Invariance and Symmetry in Evolutionary Dynamics

Simon M. Huttegger, Hannah Rubin, and Kevin J. S. Zollman

Abstract. The concept of fitness is central to evolutionary biology. Models of evolutionary change typically use some quantity called "fitness" which measures an organism's reproductive success. But what exactly does it mean that fitness is such a measure? In what follows, we look at the interplay between abstract evolutionary models and quantitative measures of fitness and develop a measurement-theoretic perspective on fitness in order to explore what makes certain measures of fitness significant.

1 Introduction

The concept of fitness is central to evolutionary biology. Fitness is a measure of an organism's reproductive success and is thus a crucial element of the theory of natural selection. But what exactly does it mean that fitness is such a measure?

The aim of this paper is to develop a measurement-theoretic perspective on fitness in order to explore what makes certain measures of fitness significant. We are hereby guided by the pioneering work of Wagner (2010),[1] who writes that

> Fitness is not a primary observable feature of organisms, such as their weight, color, and chemical composition that are defined independently of any further biological insight. Fitness is a concept that plays a role in explaining evolutionary dynamics ... For that reason the definition, measurement and the mathematical properties of fitness cannot be considered separately from the theory of natural selection. In particular, the quantitative concept of fitness only makes sense if it can be inserted into mathematical models to predict or explain evolutionary change. The mathematical theory of evolution and the measurement of fitness are mutually dependent.

Here, Wagner refers to the well-known interplay between theory and measurement. Measurements of quantities are not performed in a theoretical vacuum. Rather, there is a coevolution between the development of theory and measurement practice. Theory development guides the design of measurement apparatuses, and increased experience with measurement instruments and their outcomes feeds back on theories and models. Once a theory is well-developed and has reached a (at least temporary) reflective equilibrium, measurement can be viewed from the perspective of that theory (see van Fraassen, 2008, for a more detailed discussion). The concepts used in measurement have a definite meaning within a theory and cannot be understood properly without the theoretical context.

In what follows we study the interplay between evolutionary models and quantitative measures of fitness. Our approach is inspired by the method of dimensional analysis in physics (Krantz et al., 1971). The basic idea of dimensional analysis is that how certain quantities are combined in equations puts coherence constraints on them. The equations we are going to consider are models of evolution. Such models typically involve some quantity called "fitness"

---

[1] For a general discussion of measurement theory in biology see Houle et al. (2011). For measurement theory more generally see Krantz et al. (1971), Suppes et al. (1989), Luce et al. (1990), and Hand (2010).

that influences the dynamical behavior of the system. We are interested in the following question: what are the properties of the dynamics that ought to be preserved in order to have a useful model of evolutionary change? As we shall see, depending on the properties that ought to be preserved, we get more or less stringent requirements on the measures of fitness used in models of evolutionary change.

2 Measurement Theory

In all evolutionary models, fitness is represented by a number. Although their familiarity often belies this fact, numbers have an enormous amount of structure that can sometimes imply stronger empirical claims than intended. To use a familiar example, if I say "it is twice as hot today as yesterday," I am saying something meaningless. If it is 30 degrees Fahrenheit today and yesterday it was 15F does not imply that it is meaningful to say today is "twice as hot" as yesterday. This is easily illustrated by the fact that 30F is approximately -1C and 15F is approximately -9.5C. So the statement "it is twice as hot today as yesterday" is true under one valid temperature scale but not under another equally valid scale. Some parts of the numerical structure of Fahrenheit (or Celsius) imply more than we are entitled to say.

This example illustrates that numbers can be assigned to empirical objects in completely arbitrary ways. In science, there must be constraints on how numbers are interpreted, since numerical assignments are supposed to capture something about empirical reality. Otherwise, results of measurements or models may just be due to the mathematical structures we use and have nothing to do with the empirical reality we are really interested in. In the case of fitness, we are interested in how much structure our theoretical apparatus requires, and in the subsequent sections we will turn to that. But, first, we need to quickly review how these questions are addressed.

The theory of measurement offers a systematic approach to the study of the properties of numbers used in measurement. Measurement theory goes back to at least von Helmholtz (1887), and has made significant advances in the second half of the 20th century. Of particular importance is the *representational theory of measurement*, which was developed mainly in mathematical psychology (Krantz et al., 1971; Suppes et al., 1989; Luce et al., 1990). The representational theory of measurement aims at identifying qualitative axioms—i.e. principles that do not involve numbers—that capture a measurement concept, and then trying to establish that there *exists* a numerical representation of the concept that reflects its qualitative properties. Well known examples are the measurement of physical quantities like length, mass, or time (Hölder, 1901).

The representational theory of measurement does not give answers to all philosophically significant aspects surrounding measurement.[2] We set these issues aside here. What is important for our project is that the representational theory of measurement is very well suited for exploring the relationship between invariance and measurement. This relation was discussed extensively by Stevens (1946), who identified a number of *scale types*. These scale

---

[2] For discussions see Chang (2004) and van Fraassen (2008). See Baccelli (forthcoming) for an analysis of the role of the representational theory.

types identify what parts of numerical structure correspond to empirical claims and what parts of the structure should be ignored. This work was later substantially extended by measurement theorists (Narens, 2002).

The three most common scale types are *ordinal scales*, *interval scales*, and *ratio scales*. Each one is based on an assignment of real numbers to objects that are ordered according to a binary weak ordering relation (i.e. a relation that is connected and transitive). "At least as long as" is one canonical example. A numerical assignment that captures just the order is an *ordinal scale*. More precisely, suppose R is a weak ordering of a set of objects; aRb means that object a is not ranked above object b. Then an ordinal scale, s, assigns numbers to all objects such that $s(a) \leq s(b)$ if and only if aRb. This is the sense in which the scale s represents the relation R.

Given a particular s that represents R, one can ask how s might be transformed into another numerical assignment s' while continuing to represent R. In the case of an ordinal scale, any *strictly monotonic* transformation of s is an admissible representation of R. As a result, ordinal scales are unique only up to strictly monotonically increasing transformations. This implies that, in order to be meaningful, numerical statements must be *invariant* relative to monotonically increasing scale transformations. In other words, numerical statements must be true about all ordinal scales that represent R. Otherwise, they go beyond how R orders empirical objects and rely on a particular choice of scale.[3] For instance, ordinal scales do not allow the calculation of expected fitness since the values of expected fitness can be changed arbitrarily under strictly monotonic transformations.

Interval scales have more structure than ordinal scales. In addition to representing the ordering relation, they preserve *ratios of differences*. This implies that interval scales are unique up to *positive affine transformations*. Starting with any scale s, we can multiply s with an arbitrary positive number and add an arbitrary number to get another admissible scale. An example of an interval scale is the measurement of temperature in Celsius or Fahrenheit. This is why statements like "twice as hot" are meaningless, because that comparison is not invariant over positive affine transformations.

Positive affine transformations correspond to choosing different zero and units of a scale. *Ratio scales* have a natural zero, so the only conventional aspect is choosing a unit. Thus, a ratio scale is unique up to *multiplication by a positive number*. This implies that ratios of scale values are invariant across all scales. Examples of ratio scales are length or weight.

There are scales with no transformations that turn it into another admissible scale, such as counting the number of objects in some set. They are known as *absolute scales*. For absolute scales all numerical statements are invariant.

In their famous development of utility scales, von Neumann and Morgenstern provide a strategy for transforming ordinal preferences over lotteries into an interval scale (von Neumann and Morgenstern, 1944). The underlying idea is that if we take a series of objects

---

[3] The role of invariance in discussions of meaningfulness is investigated in Narens (2002). Mitteroecker and Huttegger (2009) and Huttegger and Mitteroecker (2011) discuss invariance and meaningfulness in a biological context. See also Houle et al. (2011).

and also a randomizing device, and if people obey a set of "consistency" axioms, we can extract more information that allows us to utilize an interval scale.

Traditional economics has, for some time, built itself on von Neumann–Morgenstern utility. Modern economic game theory is no exception, and the fundamental predictions of game theory are invariant over legitimate transformations of utility. If one identifies all the Nash equilibria of a game for players with particular utility functions, those Nash equilibria are invariant over all positive affine transformations of the players' utilties. Traditional economic game theory does not step beyond the limitations imposed by the utility theory on which it is based.

So far we have begun with a scale—ordinal, interval, or ratio—and discussed what scientific claims would be regarded as meaningful on those scales. We can also go in the reverse direction: we can begin with a scientific theory where we designate certain classes of claims as meaningful and then ask: what scales would be required for the fundamental quantities to underwrite our claims of meaningfulness? This mode of inquiry allows one to determine the commitments that a scientific theory makes regarding the structure of its fundamental parts.

By way of illustration, we could begin with traditional game theory and designate that all Nash equilibria are "meaningful." We could then ask, what is the weakest scale for the utilities that would warrant those claims of meaningfulness? In this case, the answer would be the interval scale. Utility could not be an ordinal scale because some monotonic transformations of the underlying utility functions would alter the location of mixed strategy Nash equilibria.

Game theory was imported into biology by Maynard Smith and Price (1973) and the first formal model of evolutionary change in game theory was presented by Taylor and Jonker (1978). The move from economics to biology appears straight-forward: the concept of utility is replaced by fitness and the concept of maximization by rational choice is replaced with the optimizing force of evolution by natural selection. As we will see, things are slightly more complicated.

3 Evolutionary Dynamics

The basic evolutionary model we are going to consider is known as the *replicator dynamics* in evolutionary game theory and as *Wright's selection equation* in population genetics (Taylor and Jonker, 1978; Hofbauer and Sigmund, 1998; Wagner, 2010). We focus on the continuous-time version of this dynamics, but the discrete time version can be treated in the same way.[4]

Suppose there are n traits, $s_1, \ldots, s_n$, in a population. Traits may be behavioral strategies or other phenotypic characteristics. In population genetics they represent alleles. We denote the relative frequency of individuals having trait i by x_i. The population state is given by the vector of relative frequencies $\boldsymbol{x} = (x_1, \ldots, x_n)$. Let $f_i(\boldsymbol{x})$ be the fitness of trait i if the population is in state $\boldsymbol{x}$. We make the standard assumption that f_i is a linear function of the basic fitness type i gets from interaction with type j, that is, $f_i(\boldsymbol{x}) = \sum_j f_i(j) x_j$. For now, we don't know what exactly fitness is. All we assume is that it is some real-valued function.

---

[4] Although the conclusions for the discrete time version will be different.

According to the replicator dynamics, the rate of change in trait i is given by the following system of ordinary differential equations:

$$\dot{x}_i = x_i \left(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \right), \tag{1}$$

where $\bar{f}(\boldsymbol{x}) = \sum_i x_i f_i(\boldsymbol{x})$ is the average fitness in the population. The replicator dynamics is a particular instance of the idea that traits with above average fitness increase in frequency, whereas traits with below average fitness decrease, which is the fundamental idea of natural selection.

A common variant of the replicator dynamics is the adjusted replicator dynamics (Maynard Smith, 1982):

$$\dot{x}_i = \frac{x_i \left(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \right)}{\bar{f}(\boldsymbol{x})} \tag{2}$$

The dynamics (2) results from the replicator dynamics by re-scaling with the average fitness. We will see below that the invariance properties of the replicator and the adjusted replicator dynamics can be quite distinct.

We will also consider the *two-population replicator dynamics*. The number x_i represents the relative frequency of individuals with trait i in population one, and the number y_i represents the relative frequency of individuals with trait j in population two. The state of the system is given by $\boldsymbol{x} = (x_1, \dots, x_n)$ and $\boldsymbol{y} = (y_1, \dots, y_m)$, where n and m are the number of traits in populations one and two, respectively. The two populations interact with one another, giving rise to the system

$$\dot{x}_i = x_i \left(f_i(\boldsymbol{y}) - \bar{f}(\boldsymbol{x}, \boldsymbol{y}) \right)$$
$$\dot{y}_i = y_i \left(f_j(\boldsymbol{x}) - \bar{f}(\boldsymbol{y}, \boldsymbol{x}) \right) \tag{3}$$

Here, $f_i(\boldsymbol{y})$ is the fitness of trait i in population one if population two is in state $\boldsymbol{y}$, $f_j(\boldsymbol{x})$ is the fitness of trait j in population two if population one is in state $\boldsymbol{x}$, and $\bar{f}(\boldsymbol{x}, \boldsymbol{y}), \bar{f}(\boldsymbol{y}, \boldsymbol{x})$ are the respective average fitness in populations one and two.

The replicator dynamics has been widely used in population genetics and evolutionary game theory. Besides obvious applications in evolutionary biology, it is also an important in studies of cultural evolution (Weibull, 1995; Hofbauer and Sigmund, 1998).

The replicator dynamics models a selection process in an explicitly dynamical way. Often, researchers prefer to identify the *evolutionarily stable states* of a population. In the present setting, a state $\boldsymbol{x}$ is evolutionarily stable if the following two conditions hold:

(i) $f(\boldsymbol{x}, \boldsymbol{x}) > f(\boldsymbol{y}, \boldsymbol{x})$
(ii) If $f(\boldsymbol{x}, \boldsymbol{x}) = f(\boldsymbol{y}, \boldsymbol{x})$, then $f(\boldsymbol{x}, \boldsymbol{y}) > f(\boldsymbol{y}, \boldsymbol{y})$

The first condition says that x is a best response to itself (i.e., it is a symmetric Nash equilibrium). Thus, in a majority x population, no other trait does better against the majority. The second condition says that if there is an alternative best response y, x has a higher fitness when interacting with y. Hence, a population that mainly consists of x individuals cannot be invaded by a small number of individuals that exhibit an alternative best response y, assuming the population is large and interactions are uncorrelated.

Evolutionarily stable states are asymptotically stable rest points of the replicator dynamics (1), but the converse is not true (Hofbauer and Sigmund, 1998). Thus, already at the level of the replicator dynamics evolutionary stability does not capture every type of stable rest point. For this and other reasons, the concept of evolutionary stability should be used with caution (Huttegger and Zollman, 2013; Rubin, 2016). But despite limitations, evolutionary stability is a useful and influential concept in evolutionary biology.

4 Invariance and Evolutionary Dynamics

We are now in a position to study fitness as used in the replicator dynamics in terms of invariance. Okasha (2018, §6.6) discusses fitness functions along broadly similar lines, but we will provide a more thorough analysis here.

Models in the biological and social sciences, unlike models in physics and chemistry, are often not supposed to provide precise numerical predictions. Instead, the predictions of evolutionary models are often thought of as qualitative, in that they predict end states or directions of evolution, but not necessarily the state of a given population at a particular time. Thus the idea of the invariances governing models such as the replicator dynamics has to be modified. We need to clarify which aspects of a dynamical model should be preserved. We will consider several possibilities. Once the invariances have been determined, we then show what constraints this places on the fitness function used in the model.

We start with the one-population replicator dynamics. The system (1) defines a set of trajectories on a simplex of dimension $n - 1$ (n is the number of strategies of the underlying game). There are various aspects of that system that we might want to preserve:

1. The solution trajectories should be preserved together with the velocity with which they are traversed.
2. Only solution trajectories should be preserved.
3. Certain qualitative features, such as the location of rest points and their stability properties, should be invariant.

The first case is the most restrictive. Suppose the fitness functions in (1) are changed in some way. Then the trajectories will be altered unless the change involves only multiplication of fitnesses by a positive real number.[5] That is, if we multiply fitnesses by $a > 0$, we have

---

[5] Multiplication by a negative real number would leave the trajectories intact, but switch their direction.

$$\dot{x}_i = ax_i\Big(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\Big), \tag{4}$$

This system has the same trajectories as the original one, but the trajectories are traversed slower or faster depending on whether a is less than or larger than 1. To see why, notice that multiplying with a leaves the direction of the vectors in the vector field generated by (1) the same (Weibull, 1995, §3.1.2), but the length of the vectors is determined by the difference in fitness which will be altered by a. Thus, if we want to preserve the trajectories and the velocity, then the fitness function f has to be an absolute scale.

Consider the simple coordination game pictured in Figure 1a. The trajectories and speed of change are pictured in Figure 1b. If fitness is multiplied by a constant $0 < a < 1$, the trajectories remain the same, but the speed of motion is reduced. Conversely if they are multiplied by $a > 1$, the speed of motion is increased.

	A	B
A	1,1	0,0
B	0,0	1,1

(a) A simple two strategy coordination game
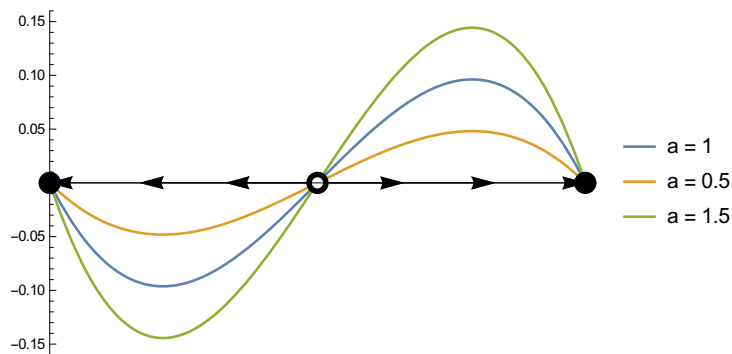
(b) An illustration of the trajectories and speed for the one-population replicator dynamics. The *x*-axis represents the proportion of the population playing *A*. The *y*-axis represents the rate of change of the population.

Figure 1: An example of the change in speed of evolution as fitness is multiplied by a positive constant $a > 0$.

The situation is different in the adjusted replicator dynamics (2). If f is a ratio scale, then not just the trajectories but also the time scale factor are invariant. This follows since the scaling factor $a > 0$ appears in both the numerator and the denominator of (2). The velocity at which the system moves along the orbits is the same for all admissible scales. Thus, *what is invariant under an evolutionary dynamics depends on the evolutionary model*.

Therefore, if we are interested in preserving both the trajectories and the speed of evolution, the replicator dynamics requires that fitness represents an absolute scale, while the adjusted replicator dynamics requires only a ratio scale.

Suppose that we are not concerned with the speed at which trajectories are traversed. Then f can be multiplied by a positive real number without having an inadmissible numerical representation of fitness. Fitness is being measured on a ratio scale for both dynamics. As we have just seen, in this situation all admissible numerical representations leave the trajectories of the system invariant. This has important consequences. Not only does a change of scale not affect the number, location, and stability properties of rest points or other regions of state space (such as periodic orbits), but it is also true that the basins of attraction, and thus their sizes, are preserved. Statements such as claiming that one basin of attraction is larger than another are thus meaningful for ratio-scaled fitness.

For fitness measured on a ratio scale, then, the trajectories of the replicator dynamics are invariant under admissible scale transformations up to a change of time scale, and the adjusted replicator dynamics is fully invariant under admissible scale transformations. If f is an interval scale, the trajectories of the replicator dynamics are again invariant under admissible scale transformations. This follows since

$$x_i \left(a f_i(\boldsymbol{x}) + b - \sum_i a f_i(\boldsymbol{x}) - b \right) = a \left(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \right),$$
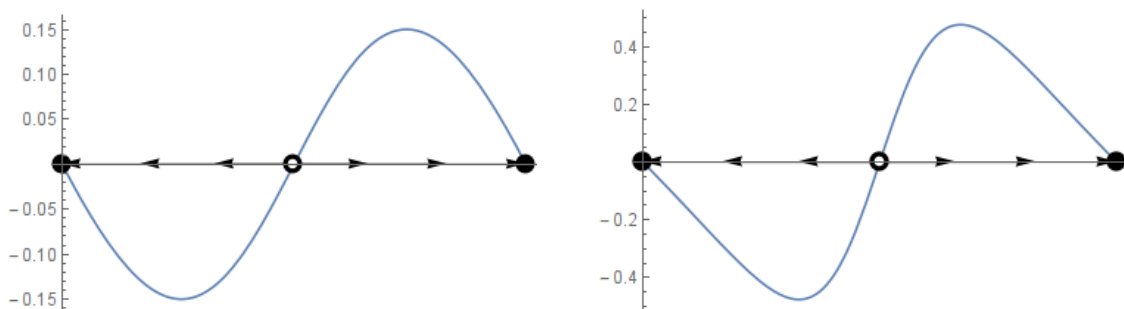
so that the same trajectories are traversed at a different velocity after a change of scale. Hence, for the invariance properties of the replicator dynamics it doesn't matter if the fitness function is a ratio scale or an interval scale. For the adjusted replicator dynamics, however, it does. Consider the change of scale to $f' = a f_i + b$, where $a > 0$. Then (2) becomes

$$\dot{x}_i = \frac{x_i \left(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \right)}{\bar{f}(\boldsymbol{x})} = \frac{a x_i (f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}))}{a \bar{f}(\boldsymbol{x}) + b} = \frac{x_i (f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}))}{\bar{f}(\boldsymbol{x}) + \frac{b}{a}}.$$

The term $\frac{b}{a}$ in the denominator can change the evolutionary dynamics significantly. If b is negative and is of significantly larger magnitude than a it may even happen that

$$\bar{f}(\boldsymbol{x}) + \frac{b}{a}$$

is negative for some $\boldsymbol{x}$, in which case the scale transformation would partially invert the direction of vectors in the vector field. Changes in b can radically alter the dynamic trajectories and rest points (see Figure 2).

(a) Coordination game with $a = 1$ and $b = 0$ (b) Coordination game with $a = 1$ and $b = -0.4$

(c) Coordination game with $a = 1$ and $b = -0.75$ (d) Coord. game with $a = 1$ and $b = -1.25$
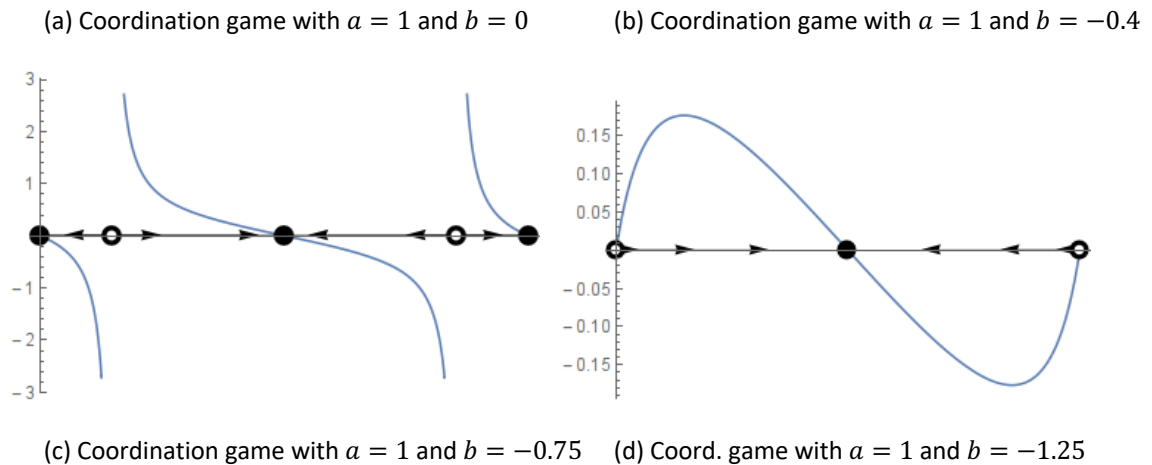
Figure 2: Illustration of the sensitivity of the adjusted replicator dynamics to the addition of a negative constant to all fitness values

This result also qualifies the well-known fact that the orbits of the replicator dynamics and the adjusted replicator dynamics are the same, up to a change of time scale (Cressman, 2003, §2.1). We see that this is true only if fitness is measured on an absolute or on a ratio scale.

This also shows how complicated the relationship is between the two dynamics. If we want to preserve both the trajectories and the speed with which they are traversed, the adjusted replicator dynamics is more permissive—it requires less from our fitness measurements than does the replicator dynamics. On the other hand, if we only care about the trajectories, but not the speed of traversal, the adjusted dynamics is less permissive—it requires more from our fitness measurements relative to the replicator dynamics.

Ordinal scales have weaker invariance properties since they are associated with a larger class of transformations. This is reflected in evolutionary theory. Let's look at the concept of evolutionary stability first. Recall that in the definition of evolutionary stability, only the ordering of fitness plays a role. But it is not quite correct that it is enough to measure fitnesses on an ordinal scale. As long as we only compare the fitnesses of pure strategies, it is true that an ordinal scale is sufficient. But as soon as we introduce mixed strategies (which is unavoidable in evolution), we need to be able to calculate expected payoffs. This requires fitness to be measurable on at least an interval scale.

For the same reasons there seem to be no interesting invariance properties of the replicator dynamics and the adjusted replicator dynamics for ordinal scales. Nash equilibria, and thus the rest points of both types of replicator dynamics, are not invariant under admissible scale transformations. Also, stability properties of rest points can change due to the arbitrariness of expected payoff calculations in the context of ordinal scales. These conclusions are slightly different when fitness is frequency-independent, i.e. when the fitness of a strategy does not depend on the population composition. In this case, ordinal scales do preserve rest points and their stability properties, though the trajectories and the speed at which they are traversed can change.

Consider for example the two games pictured in figure 3.[6] In this example, an ordinal-rank-preserving transformation affects the stability of a mixed rest point while also completely removing another. In this example the point (1/2, 1/2, 0) is a Nash equilibrium in the original game which is an asymptotically stable rest point in the replicator dynamics. After transformation, that point remains a rest point in the replicator dynamics, but it is no longer a Nash equilibrium and is therefore unstable.
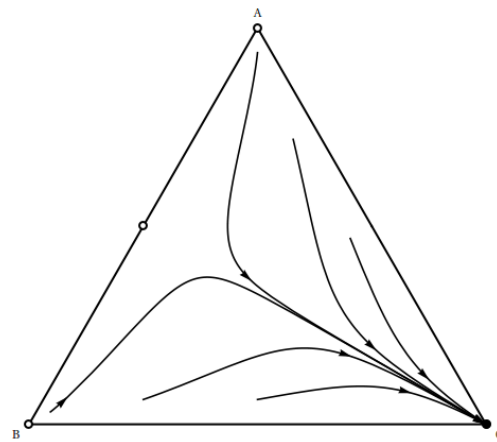
	A	B	C
A	0	10	0
B	10	0	0
C	7	2	1

(a) A three strategy game with a stable rest point at *(1/2, 1/2, 0)*

	A	B	C
A	0	10	0
B	10	0	0
C	7	6	1

(b) A three strategy game which is an ordinal transformation of the game to the left, but with an unstable rest point at *(1/2, 1/2, 0)*
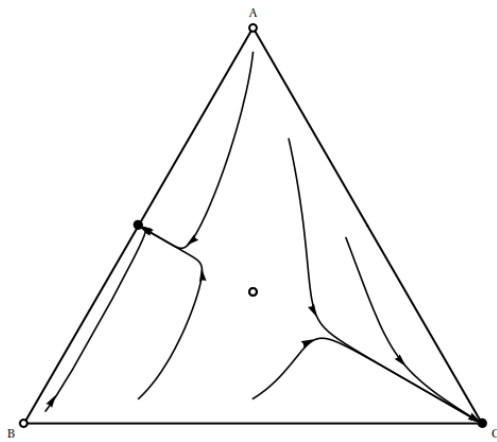
(c) An illustration of the replicator dynamics for the game (a)

(d) An illustration of the replicator dynamics for the game (b)

Figure 3: An example of two games where one is an ordinal transformation of the other but where the stability properties of a rest point are altered through changing the Nash equilibria.

Even if we restrict ourselves to situations where the monotonic transformation preserves the location of Nash equilibria, we can change their stability properties. Consider the two variations of the classic game Rock-Paper-Scissors pictured in Figure 4. Each game is a monotonic transformation of the other and both games feature unique Nash equilibria at (1/3, 1/3, 1/3). However, in the left hand game, the equilibrium is unstable and in the right hand game the equilibrium is asymptotically stable and a global attractor.

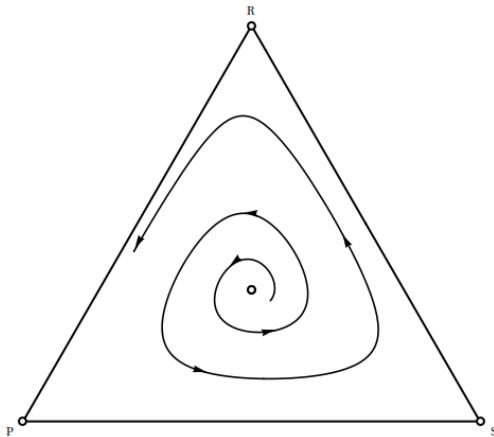|	R	P	S
R	0.25	-1	1
P	1	0.25	-1
S	-1	1	0.25

(a) A version of Rock-Paper-Scissors where the mixed strategy Nash equilibrium is unstable

	R	P	S
R	-0.25	-1	1
P	1	-0.25	-1

| S | -1 | 1 | -0.25 |

(b) A version of Rock-Paper-Scissors where the mixed strategy Nash equilibrium is asymptotically stable and a global attractor

(c) An illustration of the replicator dynamics for the game (a)
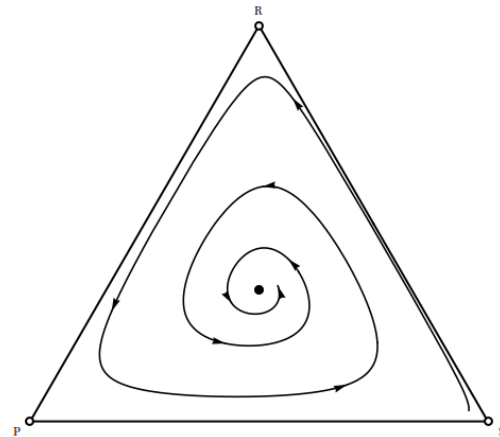
(d) An illustration of the replicator dynamics for the game (b)

Figure 4: An example of two games where one is an ordinal transformation of the other but where the stability properties of a Nash equilibrium are altered.

If there are good reasons to restrict the class of ordinal transformations in certain ways, there are more interesting invariances. Consider game dynamics of the form

$$\dot{x}_i = x_i g_i(\boldsymbol{x}),$$

where the functions g_i are continuously differentiable and generate a dynamics in the simplex. The function g_i is called the *growth rate* of i. Consider the class of all transformations that preserve the order of growth rates. This leads to the class of *payoff monotonic dynamics* (Hofbauer and Sigmund, 1998, §8.2.). Every payoff monotone dynamics has the same rest points as the replicator dynamics. Payoff monotone dynamics share many stability properties of the replicator dynamics, such as that strict Nash equilibria are asymptotically stable.

Importantly, however, payoff monotone dynamics don't all generate the same trajectories. Hence basins of attraction can change under admissible changes of the dynamics. This can significantly impact the conclusions that we draw from a model. A rest point with a large basin of attraction is usually interpreted as being significant since it attracts a large set of initial condition; if initial conditions are chosen at random, such a rest point will be observed more often than others with a smaller basin of attraction. If the underlying fitness function gives rise to a payoff monotone dynamics, no such conclusions are justified.

In two population models things are a bit different. Consider ratio scales. Changing the scale in population one by a factor of $a_1 > 0$ and in population two by a factor of $a_2 > 0$ results in the system

$$\dot{x}_i = a_1 x_i (f_i(\boldsymbol{y}) - \bar{f}(\boldsymbol{x}, \boldsymbol{y})),$$

$$\dot{y}_j = a_2 y_j \left(f_j(\boldsymbol{x}) - \bar{f}(\boldsymbol{y}, \boldsymbol{x}) \right).$$

Since the two populations interact with each other, this does not just lead to a change in time scale of the overall dynamics. Consider a two-population, two-strategy game. In this case, the two-population replicator dynamics (3) is two-dimensional:

$$\dot{x} = x\left(f(y) - \bar{f}(x,y)\right)$$
$$\dot{y} = y\left(f(x) - \bar{f}(y,x)\right)$$

The direction and rate of change is determined by the ratio $\dot{x}/\dot{y}$. This ratio depends on the factors a_1, a_2, since in general

$$\frac{x\left(f(y) - \bar{f}(x,y)\right)}{y\left(f(x) - \bar{f}(y,x)\right)} \neq \frac{a_1 x(f(y) - \bar{f}(x,y))}{a_2 y(f(x) - \bar{f}(y,x))}$$

As a consequence, a change of ratio scale in the two population replicator dynamics can change the trajectories of the system.[7]

Thus, basins of attraction are not invariant under those changes. However, close to rest points the effects of a_2, a_2 diminish, and so the stability of rest points will not be affected. For interval scales the same is true, since the additive factors in the change of scale cancel out when taking the payoff difference between the expected fitness of a strategy and the average fitness in the population.

As an example, consider the two versions of the "Chain-Store game" pictured in Figure 5. Each game is constructed by multiplying the row players fitness by a constant (either 0.1 or 10). This game features one asymptotically stable Nash equilibrium and a connected set of Nash equilibria. As can be seen by the phase portraits, the basins of are substantially altered by multiplying the row player's fitness by a positive constant.
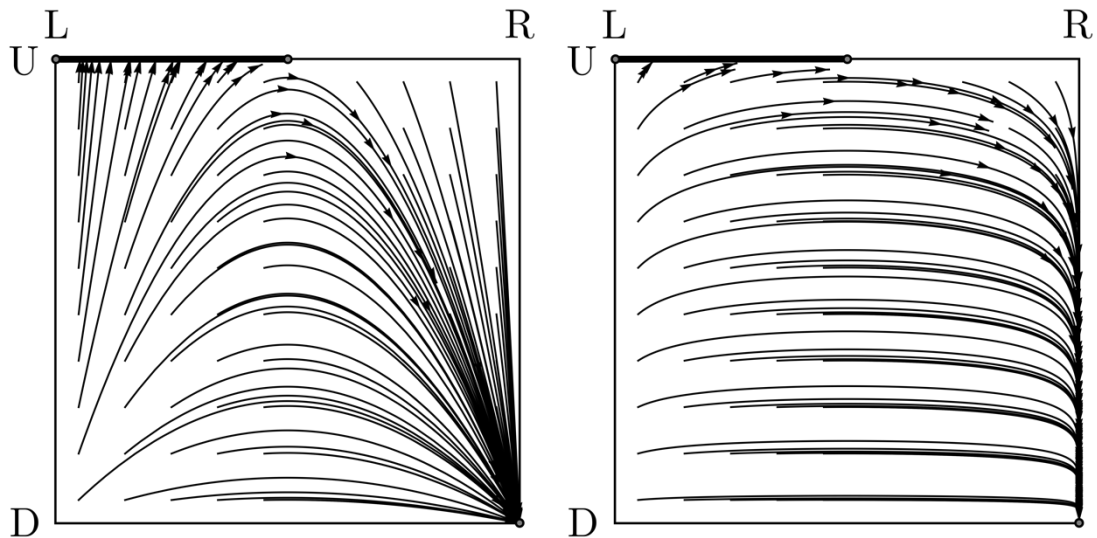
	L	R
U	(0,1)	(0,1)
D	(-1,-1)	(1,0)

(a) The Chain Store Game

	L	R
U	(0,1)	(0,1)
D	(-0.1,-1)	(0.1,0)

(b) The Chain Store Game where row's payoffs have been multiplied by 0.1

---

[7] Cf. the work of Bergstrom and Lachmann (2003) and Bruner (2019) and O'Connor and Bruner (2017).

(c) An illustration of the two population replicator dynamics for the game (a)

(d) An illustration of the two population replicator dynamics for the game (b)

Figure 5: An example of how multiplication of one players payoffs by a positive constant alters trajectories and the relative sizes of basins of attraction for a game. For illustration, the trajectories plotted in (c) and (d) start at identical points.

In sum, then, for ratio scales and interval scales only certain qualitative features of the dynamics remain invariant under the two-population replicator dynamics. The same is not true for the two-population adjusted replicator dynamics[8] in the case of ratio scales, since the factors a_1, a_2 cancel out in each population. For interval scales, the two-population adjusted replicator dynamics can exhibit a wide variety of behaviors depending on the choice of scale.

This result is especially concerning for some uses of the two population replicator dynamics. While the joint outcomes can may be on an interval scale, the individual outcomes of each population cannot be measured using scales that are independent of one another. When the two-population replicator dynamics is used to model two groups in a single species, this may be appropriate. If, however, the two populations are two distinct species one might worry about the appropriateness of this assumption.

5 Inclusive Fitness

A type of transformation that does not correspond neatly to a category of scale type is the transformation of personal, or neighbor-modulated, fitness into inclusive fitness. To calculate neighbor-modulated fitness one simply sums up all the fitness effects an organism is expected to have from their social interactions (relevant to the trait of interest). Then, to arrive at inclusive fitness, rather than multiplying or adding a constant to a measurement of fitness, one subtracts and adds fitness effects in a more complicated way (Hamilton, 1964; Frank, 2013).

---

[8] For that dynamics, just divide the right hand sides of (3) by the respective average population payoffs.

We can look at a simple example to see how this works. Inclusive fitness calculations are often used to determine when *altruistic* traits are favored by selection. These are traits where there is some cost, c, incurred by an organism and some benefit, b, conferred to another organism. The neighbor-modulated fitness of altruism and non-altruism are:

$$f_a = f_0 + P(A_{-i}|A_i) \cdot b - c \tag{5}$$

$$f_n = f_0 + P(A_{-i}|N_i) \cdot b \tag{6}$$

where $P(A_{-i}|A_i)$ is the conditional probability that an altruist interacts with another altruist and $P(A_{-i}|N_i)$ is the conditional probability that a non-altruist interacts with an altruist—these capture the probability each type receives a benefit. Further, f_0 is the baseline fitness that an organism receives. By contrast, the inclusive fitness of these traits (using w now to distinguish from neighbor-modulated fitness) are:

$$w_a = f_0 + [P(A_{-i}|A_i) - P(A_{-i}|N_i)] \cdot b - c \tag{7}$$

$$w_n = f_0 \tag{8}$$

where $P(A_{-i}|A_i) - P(A_{-i}|N_i)$ is the 'relatedness', R, between two organisms.

In this case, the transformation first strips all components which are due to the individuals' social environment (the terms consisting of b and the probability of receiving it) then augments each fitness calculation by the relatedness-weighted benefit or harm the individual causes to their social partners' fitness ($R \cdot b$ for the altruists, who cause a benefit, and 0 for the non-altruists who do not affect their social partner's fitness). In the end, this essentially amounts to subtracting the term '$P(A_{-i}|N_i) \cdot b$' from both neighbor-modulated fitness calculations.

This changes fitness in a non-constant way—the exact effect will depend on the population composition, as $P(A_{-i}|N_i)$ will generally change with the number of altruists—but in certain cases the evolutionary trajectories will be invariant under this sort of transformation. Under certain assumptions, namely *actor's control* and *weak additivity*,[9] this transformation can be shown to give the same prediction for the direction of evolutionary change as the original neighbor-modulated fitness payoffs (Queller, 1992; Birch, 2016).[10] Thus, the location of rest points and their stability properties are preserved. In the context of the replicator dynamics with two possible traits, it has been proven further that (given the above assumptions) both the solution trajectories and the speed at which they are traversed are unaltered by this transformation, when relatedness is constant (Van Veelen, 2011) or when there are pairwise interactions (Rubin, 2018).

---

[9] Roughly, these two conditions correspond to requiring that the fitness effects on the recipient do not depend on the recipient's genotype/phenotype and fitness effects from all an organism's social interactions can simply be added up (Birch, 2016).

[10] These proofs make use of the Price equation, not the continuous time replicator dynamics, but the two dynamics are equivalent when there are a finite number of traits (Rubin, 2018).

This is somewhat surprising because neighbor-modulated fitness and inclusive fitness are not positive affine transformations of one another, but they are monotonic transformations of one another. This might lead one to think of neighbor-modulated and inclusive fitness as on the same ordinal, but not interval, scale. However, it turns out that more of the structure of the replicator dynamics is preserved than with an ordinal scale. In particular, when it comes to the replicator dynamics, it looks more like an interval or even an absolute scale.

Consider a simple example first discussed by Skyrms (1994). Imagine a population where two individuals are paired to interact at random. If an altruist is initially paired with an altruist, the two interact. Any individual (altruist or non-altruist) who initially pairs with a non-altruist can refuse and seek another pairing. They are then paired again, but they cannot refuse the second pairing. This introduces a certain level of correlation. But the exact values of $P(A_{-i}|A_i)$ and $P(A_{-i}|N_i)$ will depend on the proportion of altruists in the population. In particular:

$$P(A_{-i}|A_i) = p + \frac{(p - p^2)(1 - p)}{1-p^2}$$

$$P(A_{-i}|N_i) = \frac{p - p^2}{1 - p^2}$$

With this simple example, it is clear that inclusive fitness and neighbor-modulated fitness are not equivalent. Figure 6 shows how the fitness of altruists and non-altruists compare for various population proportions.

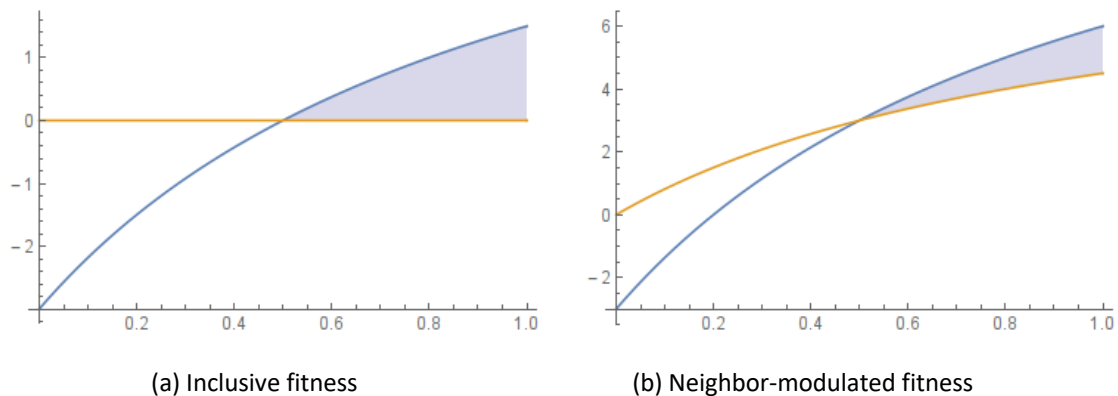(a) Inclusive fitness   (b) Neighbor-modulated fitness

Figure 6: An example of how neighbor-modulated and inclusive fitness change as the population composition changes. In this example, $f_0 = 0, b = 9$, and $c = 3$. The blue line represents the fitness of the altruistic type and the yellow line represents the fitness of the non-altruistic type. The x-axis is the proportion of altruists in the population, p.

Despite the fact that the scales are obviously different from one another, they nonetheless generate identical evolutionary histories in the replicator dynamics (see figure 7).[11] This is because the difference between altruists' and non-altruists' fitness is the same regardless of which one uses neighbor-modulated or inclusive fitness for all population compositions. So, even though we are performing a transformation which is inadmissible when using an

---

[11] For a discussion see Rubin (2018).

absolute scale, evolutionary predictions based on the replicator dynamics have the same invariance properties as when fitness is measured using an absolute scale.
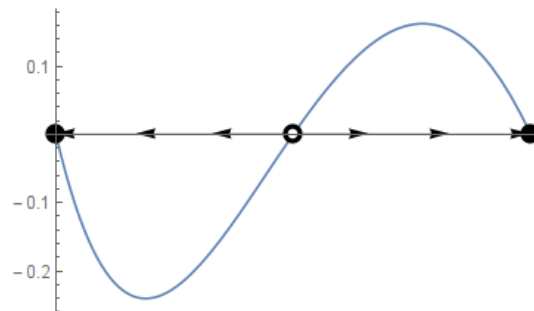

Figure 7: The replicator dynamics for both inclusive fitness and neighbor-modulated fitness in the example where $f_0 = 0, b = 9$, and $c = 3$.

Additionally, since inclusive fitness enters into the dynamics in the same way as any other calculation of fitness (namely, it is used to calculate fitness of each strategy and the average fitness of the population, which are then plugged into the dynamical equations), whatever invariance there is in the dynamics, this invariance exists when fitness is thought of as inclusive fitness. For instance, with the replicator dynamics, one can add or multiply inclusive fitness by any number without changing the solution trajectories. Further, the choice of f_0 (or whether to include it at all) is irrelevant to the evolutionary trajectories as one can add any constant without changing the solution trajectories or the speed at which they are traversed. If one is using the adjusted replicator dynamics, on the other hand, the background fitness can have important effects on the evolutionary trajectories.

6 Conclusion

Let's return to our main question: Given that we wish to preserve certain features of an evolutionary dynamics, what constraints does this put on the underlying fitness function?

Our results entail the following. Consider the standard one population replicator dynamics. In case we wish to preserve both the trajectories and the time scale, fitness must be measured on an absolute scale. If we wish to preserve the trajectories but not the time scale, it is sufficient that fitness be measured on a ratio or an interval scale. Ordinal scales don't give rise to any interesting invariances, unless there is a way to restrict admissible scales to those that respect the ordering given by expected fitness. In the two-population replicator dynamics the invariance of trajectories requires an absolute fitness scale. For ratio and interval scales, trajectories and basins of attraction may change.

The situation is different in the adjusted replicator dynamics. If we wish to preserve both the trajectories and the time scale of the dynamics, it is enough that fitness be measured on a ratio scale. On the other hand, interval scales don't lead to any interesting invariance. They may even lead to changes of basic qualitative features of the dynamics, such as the direction of the vector field.

Not every transformation biologists' use in describing evolutionary change can be captured by one of these measurement scales. In the case of inclusive fitness, described in section 5,

one transforms personal fitness functions in a non-constant way. Yet, in certain cases, this can lead to invariances in the evolutionary dynamics. What assumptions are needed to prove these invariances (and how restrictive they are) depends on whether one cares only about the location and stability of rest points or if one cares additionally about the speed at which the trajectories are traversed.

Many applications of evolutionary theory use an absolute measure of fitness, namely the number of offspring or the expected number of offspring. This leads of course to very strong invariance properties. The disadvantage is that just counting offspring or calculating the expected number of offspring is a very crude concept of fitness in situations in which we wish to determine the contribution of specific traits or strategies to fitness. For instance, Wagner (2010) has proposed a measure of fitness that captures the competitive ability of traits. His approach is based on Luce's theory of choice probabilities (Luce, 1959). Wagner develops an analogous theory for evolution, where probabilities reflect a trait's propensity to out-compete other traits. He proves that under certain conditions this gives rise to a ratio scale measure of fitness. As we have seen, this is associated with strong invariance properties in one-population replicator dynamic models, and preserves qualitative features in two-population models.

In cultural evolution one usually does not have access to measures of fitness with strong scale invariance properties. Applications in economics usually assume that fitness is a standard utility function, hence an interval scale. In these cases, the replicator dynamics has reasonable invariance properties. In applications of evolutionary game theory outside of economics it is often not so clear what fitness is. The dangerous case is that fitness is only thought of as an ordering of outcomes. As mentioned above, without any special restrictions this does not lead to any interesting conclusions regarding the significance of certain population states.

References

Baccelli, J. (2019). Beyond the metrological point of view. Forthcoming in *Studies in History and Philosophy of Science A*.

Bergstrom, C. T., & Lachmann, M. (2003). The Red King effect: when the slowest runner wins the coevolutionary race. *Proceedings of the National Academy of Sciences*, 100(2), 593–598.

Birch, J. (2016). Hamilton's two conceptions of social fitness. *Philosophy of Science*, 83(5):848–860.

Bruner, J. (2019). Minority (dis)advantage in population games. *Synthese*, 196:413–427.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press, Oxford.

Cressman, R. (2003). *Evolutionary Dynamics and Extensive Form Games*. MIT Press, Cambridge, MA.

Frank, S.A. (2013) Natural selection. VII. History and interpretation of kin selection theory. *Journal of Evolutionary Biology*, 26:1151–1184.

Hamilton, W.D. (1964). The Genetical Evolution of Social Behavior I and II. *Journal of Theoretical Biology*, 7:1–16.

Hand, D. J. (1998). *Measurement: Theory and Practice*. Wiley, New York.

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Maß. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physikalische Classe*, pages 1–64.

Houle, D., Pélabon, C., Wagner, G. P., and Hansen, T. F. (2011). Measurement and meaning in biology. *The Quarterly Review of Biology*, 86:3–34.

Huttegger, S. M. and Mitteroecker, P. (2011). Invariance and meaningfulness in phenotype spaces. *Evolutionary Biology*, 38:335–351.

Huttegger, S. M. and Zollman, K. J. S. (2013). Methodology in biological game theory. *The British Journal for the Philosophy of Science*, pages 637–658.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement, Vol. I. Additive and Polynomial Representations*. Academic Press, San Diego. Reprinted by Dover 2007.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York.

Luce, R. D., Krantz, D. H., Suppes, P., and Tversky, A. (1990). *Foundations of Measurement, Vol. III. Representation, Axiomatization, and Invariance*. Academic Press, San Diego. Reprinted by Dover 2007.

Maynard Smith, J. and Price, G. R. (1973). The logic of animal conflict. *Nature*, 246:15.

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

Mitteroecker, P. and Huttegger, S. M. (2009). The concept of morphospaces in evolutionary and developmental biology: Mathematics and metaphors. *Biological Theory*, 4:54–67.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton NJ.

Narens, L. (2002). *Theories of Meaningfulness*. Lawrence Erlbaum and Associates, Mahwah, NJ.

O'Connor, C. and Bruner, J. (2017). Dynamics and diversity in epistemic communities. *Erkenntnis*, 84:101–119.

Okasha, S. (2018). *Agents and Goals in Evolution*. Oxford University Press, Oxford.

Queller, D. C. (1992). Quantitative genetics, inclusive fitness, and group selection. *The American Naturalist*, 139(3):540–558.

Rubin, H. (2016). The phenotypic gambit: selective pressures and ESS methodology in evolutionary game theory. *Biology & Philosophy*, 31(4):551–569.

Rubin, H. (2018). The Debate over Inclusive Fitness as a Debate over Methodologies. *Philosophy of Science*, 85(1):1–30.

Sandholm, W. H., Dokumaci, E., and Franchetti, F. (2012). Dynamo: Diagrams for Evolutionary Game Dynamics. http://www.ssc.wisc.edu/ whs/dynamo.

Savage, L. J. (1954). *The Foundations of Statistics*. Dover Publications, New York.

Skyrms, B. (1994). Darwin Meets the Logic of Decision: Correlation in Evolutionary Game Theory *Philosophy of Science* 61:503–528.

Stevens, S. S. (1946). On the theory of scales or measurement. *Science*, pages 677–680.

Suppes, P., Krantz, D. H., Luce, R. D., and Tversky, A. (1989). *Foundations of Measurement, Vol. II. Geometric, Threshold, and Probabilistic Representations*. Academic Press, San Diego, CA. reprinted by Dover 2007.

Taylor, P. D. and Jonker, L. (1978). Evolutionarily Stable Strategies and Game Dynamics. *Mathematical Biosciences*, 40:145–156.

Van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford University Press, Oxford.

Van Veelen, M. (2011). The replicator dynamics with n players and population structure. *Journal of theoretical biology*, 276(1):78–85.

von Helmholtz, H. (1887). Zählen und Messen erkenntnistheoretisch betrachtet. *Philosophische Aufsätze Eduard Zeller gewidmet*.

Wagner, G. P. (2010). The measurement theory of fitness. *Evolution*, 64:1358–1376.

Weibull, J. (1995). *Evolutionary Game Theory*. MIT Press, Cambridge, Mass.