The general form of Hamilton’s rule makes no predictions and cannot be tested empirically

Martin A. Nowak, Alex McAvoy, Benjamin Allen, and Edward O. Wilson

Hamilton’s rule asserts that a trait is favored by natural selection if the benefit to others, B, multiplied by relatedness, R, exceeds the cost to self, C. Specifically, Hamilton’s rule states that the change in average trait value in a population is proportional to \( BR - C \). This rule is commonly believed to be a natural law making important predictions in biology, and its influence has spread from evolutionary biology to other fields including the social sciences. Whereas many feel that Hamilton’s rule provides valuable intuition, there is disagreement even among experts as to how the quantities B, R, and C should be defined for a given system. Here, we investigate a widely endorsed formulation of Hamilton’s rule, which is said to be as general as natural selection itself. We show that, in this formulation, Hamilton’s rule does not make predictions and cannot be tested empirically. It turns out that the parameters B and C depend on the change in average trait value and therefore cannot predict that change. In this formulation, which has been called “exact and general” by its proponents, Hamilton’s rule can “predict” only the data that have already been given.

Contributed by Edward O. Wilson, April 13, 2017 (sent for review February 2, 2017; reviewed by Michael Doebeli and Jan Rychtar)

Hamilton’s rule is a widely known concept in evolutionary biology. It has become standard textbook knowledge and is encountered in undergraduate education. For many, Hamilton’s rule expresses the intuition that cooperation evolves more easily if the benefit to others, B, multiplied by relatedness, R, exceeds the cost to self, C. Specifically, Hamilton’s rule states that the change in average trait value in a population is proportional to \( BR - C \). This rule is commonly believed to be a natural law making important predictions in biology, and its influence has spread from evolutionary biology to other fields including the social sciences. Whereas many feel that Hamilton’s rule provides valuable intuition, there is disagreement even among experts as to how the quantities B, R, and C should be defined for a given system. Here, we investigate a widely endorsed formulation of Hamilton’s rule, which is said to be as general as natural selection itself. We show that, in this formulation, Hamilton’s rule does not make predictions and cannot be tested empirically. It turns out that the parameters B and C depend on the change in average trait value and therefore cannot predict that change. In this formulation, which has been called “exact and general” by its proponents, Hamilton’s rule can “predict” only the data that have already been given.

evolution | cooperation | kin selection | sociobiology

**H**amilton’s rule is a well-known concept in evolutionary biology. It has become standard textbook knowledge and is encountered in undergraduate education. For many, Hamilton’s rule expresses the intuition that cooperation evolves more easily when there are frequent interactions among relatives, because relatives are likely to share the cooperative trait. However, Hamilton’s rule goes beyond this intuition by postulating a quantitative condition, \( BR - C > 0 \), which is said to predict whether or not a trait will be selected. Specifically, it is claimed that the change in average trait value from one time point to the next is proportional to \( BR - C \).

We immediately encounter the question of how the “benefit,” B, the “relatedness,” R, and the “cost,” C, are calculated for a given system. Surprisingly, there is no consensus about the correct method. A variety of derivations have been proposed over the years (1–10), which define B, R, and C in distinct (nonequivalent) ways. In the empirical literature, peer reviewers often disagree over which method should be used in a particular manuscript (11).

A number of recent papers (7, 9, 10) have endorsed a particular formulation (4, 5) as the exact, general, and even “canonical” version of Hamilton’s rule. This formulation, called “Hamilton’s rule—general” (HRG) (12, 13), is claimed to be as general as natural selection itself (7, 14). The derivation, which we recapitulate below, is simple and contains only a few steps.

The mathematical investigation of HRG reveals three astonishing facts. First, HRG is logically incapable of making any prediction about any situation because the benefit, B, and the cost, C, cannot be known in advance. They depend on the data that are to be predicted. At the outset of an experiment, B and C are unknown, and so there is no way to say what Hamilton’s rule would predict. Once the experiment is done, HRG will produce B and C values in retrospect such that \( BR - C \) is positive if the trait in question has increased and negative if it has decreased. But these “predictions” are merely rearrangements of the data that have been collected and already contain information about whether or not the trait has increased. In particular, the parameters B and C depend on the change in average trait value.

The second astonishing fact of HRG is that the prediction, which exists only in retrospect, is not based on relatedness or any other aspect of population structure. A common interpretation of the terms in Hamilton’s rule is that R quantifies the population structure, whereas B and C characterize the nature of the trait. But the derivation shows that this interpretation is wrong. All three terms, B, R, and C, are functions of population structure, whereas the overall value of \( BR - C \) is functionally independent of population structure. Any information about who interacts with whom cancels out when calculating the value of \( BR - C \).

The third fact of HRG is that no conceivable experiment exists that could test (or invalidate) this rule. All input data, whether they come from biology or not, are formally in agreement with HRG. This agreement is not a consequence of natural selection, but a statement about a relationship between slopes in multivariate linear regression. This relationship between slopes has been known in statistics at least since 1897 (15).

**Derivation of HRG**

We recapitulate the derivation of HRG given in refs. 4, 5, 7, 9, and 10. We also provide explicit algebraic formulas for B and C that result from this derivation.

We imagine a population of n individuals at a given point in time. Each individual i has a fitness value, \( w_i \). The list \( w = (w_1, w_2, \ldots, w_n) \) is the collection of fitness values in the population, which can be interpreted as expected or realized number of offspring in the next generation. If the total population size is constant, which is assumed for simplicity, then the average fitness is \( \bar{w} = 1 \).

**Significance**

Hamilton’s rule is a well-known concept in evolutionary biology. It is usually perceived as a statement that makes predictions about natural selection in situations where interactions occur between genetic relatives. Here, we examine what has been called the “exact and general” formulation of Hamilton’s rule. We show that in this formulation, which is widely endorsed by proponents of inclusive fitness theory, Hamilton’s rule does not make any prediction and cannot be tested empirically. This formulation of Hamilton’s rule is not a consequence of natural selection and not even a statement specifically about biology. We give simple and transparent expressions for the quantities of benefit, cost, and relatedness that appear in Hamilton’s rule, which reveal that these quantities depend on the data that are to be predicted.

Author contributions: M.A.N., A.M., B.A., and E.O.W. designed research, performed research, analyzed data, and wrote the paper.

Reviewers: M.D., University of British Columbia; and J.R., The University of North Carolina at Greensboro.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. Email: ewilson@oeb.harvard.edu.
Although the first two lists determine the value of \( BR - C \), as well as the overall genetic change in the population, a third list is required to determine the individual values of \( B, R, \) and \( C \). Because Hamilton’s rule is meant to describe phenomena such as kin selection or social evolution, the third list contains information about interactions between individuals. For each individual \( i \), the quantity \( h_i \) represents the average trait value of \( i \)’s interaction partners. The list \( h = (h_1, h_2, \ldots, h_n) \) summarizes the interactions in the population at the given time.

Some further notation is needed. The variance of a list of numbers is \( v_g = \sum_{i=1}^{n} g_i^2/n - \bar{g}^2 \). The covariance of two lists, \( g \) and \( w \), is \( c_{gw} = \sum_{i=1}^{n} g_i w_i/n - \bar{g} \bar{w} \). Note that \( c_{gg} = v_g \).

Because \( w = 1 \), the change in average trait value is

\[
\Delta \bar{g} = c_{gw}. \tag{1}
\]

HRG defines the parameter \( R \) as the slope of a best-fit line for the data in the plane \( h \) vs. \( g \) (Fig. 1E). The formula for the slope is

\[
R = \frac{c_{gh}}{v_g}. \tag{2}
\]

The covariance, \( c_{gh} \), contains terms of the form \( g_i h_i \), which is the product of the trait value of individual \( i \) and the average trait value of the interaction partners of individual \( i \). Therefore, \( c_{gh} \) and consequently \( R \) depend on population structure.

The derivation continues by calculating a best-fit plane to the data given in the 3D space \( w \) vs. \( g \) and \( h \) (Fig. 1F). The parameters \( B \) and \( C \) are defined by the slopes of this plane. The algebraic formulas are

\[
B = \frac{v_g c_{hw} - c_{gh} c_{gw}}{v_g v_h - c_{gh}^2}. \tag{3a}
\]

\[
C = -\frac{v_h c_{gw} - c_{gh} c_{hw}}{v_g v_h - c_{gh}^2}. \tag{3b}
\]

These expressions can be written as

\[
B = \frac{c_{hw} - R \Delta \bar{g}}{v_h - R^2 v_g^2}; \tag{4a}
\]

\[
C = -\frac{c_{gw} - R c_{hw} v_g}{v_g v_h - R^2 v_g^2}. \tag{4b}
\]

The parameters \( B \) and \( C \) depend on \( R \) and therefore on population structure. Therefore, “benefit” and “cost” are not just properties of a trait but depend on relatedness. It follows that changing relatedness would typically also change benefit and cost. This dependency is at odds with many empiricists’ intuition about Hamilton’s rule.

Both \( B \) and \( C \) also depend on \( \Delta \bar{g} \), which is the change in average trait value. The change in average trait value is supposed to be predicted by Hamilton’s rule, but the parameters \( B \) and \( C \) depend on this quantity. Therefore, it makes no sense to claim that HRG makes a prediction about \( \Delta \bar{g} \).

The derivation is completed by calculating the term, \( BR - C \), which leads to

\[
BR - C = \frac{c_{gw}}{v_g}. \tag{5}
\]

Note that all \( h_i \) values have canceled out in this calculation. Therefore, the prediction \( BR - C \), which is equal to \( c_{gw}/v_g \), has discarded all information about the \( h \) list, which denotes the trait list of interaction partners. Whereas the individual values of \( B \), of \( R \), and of \( C \) contain the \( h \) list, the value of \( BR - C \) does not.

Any collection of numbers that is used for the \( h \) list, as long as the denominators in Eqs. 3a and 3b are nonzero, gives the same value of \( BR - C \). For example, one can use the digits of \( \pi \) and obtain the same prediction for the change in average trait value (Fig. 2). The numbers of the \( h \) list affect the individual values of \( B \) and \( C \), but they do not affect the value of \( BR - C \).
Hamilton's rule is commonly thought to capture the idea that cooperative behaviors can be selected if the benefits go to close relatives. In this understanding, Hamilton's rule is believed to arise because interactions occur between relatives. The parameters \( B \) and \( C \) depend on all three lists. In particular, \( B \) and \( C \) also depend on the change in trait value. Therefore, they cannot be used to predict that change in any meaningful way.

The formal correctness of HRG is established by Eqs. 1 and 5. Because the variance \( v_x \) must be positive for the formalism to make sense, the sign of \( BR - C \) is the same as the sign of \( \Delta g \). But we have seen that \( B \) and \( C \) depend on \( \Delta g \). Therefore, \( BR - C \) cannot predict or explain \( \Delta g \). The expression \( BR - C \) is simply an extended form of writing the ratio \( \Delta g / v_x \), in which we add an arbitrary list that subsequently cancels out again.

**Slopes and Statistics**

The relationship between the slopes of the various linear regressions, expressed by HRG (Eq. 5), is not a consequence of biology and not a discovery of inclusive fitness theory. It is a fact of multivariate statistical analysis, the proof of which we recall in Appendix.

In a linear regression of \( w \) vs. \( g \), the slope of the line is \( m_{wg} = c_{wg} / v_g \). Likewise, a linear regression of \( w \) vs. \( h \) gives slope \( m_{wh} = c_{wh} / v_h \). Furthermore, the linear regressions of \( h \) vs. \( g \) and of \( g \) vs. \( h \) lead to the slopes \( m_{gh} = c_{gh} / v_h \) and \( m_{hg} = c_{hg} / v_h \), respectively.

In three dimensions, the multivariate regression of \( w \) vs. both \( g \) and \( h \) leads to a plane given by two slopes \( M_h \) and \( M_g \). It is standard textbook knowledge (16) that the relationship between the slopes is given by the linear system of equations

\[
\begin{align*}
    m_{wg} &= M_g + M_h m_{hg}; \\
    m_{wh} &= M_h + M_g m_{gh}.
\end{align*}
\]

(6a) \( \text{See Fig. 1 for a graphical depiction of each term.} \)

Inclusive fitness theorists call Eq. 6a “Hamilton's rule” by setting \( B = M_i \), \( C = -M_g \), and \( R = m_{wg} \). These quantities are interpreted as benefit, cost, and relatedness (7, 9) and are used to classify behaviors as “altruistic,” “mutually beneficial,” “selfish,” or “spiteful” (17). If \( B \) and \( C \) are positive, the situation is classified as altruism; if furthermore \( BR - C > 0 \), then it is concluded that the altruistic trait increases because interactions occur between close relatives.

However, it is well understood in statistics that relationships such as Eq. 6 do not themselves imply causality (18, 19). Whereas there can be causal relationships between dependent and independent variables in a linear model, these relationships cannot be deduced from linear regression alone (20, 21). Therefore, without further assumptions or information, the meanings attached to the terms in HRG have no basis in mathematics or statistics. Moreover, the derivation of HRG does not take into account any aspect of the mechanism that leads to a change in trait value and therefore cannot return a description of that mechanism (Fig. 3) (22). It merely defines quantities \( B \), \( R \), and \( C \) as functions of \( w \), \( g \), and \( h \) such that \( BR - C \) is proportional to the change in trait frequency, \( \Delta \gamma \).

**Benefit and Cost Need Not Make Sense**

Although HRG does not explain or predict the change in average trait value, it could be the case that the parameters \( B \) and \( C \) provide some biological insights. In a previous paper, we showed that it is easy to envisage biological processes that are mischaracterized by the resulting values of \( B \) and \( C \) (22). In Fig. 3, we provide further examples of this kind.

The parameters \( B \) and \( C \) also behave unbiologically in the following way: A small change in a single \( g_i \) or \( h_i \) value can make \( B \) and \( C \) jump from a very large negative value to a very large positive value (Fig. 4). For example, a small deviation in the measurement of an empirical system could change the assessment of a behavior from tremendously helpful (say, \( B = 10^{10} \)) to tremendously harmful (\( B = -10^{10} \)). Such quantities are not biologically meaningful.

**Discussion**

Hamilton's rule is commonly thought to capture the idea that cooperative behaviors can be selected if the benefits go to close relatives, because these relatives are likely to share genes for cooperation. In this understanding, Hamilton's rule is believed...
to make important, testable predictions for the evolution of social behavior: A trait is selected if benefit times relatedness exceeds cost. Clearly, such a simple and seemingly plausible statement has great intuitive appeal.

However, any intuition can only be as good as its mathematical or biological underpinning. The purpose of this article has been to clarify the mathematical derivation of Hamilton’s rule that has been endorsed as exact, general, and canonical by the inclusive fitness community (HRG) (4, 5, 7, 9, 10). The derivation of HRG is encapsulated in Eqs. 1–5. Any collection of triples can be used as input data and will turn out to be in “agreement with Hamilton’s rule” as long as the relevant denominators are nonzero. If the denominators are zero, the quantities and are undefined. The data can come from any experiment, from any theory, from a deliberate or erroneous variation of either, or be completely imaginary. All such data, biological or not, will behave as “predicted by Hamilton’s rule.” Clearly, HRG is not a statement about biology and not a consequence of natural selection.

The predictive power of HRG is equivalent to the following example: If you give me the shoe sizes and heights of a group of people, then I can predict the heights. My algorithm also works if you gave me the wrong shoe sizes.

That HRG has no predictive power has been previously noted by ourselves (22) and others (12, 23, 24), yet HRG is credited with making a variety of empirical predictions (7, 9, 14).

Much like the Price equation (25–27), HRG provides a functional relationship between quantities that are obtained from a population at two successive points in time. Whereas the change in trait frequency, need not be independent of in a statistical sense, the derivation of HRG takes neither statistical relationships nor any information about suitability of a linear model into account (7, 22). Starting from three lists, and , it fits a linear model of vs. and and finds and , and such that is proportional to . Notably, and are themselves functions of . The result is an algebraic expression for that is (functionally) independent of (Fig. 2). In analogy to our previous example, although shoe size and height could be correlated, if we already know the heights, the shoe sizes are not needed to determine heights.

In short, there is a startling discrepancy between the common intuitive understanding of Hamilton’s rule and the derivation of this rule that has been described as exact and general. In some cases, this discrepancy can be seen within a single paper. For example, ref. 7 uses 18 different variations of “Hamilton’s rule correctly predicts...” in reference to HRG, which makes no prediction at all.

Although HRG is the only formulation of Hamilton’s rule that is claimed to be exact and general, there are other approaches that define benefit, cost, and relatedness in different ways. For example, benefit and cost can be properties of individual phenotypes, and relatedness can be defined using common ancestry (1, 3, 12, 13, 24, 28). This approach, “Hamilton’s rule—special (HRS) (12, 13), has the advantage of making testable predictions, because the benefit and cost of a phenotype can be determined in advance. However, it is easy to show that HRS holds only for special cases and not in general (12, 28–31).

The existence of these conflicting definitions makes it impossible to meaningfully test or falsify Hamilton’s rule. Any theoretical or empirical result that appears to violate Hamilton’s rule can be reanalyzed using HRG to show that the outcome is “as predicted by Hamilton’s rule.” Indeed, this pattern has been repeated many times in the literature (7, 14, 28, 32–34). It appears that there are no real or hypothetical data that the inclusive fitness community would accept as a violation of Hamilton’s rule.

Some papers attempt to empirically test Hamilton’s rule (35–40). Tests of Hamilton’s rule are typically done by experimentally determining the benefits and costs of a phenotype and quantifying relatedness using genetic markers or pedigree. But such a procedure—while scientifically reasonable—tests only HRS, which is not the exact and general version of Hamilton’s rule. We are aware of only one paper (23) that attempts to apply HRG to an empirical system. They find, as we have shown here, that HRG does not predict any aspect of their system, but yields only a value of that coincides with the result they have already obtained.

The biological question at hand is how population structure affects the evolution of social behavior, which is a deep and important question that has been studied extensively (41–49).
The intuition that a cooperative gene can spread by preferentially conferring benefits on cobearers of this gene is correct. However, Hamilton’s rule, in its exact and general formulation, is unrelated to this biological intuition and (in general) neither predicts nor explains the evolution of social behavior.

Indeed, we should not expect that interplay of population structure and social behavior can be reduced to a simple rule with three parameters. Social interactions, which are typically multilateral (50) and nonlinear (51, 52), cannot be expressed by a single benefit and cost. Complex population structures (43, 46, 53, 54) cannot be expressed by a single relatedness quantity. Assortment among relatives often has a positive effect on cooperation (41, 44–47), but in other cases it has a negative effect (48, 55) or no effect at all (42, 45). A good understanding of these questions, like all great problems in science, will require careful empirical observation in concert with meaningful mathematics.

**Appendix**

Here, we recapitulate the derivation of HRG as a simple consequence of a well-known result in statistics. For any collection of \(n\) data points in the form of triples, \(\{(w_i, g_i, h_i)\}_{i=1}^{n}\), suppose that (i) \(m_{w,g}\) is the slope of a least-squares regression line \(w\) vs. \(g\), (ii) \(m_{w,h}\) is the slope of \(w\) vs. \(h\), (iii) \(m_{g,h}\) is the slope of \(g\) vs. \(h\), and (iv) \(m_{g,h}\) is the slope of \(g\) vs. \(h\). Furthermore, for the least-squares plane expressing \(w\) vs. \(g\) and \(h\), let (v) \(M_g\) be the slope of the line obtained by holding \(h\) constant, and let (vi) \(M_h\) be the slope of the line obtained by holding \(g\) constant (see Fig. 1 for details).

**Proposition.** These slopes satisfy the following equations:

\[
\begin{align*}
    m_{w,g} &= M_g + M_h m_{w,h}; \\
    m_{w,h} &= M_h + M_g m_{w,g}.
\end{align*}
\]

**Proof.** Suppose that \(\{(x_1, \ldots, x_n), (y_1, \ldots, y_n)\}_{i=1}^{n}\) is a collection of \(n\) data points, which can be expressed in matrix form as

\[
X := \begin{pmatrix}
1 & x_{11} & \cdots & x_{1p} \\
1 & x_{21} & \cdots & x_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & \cdots & x_{np}
\end{pmatrix}; \quad y := \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix}.
\]

If we are to fit a linear model of the form \(y = x^T \beta\) for some coefficient vector, \(\beta\), then it is a well-known result in statistics (20) that the least-squares solution satisfies the equation

\[
\left(X^T X\right) \beta = X^T y.
\]

Because all of the terms appearing in Eq. 7 are slopes, we may assume without a loss of generality that \(\bar{g} = h = \bar{w} = 0\). From Eq. 9, we obtain

\[
\begin{align*}
    m_{w,g} &= \frac{w \cdot g}{g \cdot g}; \\
    m_{w,h} &= \frac{w \cdot h}{h \cdot h};
\end{align*}
\]
Moreover, we also see from Eq. 9 that
\[
\begin{pmatrix}
0 & 0 \\
0 & g \cdot g \cdot h \\
0 & g \cdot h \cdot h
\end{pmatrix}
\begin{pmatrix}
M_{gh} \\
M_{gh} \\
M_{gh}
\end{pmatrix}
= 
\begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix},
\]
which, together with Eq. 10, gives Eq. 7. Note that Eq. 7 can be written in matrix form as
\[
\begin{pmatrix}
m_{ag} \\
m_{ah} \\
m_{bh}
\end{pmatrix}
\begin{pmatrix}
1 \\
m_{gh} \\
1
\end{pmatrix}
\begin{pmatrix}
M_{ag} \\
M_{ah} \\
M_{bh}
\end{pmatrix}
= 
\begin{pmatrix}
m_{ag} \\
m_{ah} \\
m_{bh}
\end{pmatrix}
\begin{pmatrix}
M_{ag} \\
M_{ah} \\
M_{bh}
\end{pmatrix}.
\]
Provided \(m_{ag}m_{ah} \neq 1\), we can invert this matrix to see that

\[
M_{gh} = \frac{m_{ag} - m_{ah}m_{bh}}{1 - m_{ag}m_{ah}},
\]
\[
M_{ah} = \frac{m_{bh} - m_{ag}m_{ah}}{1 - m_{ag}m_{ah}}.
\]
Because \(B = M_{h} \) and \(C = -M_{g} \) in HRG, we have Eq. 3. The only situation in which we cannot solve explicitly for \(M_{g}\) and \(M_{h}\) in Eq. 7 is if \(m_{ag}m_{ah} = 1\), which happens if and only if there exist constants \(k_{1}\) and \(k_{2}\) (not both zero) for which \(k_{1}(y_{i} - \bar{y}) = k_{2}(\bar{x}_{i} - \bar{y})\) for each \(i\). In this case, the values of \(M_{g}\) and \(M_{h}\), and therefore also the values of \(B\) and \(C\) in HRG, are undefined.

Introducing \(R' = m_{gh}\), we can write
\[
B = \frac{c_{hw} - R\Delta \tau}{v_{h} (1 - RR')},\]
\[
C = -\frac{\Delta \tau - R' c_{hw}}{v_{h} (1 - RR')},
\]
which gives a symmetric alternative to Eq. 4.

It is interesting to note that the relationships between linear regressions with one and two explanatory variables, which are captured in Eq. 13 and give rise to HRG, appeared in the statistics literature as far back as 1897 (15).

ACKNOWLEDGMENTS. The authors thank Donald Rubin, Karl Sigmund, and John Wakeley for helpful discussions.