pertain to human beings (such as groups budding and reproducing faster than their constituent elements do). That a group persists or expands is not itself an instance of group selection because the group is not, in this case, the unit of selection. Humans have a number of adaptations to groupishness, however, and these are important.

Heritability: The proportion of variance in a trait accounted for by genetic factors. It is important to note that heritability is not a fixed property. For example, as shared environments become more similar, then heritability increases. Heritability is sometimes confused with "genetic determinism" which is a largely meaningless phrase. The development of human heads can be safely said to be encoded in genes. The actual possession of a head has a heritability of 0, because variance in the number of heads possessed by a specific human (0 or 1) will be entirely explicable in terms of environmental factors.

Inclusive fitness: Individuals can affect the transmission of their genes into the next generation either directly (their own reproduction) or indirectly (that of relatives). Inclusive fitness is an extension of Darwin's principle of evolution by natural selection so as to include social behaviors – indirect as well as direct benefits. It is inclusive fitness that Darwinian individuals can be assumed to maximize – making it the most generally applicable model for explaining adaptation yet formulated. Inclusive fitness can be measured through measuring the effect on offspring in general (direct and indirect) multiplied by the degree of relatedness. As such it is foundational and true by deduction once one accepts that evolution is through differential survival of genes in the gene pool. Attempts to disprove it are thus misguided, unless one first attempts to disprove the concept of fitness or the concept of genes. Hamilton's rule for inclusive fitness is often simplified to $r\,B - C > 0$, where r is the coefficient of relatedness, B is the benefit in terms of fitness, and C is the cost to the actor.

Kin selection: A term coined by Maynard Smith (1964) to explain the indirect fitness benefits that accrue from aiding kin reproduction and to distinguish this from group selection. In one sense kin selection refers to relatedness due to common descent. However, a broader use of the term refers to the degree of shared genes at particular loci – whether or not these happened to come from shared ancestry. However, given that green-beard effects (where genes can recognize copies of themselves) are likely to be rare, the differences between these uses are unlikely to matter much in humans. Hamilton does not use the term in his writings. Inclusive fitness (unlike kin selection) does not require actual kinship, just genetically nonrandom altruism (Hamilton 1975). This could occur through situations of comparatively low dispersal, for example.

Mutualism: Behaviors that provide fitness benefits (not necessarily equal) to both actors and recipients. It is easy to mistake mutualistic behaviors for altruistic ones.

Relatedness: Although this is commonly thought to refer to shared genes, this is a simplification with consequences. A much better definition is to put relatedness in terms of the degree of genetic similarity between individuals related to the average background shared genetic similarity. See Box 1 (taken from West et al. 2011 and used with permission) for a more complete description.

Selfish: Behaviors that benefit the actor but impose a cost on the recipient.

Social behavior: Any behavior that has consequences for other individuals (Hamilton 1964). Mere presence may have consequences but it is not a behavior, per se. There are four types of consequence that can occur: altruistic, mutual, selfish, and spiteful.

Spiteful: Behaviors that impose costs on both recipient and actor. One of the strengths of Hamilton's (1970) formulation of social behaviors is its successful prediction of spiteful behaviors in cases where relatedness coefficients are negative.

Implications of Hamilton's Rule

Good theory both explains and predicts. It explains what we see and frames what kinds of questions we can ask of our observations. And, it makes (one hopes) surprising predictions about things that we might see and the things that, despite persistent search, we don't see. The more surprising the prediction, the more confident we are of the theory if it is confirmed. Darwinian evolution by natural selection explains the appearance of design in the natural world and places constraints on the sorts of traits that can exist by limiting the ways that they can come to exist. For example, Darwin could famously predict the existence of a particular kind of moth and its traits prior to its discovery, based on observations of a deep flower with nectar at its base.

Hamilton's rule makes similar surprising predictions and offers constraints on the sorts of social behaviors that can evolve. Its predictions are somewhat more complex to follow than natural selection alone, however. The rule also makes predictions, but often not the ones attributed to it.

Inclusive fitness implies that organisms can be assumed to be acting so as to maximize their average lifetime fitness (even if individuals may deviate from this). Hamilton's extension of the Darwinian insight was to realize that this included not just the fitness of the individual themselves but also of their relatives. Social behavior involves more than one entity, of course. For ease let's call them an actor and a recipient. Either one, both, or neither can benefit from the interaction. This gives us four possibilities. If both benefit, this is mutualism. If neither benefits, this is spite. If the actor gains at the expense of the recipient, then they are truly selfish (in the technical sense of the word – they may or may not have selfish motives), whereas if the actor loses out but the recipient benefits, this is altruism (West et al. 2007). Once again, "altruistic" here means what it means in a technical biological sense. For everyday usage behaviors that are mutualisms (like both parents giving loving care to a child) might be felt to be altruistic when they are mutually beneficial. It is important not to be misled (by our natural tendency to be hypervigilant for our fellow humans' potential to be exploiting us) into thinking that these behaviors are somehow not really altruistic.

Questions That Biology Can Answer

Tinbergen (1963) helpfully defined the four types of biological questions that can be asked of a trait.

1. How does the trait contribute to fitness? (Evolutionary question.)
2. How does it function? (Mechanistic question.)
3. How does the trait develop? (Ontogenic question.)
4. How did the trait evolve? (Phylogenetic question.)

There is resemblance to Aristotle's four causes (material, formal, efficient, and final) with some scholars insisting that final causes (teleology) correspond to the evolutionary explanation of a trait, and, possibly, Darwin himself may have flirted with this idea. However, teleology is end directed and evolutionary fitness is something that can only be seen in retrospect. Other than as a shorthand (i.e., organisms behave as if they are trying to maximize their fitness), the resemblance of evolutionary causation to teleology is misleading. Organisms are driven neither by an inner elan vital nor pulled by an external divine plan.

A useful general distinction following from Tinbergen (1963) is that ultimate answers are given to "why" questions, whereas proximate answers are given to "how" questions. Thus, in answer to a question about how eyes develop, a proximate answer might look at the development of eyes and ask questions such as "do neonates see color?" On the other hand, a question that asked "why do humans discriminate red and green" might make reference to our phylogenetic history and how the ability of our ancestors to discriminate ripe fruits from unripe ones increased their fitness. These would be ultimate questions.

Information and Price's Formulation of Altruism

The general application of game theory to evolutionary problems begins with the work of Smith and Price (1973) which showed how limited conflict between animals could be modeled without assuming some benefit to the species model. These insights depended in turn on Price's (1995 posthumous) formulation of Hamilton's rule. Here he realized that any useful mathematical definition of selection is needed to exclude "psychological factors of preferences and decision making" (p. 389).

This is important for a number of reasons, but one of them is that human beings – hypervigilant as they are to signs of possible fakery and betrayal in acts of apparent altruism – can be led astray by the notion that all altruism is somehow not real – by which they usually mean that the appropriate feelings associated with it may be absent in a particular case. It may well have made a lot of evolutionary sense for our ancestors to test one another in the group for the presence or absence of particular moral sentiments. Indeed, we likely still do this in the form of gossip and similar behaviors. However, the proximate moral emotions of, for instance, empathy or shame, which mediate social behaviors, are not to be confused with the ultimate causes of how the gene underlying altruistic behaviors might be selected for by evolution.

Price (1995) makes an explicit analogy with how Hartley's (1928) definition of information, which made no reference to meaningfulness, was foundational to Shannon's (1948) insights into information theory. Hartley's (1928) definition of a practical measure of information was in terms of the logarithm of the number of possible symbol sequences. It is a striking fact, not lost on contemporary physicists, that this definition of information:

$$W = K \log m$$

(where W is the speed of transmission of information, K is a constant – to be empirically determined – and m is the range of voltage levels to choose from in the signaling system)

Has so much in common with Ludwig Boltzmann's formulation of entropy

$$S = k_B \log W$$

(where S is the entropy of an ideal gas, k_B is Boltzmann's constant – an empirically determined number – and W is the number of microstates in that system).

The fact that energy and information can be put in terms of one another should alert scholars to the fact that the information referred to here does not require psychological meaning. This is important because many might otherwise assume that information requires some sort of irreducible semantic content – e.g., a mind that understands it – and it does not. In the same way, altruism (and aggression for that matter) can be modeled without any necessary ascription of proximate mechanisms by which they are manifested (such as loving feelings).

Genes, Selfish, and Otherwise

Genes can be helpfully thought of as a catalyst whose catalyzing reactions affect their own representation in the next generation (Haig 1997). Another way to think of this is that they are as replicators that build vehicles through which they interact with the world, including one another (Dawkins 1999).

Game theory has become central to modeling these complex interactions of vehicles, providing testable and often surprising predictions (Smith and Price 1973). Gene frequency can be held in various kinds of dynamic equilibrium – helpfully referred to as evolutionarily stable strategies (ESS). This is a term borrowed from game theory to describe a set of strategies adopted by actors in a population that is stable. By "stable" it is meant that an invading (and rare) alternative strategy cannot invade and become dominant.

To a first approximation, an individual's gene's success is synonymous with her own. However, this isn't necessarily true when the fitness of relatives who share particular genes by common descent is factored in, and this is when Hamilton's (1964) insight comes into play. At this point a suite of possibilities for modeling and predicting behaviors opens up. Of course, analysis may not stop there. As Haig (1997) has pointed out, the genes may well be in conflict with one another at the level of expression within the individual as well. Indeed the latter provides a powerful test of how powerful the "selfish-gene" model really is (Dawkins 1976). To illustrate this consider the conditions of Angelman and Prader-Willi syndromes – both of which are usually considered as developmental abnormalities.

Unless genes are taken to be "selfish," i.e., that they seek representation in the next generation even at the (possible) expense of their hosts, then a number of phenomena are inexplicable. A good example would be maternal/paternal drives for gene expression – genomic imprinting. Here, some proximate mechanism (typically methylation) causes some genes to prosper (be expressed) at the expense of others. Some genes know if they are derived from the father or the mother. A classic example would be genes on chromosome 15 which code for the growth of the hypothalamus (Buiting et al. 1995). The father's genes would benefit – i.e., maximize their representation – from an offspring which demands a lot from the mother, and the genes from his line try to force expression at the expense of the mother's genes. At the same time, her gene's interests (and her own) would be served by hedging her bets and not investing all in one highly demanding offspring.

Normally the conflict of these genes is held in dynamic tension – neither set winning out. However, the evidence that they reached this impasse through conflict lies in the conditions that (rarely) result if one set does happen to win out over the other. If the father's genes win, then the baby develops Angelman syndrome (and is highly demanding), whereas if the mother's win, Prader-Willi syndrome (also known as "happy puppet syndrome" for the relative undemandingness of the baby) is the result. If the fight for expression was silenced, then the offspring would be perfectly viable (Moore and Haig 1991). Therefore, the existence of the syndromes (which occur when the mechanisms are not in dynamic equilibrium) constitute evidence for selfish gene theory. They are the classic "signs of a struggle" that detectives see when they enter a crime scene; nothing else explains the syndromes in question. Note that these are not cooperative strategies because the more efficient ones (where both sides drop their weapons) are not evolutionarily stable strategies.

However, these examples are not per se social interactions, and at the level Hamilton's rule operates, it is social interactions that are the normal focus of attention.

Cooperation and the Major Biological Transitions

It might be thought that evolution by natural selection implies universal conflict. It is true that, without a struggle for resources, there is nothing that counts as outcompeting others. However, the means of achieving this is frequently cooperative in nature. Selfishness of genes does not imply that they cannot cooperate in many ways to achieve their goals. Cooperation is ubiquitous in nature. One way to see this is in terms of the eight major transitions that have occurred, each increasing the level of complexity and requiring a cooperative mechanism to do so (Smith and Szathmary 1997).

(1) Replicating molecules	=> Populations of joined molecules
(2) Independent replicators	=> Chromosomes
(3) RNA (gene and enzyme)	=> DNA and protein (genetic code)
(4) Prokaryotes	=> Eukaryotes (cells with nucleus and organelles)
(5) Asexual clones	=> Sexual reproduction
(6) Protists	=> Multicellular organisms with organs
(7) Solitary individuals	=> Colonies with sterile castes
(8) Primate societies	=> Human societies with language

Some might argue that cultural evolution also belongs in this line as the next step, but the majority view is that cultural evolution represents a separate process rather than a biological transition per se (West et al. 2011).

Direct and Indirect Fitness

Fisher (1930) made a crucial distinction between direct and indirect fitness. To increase direct fitness is to increase the representation of the actor's genes in the next generation. It is also possible to have fitness modulated by the behaviors of neighbors. For a variety of obvious reasons, neighbors are more likely to be kin and this is indirect fitness.

Relatedness

Organisms have traits, and those traits can be quantified. The trait (phenotype) can be meaningfully separated into the heritable component and environmental component. Natural selection acts upon genes and changes the average value of the phenotypic quantity of interest in populations (not individuals) as Fisher (1930) showed. This is how increases in fitness are defined. Furthermore, Fisher (1930) usefully separated such fitness increases into direct and indirect effects – the latter including improving the fitness of kin (see Fig. 1).

One of Hamilton's key insights was to realize that shared-gene underlying altruism could exist linearly as well as vertically in a population. "There is nothing special about the parent-offspring relationship except its close degree and a certain fundamental asymmetry" (Hamilton 1964, pp. 1–2). In Hamilton (1970) kinship by common descent can be calculated from shared genealogy. Subsequently it was appreciated that the direction of selection for social behaviors could be driven by appropriate statistical associations between individuals (Hamilton 1972; Price 1970).

Hamilton's early work takes an expressly population genetic approach – for example, he showed that the net effect of allele frequency on related individuals can be expressed in terms of the fitness effects on those partners in relation to their degree of relatedness. This is exactly what r B – C >0 implies. Later work (Queller 1992) showed that this is a special case of a more general model of covariant traits in a socially linked population. This later work fits Hamilton's insights into a quantitative genetics model but should not be seen as superseding these insights – rather as extending them to even further generalizability.

The "r" refers to the probability that a particular allele is shared through common descent. It does not imply anything about the proportion of shared genes in common. Humans share over 99 % of their genes – r measures the greater similarity between relatives above the background similarity between members of the same species.

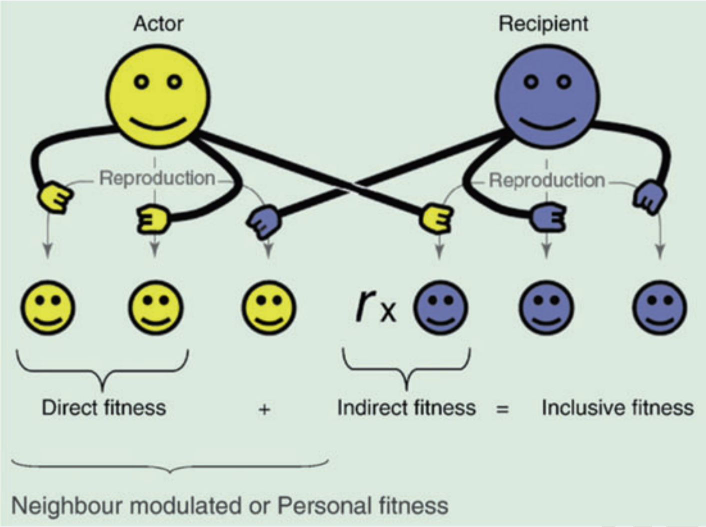
An example of r: Let us assume the frequency of a particular allele in humans is 0.9. If I have a brother, then he is related to me with a coefficient of relatedness of 0.5. For us as a pair of siblings, half the time my brother will get the allele in question from the same parental chromosome as I did and half the time from the other parental chromosome. Therefore the probability of him sharing the particular allele with me is $\frac{1+0.9}{2} = 0.95$. This is (intuitively) halfway between the background frequency of the allele and 1. It is possible that using the concept of gene frequency rather than fitness might forestall confusion.

Families and Tribes

For parents to invest in offspring is species typical across taxa, with the types of investment varying according to a host of factors. In humans it might be thought that, as human children are so labor-intensive, investment would be total and preclude other behaviors. However, there is plenty of scope for the interests of parents and offspring to diverge somewhat. Parent-offspring conflict is a theme developed by Trivers (1974). The patterns of investment are predicted to alter in accordance with the possibility of investing in further offspring and viability of the offspring in question among other things.

Altruism toward families, both immediate and extended, requires some mechanisms, and Hamilton proposed that both kin-recognition mechanisms and the natural result of populations with a high viscosity could be the local means by which these occur. The various forms of potential markers of kin recognition are a rich field of enquiry. It could involve such markers as facial features, smell, dialects, skin tone, and proximate markers of tribal allegiance. Viscous populations are those where dispersal is sufficiently low that proximate mechanisms that treat mere proximity as sufficient to cue kinship can be selected for. These will form part of what Dawkins (1999) calls an extended phenotype, and, once again, this is a rich source of potential enquiry.

It is even possible that otherwise hard to explain behaviors such as homosexual preference (as distinct from homosexual behaviors which are far more common across taxa) might be explained in terms of inclusive fitness. If so-called helpers at the nest effects could occur with humans with the benefits conferred offsetting the costs involved,



Hamilton's Rule and Theoretical Implications, Fig. 1 Inclusive fitness is the sum of direct and indirect fitness (Hamilton 1964). Social behaviors affect the reproductive success of self and others. The impact of the actor's behavior (*yellow hands*) on its reproductive success (*yellow offspring*) is the direct fitness effect. The impact of the actor's behavior (*yellow hands*) on the reproductive success of social partners (*blue offspring*), weighted by the relatedness of the actor to the recipient, is the indirect fitness effect. In particular, inclusive fitness does not include all of the reproductive success of relatives (*blue offspring*), only that which is due to the behavior of the actor (*yellow hands*). Also, inclusive fitness does not include all of the reproductive success of the actor (*yellow offspring*), only that which is due to its own behavior (*yellow hands*; adapted from West et al. 2007a). A key feature of inclusive fitness is that, as defined, it describes the components of reproductive success which an actor can influence and therefore which they could be appearing to maximize (Reprinted from West et al. (2011) with permission)

then this could explain the development and preference of such behaviors. No one has, to date, convincingly shown this to happen without other assumptions being built in to the models as yet

There is a natural consequence of favoring one's kin and tribe and that is potential disfavoring of those who are not kin or who might be threats to resources. Indeed, this is what we find. Not only is in-group favoritism species typical in humans, but there is also a marked Cinderella effect – with those who are not one's offspring – but are still making demands on one's parental care, being at increased risk of neglect or even violence (Daly and Wilson 1998).

Strong Reciprocity and Economic Games

Humans do not just help one another; they also punish those that do not help. This calls for an explanation because said punishment is costly. Evidence for these moral dispositions – such as pride, envy, spite, shame, helpfulness, and similar – can be found cross-culturally and even sometimes in prelinguistic neonates favoring puppets who display such tendencies. A large literature – too large to fully review here – involving economic games, has developed to demonstrate these tendencies across human populations (see West et al. 2011, for an extended discussion of economic games in this context). This tendency to punish transgressors and free riders in a community is sometimes referred to as "strong reciprocity."

It is especially important in considering models of such behavior to distinguish proximate mechanisms, such as feelings of moral outrage at transgressors, from ultimate causes for the selection of such mechanisms, i.e., how they might have contributed to fitness (Mayr 1961). Such selection may be in terms of altruism sensu stricto, but it is at least as likely to be evidence of a mutualistic system. While it is important that we can model human behavior in laboratory, or quasi-laboratory settings, care must be taken with conclusions that appear to be challenging Hamilton's rule. They are not; any more than the fancifully irreducibly complex systems beloved of the creationists are a challenge to evolution by natural selection. While there may appear to be an elegant symmetry between maximizing utility functions and maximizing inclusive fitness, it doesn't follow that the world acts this way. Often, Hamilton's rule is not being addressed at all.

One of the consistent findings from experimental psychology is that participants do not see themselves as isolated laboratory creatures but as full human beings who also live outside of laboratories and might meet the recipients of their generosity or cruelty at some point – whatever an experimenter might insist. Thus, the use of so-called one-shot games (especially where it appears that people are more cooperative than expected) must be carefully considered. The experimenter might see them as one shot. It does not follow that this is how the participant sees them. Natural selection works by generating proximate mechanisms that maximize inclusive fitness on average. The fact that such systems can be made to misfire does not invalidate natural selection. No one would argue that the existence of pornography undermines sexual selection, just because humans can be fooled by it into generating nonreproductive behaviors.

As an exampl[illegible]nsider the data on ultimatum games. These are where one actor makes an offer and the other participant can choose to accept or reject it. It turns out that a robust finding is that people make larger offers that standard economic theory would predict, and it could be argued that this supports a contention that humans are notably more cooperative than utility-maximizing (taken as a proxy for inclusive fitness maximizing) theory would allow for. However, as West et al. (2011) point out, the data could equally well support a rather more depressing conclusion – namely that humans are notably more antisocial (and know it) than we previously wished. If humans have evolved to recognize that other humans are especially spiteful and vicious, then they might expect low offers to be punished. Thus the surprising behavior would be the rejection of lower offers, and we could (equally) conclude than humans are more punitive than previously thought.

It is possible that scholars have been misled by the word "selfish" in the title of Dawkins (1976) work into mistaking the maximizing of representation of genes (ultimate causation) with selfish motives (proximate motivation). Expecting selfish behavior, scholars perhaps have been pleasantly surprised to find that humans are not, in fact selfish. But selfish gene theory never predicted that they were, and showing that they are not doesn't undermine the model of the selfishness of genes.

Even more common than ultimatum games is the prisoner's dilemma. This is an experimental economic situation involving a payoff matrix that allows researchers to model interactions of cooperation and defection. It is common to find it assumed that the prisoner's dilemma has proved that in one-shot games the strategy of defection (selfishness even to the point of spite) should be maximized but that cooperative tit-for-tat exchanges will thereafter defeat all other strategies in iterated interactions. Frank (1998) provides a sophisticated set of ways to model such interactions in a way that does justice to the complexities – modeling the biology of actual interactions rather than assuming that the world must be fitted, procrustean style, to the most elegant mathematical model. It turns out that neither of these assumptions is correct.

Group Selection and Selection for Groupishness

For a variety of interesting reasons, the term "group selection" has become almost synonymous with culture and morality in some quarters. In part, this confusion is that selfish genes make selfish people (and that therefore some special, extragenetic mechanism is required to make them social), but there is more to the issue than this.

Organisms cooperate, often at a cost to themselves. How, given the famous Darwinian struggle for survival, is this possible? In the 1950s and 1960s, many social behaviors were explained in terms of benefits conferred on individuals (e.g., Tinbergen 1951; Lorenz 1966). Group selection arguments, sometimes even arguments in terms of the benefit to the species as a whole, were frequently invoked. In 1962 Wynne-Edwards proposed that organisms would voluntarily reduce their own fitness – for example, by limiting their own reproduction – so that the wider group could survive. One can still hear arguments of this kind in general circulation, although more rarely among scholars.

Both logic and empiricism were fatal to these sorts of group selection arguments, however. In terms of logic, it was reasoned that any such adaptations would be ruthlessly outcompeted by intruders that exploited them. Then, when the data were explored, it was found that the sort of self-limiting behaviors predicted could be shown to not occur.

There are other forms of group selection, including ones where the unit of selection is itself the group rather than the gene or includes the gene as well in so-called multi-level selection. All scholars agree that this sort of group selection can occur; indeed the covariance equations of Price (1972) could be expanded on one side of the equation without any limit so as to encompass any unit of selection that is desired. However, such expansion comes at a cost that the effect of selection decreases exponentially with such expansion. What this means in practice is that if groups were to bud and reproduce faster than their elements do, then group selection would overtake individual selection. In the absence of this, the effects of group selection are going to be vanishingly small, and no one has convincingly shown that this could occur with human beings.

Advocates of group selection in relation to humans tend to emphasize helping rather than, for example, genocide. But genocide and other activities like suicidal sacrifice for a military cause would potentially provide some of the strongest evidence for group selection, and in humans it is certainly well attested in our histories. Critics of inclusive fitness models are sometimes explicit that they believe that such selection dooms us to being ultimately selfish (Wilson and Sober 1998). But, "selfish" here is equivocal. Humans are very interested in the motives of one another. For example, we are intensely interested in whether someone's motives can be trusted. To say that someone is selfish is tantamount to saying that, at crucial moments, they cannot be relied upon. But the selfishness of genes – their blind replication at the expense of others (unless those other genes produce mutual benefit) – must not be confused with selfish motives. Genes have no motives.

What Hamilton's rule explains is not that all so-called altruistic motives are at heart a sham. On the contrary, it explains how genuine empathy, self-sacrifice, and love can exist without supernatural intervention. By analogy, if someone were to see a brain scan of a loved one and noted that their ventral-tegmental area was firing strongly in response to a stimulus of them, would they conclude that "love was not real" because they could see its activity in the brain? They would be foolish if they did – because what they have just seen is actual evidence of just what a real thing love is.

Au8

Philosophers have struggled for millennia to try to make sense of our sense of morality. Where could it have come from? Is it god? Is it reason? Is morality somehow part of the fabric of the universe? What Hamilton's rule demonstrates is that some aspects of morality – those that give rise to empathy, for example – are indeed part of the fabric of the universe. This is not the whole of morality, of course. And Hamilton's rule certainly provides no guidance (or possibly only provides bad guidance) on a human's actual conduct. However, it provides a non-supernatural source for our proximate social feelings toward one another.

When something persistently reoccurs in human thought, it's likely not simply a misunderstanding but revealing of underlying cognitive architecture. As Lewis Wolpert famously put it, science is profoundly unnatural. We have only been performing science (rather than mere data collection about local regularities) in recent times. The conclusions of science are typically counter-intuitive and require a record of testing and failures to build up gradually. Scientific conclusions often jar with our sensibilities.

For instance, various forms of Lamarckism (inheritance of acquired characteristics) keep cropping up in each generation perhaps because humans, as obligate investors in their children, probably cannot quite bring themselves to believe that the minute details of what they do as parents – unless they do some truly ghastly things – are normally swamped by the effects of genes. Similarly, perhaps it is the case that group selection keeps recurring as a plausible source of human moral behavior because it feels intuitively evident that we often suppress our needs for the sake of the group.

Indeed, some prominent scholars have made such introspection a major pillar of their arguments for group selection (Wilson and Sober 1998). But intuition and introspection are very misleading here. Given our species' long-documented capacity to have major sections of cognitive architecture opaque to others, it would be premature indeed to consider introspection as final. As Marvin Minsky memorably put it in an interview with Ken Campbell, "all parts of your mind are treating the other parts like tiny robots and finding ways to trick them." There is absolutely no reason to treat our inner emotions about the "good of the group" as anything more than an effective way to get ourselves to advertise ourselves as selfless members of the said group. Indeed, given that the most effective way to fool others is to sincerely believe it oneself, we have very strong reasons to be deeply suspected of introspections in this particular area most of all.

An important implication raised by Hamilton (1975) is that altruism in semi-isolated groups depends on the migration rates rather than the size of the groups. This is a rather surprising finding, but Hamilton showed that the interrelatedness of the groups will gradually tend toward the level of siblinghood if there is, for example, just one migrant per two generations.

Misconceptions About Inclusive Fitness

As well as stating what Hamilton's rule is, it is important to state what it is not. Misconceptions about inclusive fitness abound, leading to a number of attempts to rectify this in the theoretical literature (Dawkins 1979; Griffin et al. 2002; Park 2007; West et al. 2011). Some of the most common of these (that particularly matter in relation to human beings) are:

1. That cooperation is altruistic. This misconception may partly be due to the fact that humans have such exquisitely sensitive reactions to the potential cheaters and traitors in a group. This likely results in the notion that a suggestion that a behavior is somehow not really altruistic (because the actor also benefitted) becomes conflated with the idea that the cooperation is itself not genuine. In fact, many cases that are described in the literature as "altruistic" are in fact mutualistic; both actor and recipient benefit from them. It may be better to call "reciprocal altruism" by other names to emphasize this fact (West et al. 2011). Reciprocity (Trivers 1971) can be direct or indirect (mediated, e.g., through reputation in humans), but it is not altruistic, in the strict sense required by Hamilton's rule, if both parties increase fitness through the behaviors. There are a multitude of mechanisms for enforcing cooperation and these have been found across taxa. The most common is simply punishing transgressors (West et al. 2011). None of these mechanisms require Hamilton's rule or require elaborate new types of explanations to be able to explain their occurrence (although they may well require sophisticated and sensitive modeling, of course).
2. That relatedness is merely about shared genes. This is incorrect. Only genes that influence behavior (through possible altruism) can fall under the kind of selection that Hamilton's rule describes. For instance, two clones who had no altruistic genes would not aid one another simply because they were genetically identical. Nor is altruism somehow proportional to shared genes. Humans and mustard grass

share 15 % of their genes but do not show any measurable degree of altruism toward one another. As stated clearly by Krebs (1987), "The reason why relatedness is important…the coefficient of relatedness between two individuals is equivalent to the probability that they share the gene for altruism, not because they share a high proportion of other non-altruistic genes" (p. 93).

One consequence of this is that simplistic models of individuals doling our lumps of altruism in proportion to the degree of relatedness of family members are very unlikely to be empirically verified. Generational asymmetries in investment and reproductive value are likely to be important variables, for instance.

3. That there is something special about siblinghood, cousinhood, or other similar kin relationships. This is also untrue. One of Hamilton's insights was to appreciate that parent-offspring relatedness was only one way in which relatedness might matter. In a diploid species (such as humans), each offspring has a 50 % chance (ignoring complications) of inheriting a specific altruistic gene. So as long as the benefits bestowed on the offspring outweighed the costs incurred in fitness by the parent multiplied by 50 % (the coefficient of relatedness), the gene could be selected for. Indeed, non-kin who happened to share a cooperative gene could act altruistically, in principle. While it is unlikely that this happens in humans, limited dispersal patterns and viscous populations can frequently mean that groups of humans who are apparently not close kin are still highly related to one another.
4. That kin selection is group selection. Selection for groupishness is not the same as group selection. There is no space here to explore all the variety of possible meanings of the term "group selection," but some have been discussed above. It is worth noting that evolutionary biologists in general accept the fact that group adaptations (which is one of the meanings attached to the term) can only occur in very specific circumstances that do not apply to humans such as in communities of clones or in situations with no within-group competition. Many evolutionary biologists would argue that putting things in terms of group selection adds nothing in terms of explanation but carries a potential for confusion (see West et al. 2011 for an extended discussion).
5. That kin selection requires the ability of genes to recognize one another. This putative property is sometimes referred to as a "green-beard" ability (Dawkins 1976). If the gene that underlays an altruistic behavior also pleiotropically produced both visible markers (e.g., the eponymous green beards) and the preference for such markers, then the genes could aid each other more directly. This appears to be very rare in nature. However, kin discrimination can occur through a variety of proximate cues. The most obvious of these is a shared early environment. In birds this would typically be a nest, but there is plenty of evidence (such as the famous Westermarck effect that prevents siblings' sexual interest in one another) of humans' assuming (not necessarily consciously) that those they grew up with are close kin. Other likely sources of interest for humans might include the ability of fathers to discriminate likely offspring, patterns of investment that reflect degrees of paternity uncertainty, and the role of infanticide and natal neglect (see West et al. 2011 supplementary material for an extended discussion of the research in this field).
6. That animals, and humans prior to arithmetical ability, need to be able to consciously calculate relatedness (Sahlins 1977) for Hamilton's rule to apply. No conscious calculations are required here; any more than spiders are required to be able to perform Weyrauch's formula of load bearing to be able to build their webs. Mathematics may be used to model and predict behaviors, but not necessarily the mechanisms by which those behaviors occur.
7. That altruistic behavior is too complex to be captured by a single gene and that therefore there cannot be a "gene for altruism." This is misleading. Fisher (1930) noted that phenotypically neutral genes were likely to be very rare in practice. Behaviors grow out of complex

interactions of genes, not one single "gene for X." By way of example, a behavior that involves the feeding of chicks in the nest probably relies on a complex interplay of many genes working through proximate rules such as "feed whatever is in your nest, has a large patch of yellow, and is making a noise." This rule can be exploited by, e.g., a cuckoo in a reed warbler nest. However, a mutant gene that caused the reed warbler to treat its younger siblings as its offspring (say) would be an altruistic gene in the strict sense; it reduces the older reed warbler's fitness but increases that of its siblings. Such a gene would not create the feeding behavior from nothing; it would build on existing behaviors (Dawkins 1979).

8. That Hamilton's rule predicts specific interactions between individuals. For example, it is not true (despite Haldane's famous quip) that humans regularly give their lives for two brothers or eight cousins. Neither does Hamilton's rule predict that they will (or should). Despite this, it is common to see Hamilton's rule presented in undergraduate textbooks as something that will predict specific altruistic acts (see Park 2007, for extended discussion and examples of this misconception occurring). Hamilton's rule describes the circumstances under which a particular altruistic gene can be selected for, not proximate instances of behavior.

Conclusion

Hamilton's rule (1964) is a foundational, axiomatic extension of Darwin's (1859) insights concerning how species develop through natural selection. Where Darwin (1859) explained the apparent miracle of design without recourse to the supernatural, Hamilton 1964) explained the underlying apparent miracle of morality – i.e., altruism – without recourse to anything other than the components of natural selection. This insight isn't the whole of moral behavior of course. Human morality also requires reason to, for example, extend thought and behavior in logically consistent terms. Hamilton's rule certainly does not itself provide a justification for behaviors. Indeed, inclusive fitness would seem to promote (say) nepotism, and this tendency is not a justification – rather the reverse.

One of the things that humans intent on building a better world would be wise to do is to pay attention to the grain of human nature rather than be in denial of it. Humans are not slaves to their genes but their genes do keep culture on a leash, to echo E. O. Wilson's memorable phrase. Hamilton's rule delineates one of the most important ways in which this occurs. Does this make human morality some sort of mistake, as some people seem to fear? In short, do the nihilists (as H. P. Lovecraft joked) have a point when they say "The world is indeed comic, but the joke is on mankind." Not in the least. The recognition that our (proximate) moral sensibilities evolved in strict accordance with the known rules of biology means that they are real things.

More than that in principle, this realization gives us ways to identify and perhaps deal with those who do not share those proximate sensibilities. Those who lack empathy, for instance, are in principle just as disabled as those born without eyes. Biology is silent on the rational application of such moral sensibilities as shame, pride, and the desire to protect others, however. The rational application of these sensibilities in individual morality, or in the large-scale coordinations that politics requires, is a very human ability too and relies on our ability and need for reason and consistency. Only a highly simplistic moral philosophy would assume that feelings and sentiments alone were the whole of human ethics.

For biologists (and psychologists who accept that psychology must be at a bare minimum consilient with biology), then Hamilton's rule represents a powerful tool. As with all powerful tools, the potential can go both ways. Although it might seem daunting to face up to the challenges that the mathematical formulations require of us to model human behavior, it is also worth bearing in mind that human minds are a collection of complex kludges that evolved over millions of years in response to many conflicting pressures. The promise of an elegant predictive mathematical tool in the manner of the theoretical physicists is a tempting goal, though probably never attainable. That said, Hamilton's rule probably comes as close to being such a realization of the Ionian enchantment – the unification of all sciences through mathematics – as we are ever likely to get in behavioral science.

Cross-References

- ▶ Hamilton's Rule
- ▶ Kin-Recognition and Classification in Humans
- ▶ Life or Death Helping in Humans
- ▶ Patterns of Helping Among Los Angeles Women

References

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D., & Horsthemke, B. (1995). Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nature Genetics, 9*(4), 395–400.

Daly, M., & Wilson, M. (1998). *The truth about Cinderella: A Darwinian view of parental love*. New Haven: Yale University Press.

Darwin, C. (1859/1991). *On the origin of species by means of natural selection* (p. 502). London: Murray.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Dawkins, R. (1979). Twelve misunderstandings of kin selection. *Zeitschrift für Tierpsychologie, 51*(2), 184–200.

Dawkins, R. (1999). *The extended phenotype: The long reach of the gene*. Oxford Paperbacks.

Fisher, R. A. (1930). *The genetical theory of natural selection: A complete variorum edition*. Oxford: Oxford University Press.

Frank, S. A. (1998). *Foundations of social evolution*. Princeton: Princeton University Press.

Griffin, A. S., & West, S. A. (2002). Kin selection: Fact and fiction. *Trends in Ecology & Evolution, 17*(1), 15–21.

Haig, D. (1997). *The social gene. Behavioural ecology: An evolutionary approach* (pp. 284–304). Oxford: Blackwell Science.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. & II. *Journal of Theoretical Biology, 7*(1), 17–52.

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

Hamilton, W. D. (1975). Innate social aptitudes of man: An approach from evolutionary genetics. *Biosocial Anthropology, 133*, 155.

Hartley, R. V. (1928). Transmission of information. *Bell System Technical Journal, 7*(3), 535–563.

Krebs, D. (1987). *The challenge of altruism in biology and psychology. Sociobiology and psychology: Ideas, issues, and applications* (pp. 81–118). Hillsdale: Erlbaum.

Lorenz, K. Z. (1966). Evolution of ritualization in the biological and cultural spheres. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences, 251*(772), 273–284.

Mayr, E. (1961). Cause and effect in biology kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science, 134*(3489), 1501–1506.

Moore, T., & Haig, D. (1991). Genomic imprinting in mammalian development: A parental tug-of-war. *Trends in Genetics, 7*(2), 45–49.

Park, J. H. (2007). Persistent misunderstandings of inclusive fitness and kin selection: Their ubiquitous appearance in social psychology textbooks. *Evolutionary Psychology, 5*(4), 860–873. 147470490700500414.

Price, G. R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics, 35*(4), 485–490.

Price, G. R. (1995). The nature of selection. *Journal of Theoretical Biology, 175*(3), 389–396.

Queller, D. C. (1992). A general model for kin selection. *Evolution, 46*, 376–380.

Shannon, C. E. (1948/2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3–55

Smith, J. M., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*, 15.

Smith, J. M., & Szathmary, E. (1997). The major transitions in evolution. Oxford University Press.

Tinbergen, N. (1951). *The study of instinct*. New York: Oxford University Press.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*(4), 410–433.

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

Trivers, R. L. (1974). Parent-offspring conflict. *American Zoologist, 14*(1), 249–264.

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior, 32*(4), 231–262.

Wilson, D. S., & Sober, E. (1998). *Unto others: The evolution and psychology of unselfish behavior* (pp. 34–36). Cambridge, MA: Harvard University Press.

Wynne-Edwards, V. C. (1962). Animal dispersion: In relation to social behaviour (No. QL752 W9 1962a).

Author Queries

Encyclopedia of Evolutionary Psychological Science
Chapter No: 1488-1

Query Refs.	Details Required	Author's response
AU1	Please check if your affiliation(s) and, if applicable, that/those of your co-author(s) are correct and complete.	
AU2	Please provide Synonyms.	
AU3	Please check if identified head levels are okay.	
AU4	Please check if edit to sentence starting "Hamilton's rule for. . ." is okay.	
AU5	Please provide details for Smith (1964), West et al. (2007a), Hamilton (1972), Price (1970), Tinbergen (1951), Sahlins (1977) in the reference list.	
AU6	Please check if formatting of table is okay.	
AU7	Hamilton (1964a) has been changed to Hamilton (1964) as per the reference list. Please check if okay.	
AU8	Please check if edit to sentence starting "On the contrary. . ." is okay.	
AU9	Please cite Wynne-Edwards (1962) in text.	
AU10	Please provide publisher location for Smith and Szathmary (1997).	

Note:
If you are using material from other works please make sure that you have obtained the necessary permission from the copyright holders and that references to the original publications are included.