How to read "heritability" in the recipe approach to natural selection

Abstract

There are two ways evolution by natural selection (ENS) is conceptualized in the literature. One provides a 'recipe' for ENS incorporating three ingredients: variation, differences in fitness and heritability. The other provides formal equations of evolutionary change and partitions out selection from other causes of evolutionary changes such as transmission biases or drift. When comparing the two approaches there seems to be a tension around the concept of heritability. A recent claim has be made that the recipe approach is flawed and should be abandoned. In this paper I show that the tension is only a superficial one. If one uses a concept of heritability strictly in line with the formal equations of evolutionary change, the recipe approach keeps its validity and generality. To demonstrate that the intuitive concept of heritability is not a general one, I use one formulation of the Price equation formulated by Okasha, show that the concept of heritability in his formulation incorporates both the intuitive notion of heritability as a measure of similarity between parent and offspring characters and a measure of persistence. I advocate that for persistence to be incorporated in the concept of heritability used in recipes for ENS in the same way heredity is, show that this is readily attainable and thereby dissolve any point of tension concerning heritability between the recipe and the analytical approach to ENS.

1. Introduction

It is widely acknowledged that for evolution by natural selection (ENS) to occur in a population of entities, three ingredients are necessary. A number of formulations that Godfrey-Smith ([2009]) calls *recipes*, incorporating these three ingredients have been proposed throughout the history of evolutionary biology (for different versions of recipes see Brandon [1990]; Darwin [1859]; Endler [1986]; Lewontin [1970], [1978], [1985]; Okasha [2006]; Ridley [1996]). A generic recipe can roughly be delineated as follows. For ENS to occur, a population of entities must exhibit 1) variation, that

leads to 2) difference in fitness among the variants and that 3) is heritable. The term *evolution* in the expression *evolution by natural selection* usually refers to changes of frequency of variants in a population over time, as classically defined in population genetics. I follow this definition here.

Recipes are supposed to apply not only to biological populations of organisms but to *any* population of entities, as well as any level of organization (Lewontin [1970]). Thus, they represent one way to abstract the Darwinian apparatus. But in recent years, evolutionary theory is undergoing another form of abstraction that relies more heavily on mathematics and has the potential to be applied widely across disciplines from evolutionary biology to economics. I call this approach the *formal* approach. The work of Okasha ([2006]), Rice ([2004]) or Gardner and Grafen ([2009]), to cite only few examples, represent this tendency well. Yet, when looked at superficially, there seem to be several points of tension between the two approaches. One of them is around the meaning of fitness. In fact, the concept of fitness, under the formal approach seems not always to match the notion of "expected reproductive output" as it is classically interpreted in most recipes (Godfrey-Smith [2009]; Ridley [1996]). But fitness will not be directly treated here as I leave it for further work.

Another apparent point of tension between the *recipe* and the *formal* approach, albeit more subtle, is around the concept of heredity, and more precisely one of its statistical measure at the population level, namely heritability. In this paper, I aim to show that the intuitive interpretation of the notion of heritability used in the recipes does not *always* match the notion of heritability used in the formal approach and that when this intuitive interpretation of heritability is used in recipes, it will lead to wrong predictions about ENS in some important cases. I show that under a concept of heritability strictly in line with the formal equations of evolutionary change, this point of tension disappears.

To get there, the remainder of the paper will be divided into 3 sections. In Section 2, I present the different approaches to heritability. I show why the concept of heritability is preferable to the concept of heredity, why the two notions of heritability, namely the variance approach and the regressions approach can largely be considered as equivalent and yet in spite of their equivalence, the regression approach remains more convenient to use for my purpose. From there I present the formulation of the Price equation proposed by Okasha in his 2006 book that makes explicit the notion of heritability. The Price equation will be the formulation of the formal approach to ENS used throughout the paper.

In Section 3, I examine what looks to be a problem for recipes for ENS identified by Earnshaw-Whyte ([2012]). According to Earnshaw-Whyte, one can find cases of ENS in which there is no heritability in the sense defined in Section 2 (i.e. the regression one), yet ENS is observed if one refers to the classical meaning of evolution in population genetics. He proposes a verbal model in which there is ENS but no heritability as classically interpreted and that the recipe approach, under his reading, is unable to account for. This leads him to conclude that the classical concept of ENS is irremediably flawed and should be abandoned. I claim that his view is too strong. Under an uncharitable reading of Earnshaw-Whyte's model, an intuitive interpretation of the recipes leads to the correct prediction. Under a more charitable reading of his model, this intuitive interpretation leads to wrong predictions of evolutionary change as Earnshaw-Whyte claims. I show however, that the only difference between the two readings of Earshaw-Whyte's model depends on an interpretation of which individuals in the model are considered as parents and which ones are considered as offspring. I claim that this difference in interpretation should not change our claims about heritability and ENS.

In Section 4, in light of the two possible readings of Earnshaw-Whyte's model, I propose that a correct interpretation of the notion of heritability includes both the offspring of the parental population and the parents that have survived after reproduction (if any) and present in the offspring population. I call this interpretation *temporal* heritability. Further on, I show that this interpretation is in accordance with the Price equation since the Price equation can describe the evolutionary change of populations of individuals over an arbitrary period of time and in which selection involves differential persistence instead of differential reproductive rates (Rice [2004], p. 169).

Finally, starting from the formulation of the Price equation given by Okasha and presented in Section 2, I show that temporal heritability can be separated into two different components: one I name *generational heritability* which corresponds to the intuitive concept of heritability and *persistence heritability* usually neglected in the recipe approach. I show that persistence heritability and generational heritability can alternately be neglected in some cases (e.g. non-overlap of generations, no reproduction), but that in most real biological populations they both need to be taken into account to predict accurate evolutionary change. Overall I concur with Bouchard ([2004], [2008], [2011]) that differential persistence is an important factor for evolution by natural selection, although we will see that the notion of heritability I use is slightly different from that of Bouchard. More recently Bourrat ([in press]) using computer simulations and Doolittle ([in press]) in a more verbal fashion have shown the importance of differential viability or persistence for reasoning about natural selection.

2. Heritability in the Price equation

2.1. Different approaches to heredity and heritability

As we have seen in the Introduction, recipes for ENS include heredity between parents and offspring as a necessary condition. But what does it mean to have heredity between parents and offspring? In its most general sense, it simply means that there is the transmission of traits between parent and offspring associated with a parent-offspring resemblance (Godfrey-Smith [2009], 24). One problem with the concept of heredity, when transposed to recipes for ENS, is that heredity is a qualitative concept. As such, it does not allow quantitative predictions of evolutionary changes due to natural selection at the population level. To make these predictions, different statistical measures of heredity, regrouped under the term of *heritability*, have been developed.

Heritability is usually defined as "a ratio of variances, specifically as the proportion of total variance in a population for a particular measurement, taken at a particular time or age, that is attributable to variation in additive genetic or total genetic values"(Visscher, Hill, & Wray [2008], p. 255). In a genetic context, it can thus either refer to the part of the variance of a phenotype in a population attributable to total genetic variance *or* to additive genetic variance. In the first case heritability does not presuppose any heredity mechanism of phenotypes. This is the notion of heritability preferred by behavioral geneticists and psychologists who are interested in knowing the contribution of genes to psychological traits (Downes [2009a]). It is usually labeled broad-sense heritability (H^2). But broad sense heritability is not the notion employed in the context of natural or artificial selection. In fact, in those later contexts, for heritability to be positive, there must be the transmission of the trait over time (usually from parents to offspring). Since genes only transmit their additive effect between generations, only the additive genetic effects are taken into account in this context. The notion of heritability relevant in this context is narrow-sense or realized heritability (h^2). But within evolutionary theory another, more abstract definition is often preferred. Heritability can be defined as "the linear regression of average offspring character on parental character, or mid-parental character if reproduction is sexual" (Okasha [2006], p. 34). This notion of heritability used

among others by Godfrey-Smith ([2007], [2009]), Okasha ([2006], [2010]), Rice ([2004]) and Roughgarden ([1979]).

Heritability can thus correspond to three distinct notions (Jacquard [1983]; Sesardic [2005]): 1) broad-sense genetic heritability, 2) narrow-sense or realized genetic heritability using the ratio of variance approach, henceforth the variance approach, and 3) heritability as a measure of parent-offspring phenotypic resemblance, henceforth the regression approach. Importantly, both the second and third notions of heritability correspond to realized heritability, the only form of heritability that matters for selection, except that the third notion can be applied to cases different from the biological ones, such as non-genetic inheritance (e.g. cultural evolution). Both notions are thus used in evolutionary theory and there are disagreements over which one of them, if any, is the most general one. Some authors consider that the regression approach is more fundamental than the ratio of additive genetic variance approach (Okasha [2006], [2010]; Rice [2004]), while others disagree. For example, Downes ([2009b]) proposes that the notion of heritability to be used should depend on one's agenda. In some cases, he argues, one of the two definitions might be more appropriate.

There are different reasons to think that the phenotypic regression approach is more general than the genetic variance approach. One is that genotypes can be treated as phenotypes (Rice [2004], p. 165). Another is that the Darwinian apparatus can be applied to a different substrate from genetic (e.g. culture). Using a definition of heritability relying on additive genetic variance will only be problematic in those cases. Finally, Rice ([2004], pp. 204-05) and Godfrey-Smith ([2007], p. 510) both argue that in some significant cases of evolution, heritability can be negative or greater than 1. These values cannot be obtained if one used the genetic variance notion of heritability where the values of heritability are comprised between 0 and 1 (included). Yet, there are cases, for example

when there is a correlation between the environment of the parents and the offspring in the population, where regression heritability might be positive but the positive value obtained is due solely to a correlation between the environment of the parents and the offspring. If one uses the regression notion of heritability in those cases, they might conclude misleadingly that the similarity between parent(s) and offspring is due to the causal role of the parent(s) when it is in fact due to the causal role of the environment. Using the variance approach does not lead to this kind of problems (at least in theory).

We can thus see that it is hard to choose between the two approaches since they both have advantages and limitations. That said, in most cases both notions will be equivalent *modulo* the assumptions that the environments of the parent(s) and offspring are not correlated and that different entities can have in some cases the same role as genes. Yet, the regression approach seems overall less rigid than the variance one, because it can readily accommodate any inheritance system. Thus, the rest of the paper will dispute the idea that the existence of parents and offspring is necessary for evolution by natural selection, the approach to heritability I will initially use will be based on the regression approach, keeping in mind its limitations.

2.2. Heritability: from recipes to the Price equation

In the first chapter of his 2006 book, Okasha proposes a bridge between the recipe and the formal approach to ENS. His aim is to formulate the Price equation in a way that will make apparent the three ingredients of the recipes for ENS. The Price equation is a simple statistical algebraic formula proposed by George Price (Price [1970], [1972]) that can be interpreted as describing the evolutionary change on a character of a population between two times. One important aspects of the equation is that it relies on no contingent biological assumptions. For that reason it always holds

true (Frank [1998]; Okasha [2006], p. 19). Both the generality and the statistical nature of the Price equation are usually considered as virtues, but the cost to pay is that applying a biological or causal interpretation using the equation can be a hard task.

The Price equation is usually formulated as follows:

$$\overline{w}\Delta\overline{z} \;=\; Cov\,(w, z) \;+\; E(w\Delta z) \qquad (1)$$

where $\Delta\overline{z}$ is the change in average character value from one generation to another; $\Delta\overline{z} = \overline{z}' - \overline{z}$ with $\overline{z}$ being the average value of the character z at the parental generation and $\overline{z}'$ the average value of the character z at the offspring generation; $\overline{w}$ is the average fitness in the population; $\Delta\overline{z}$ is the evolutionary change in character; $Cov\,(w, z)$ is the covariance the between absolute fitness w and character z of each individual; $E(w, \Delta z)$ is the average, or expected value of the quantity $w\Delta z$, i.e. the product fitness x transmission bias. If reproduction is perfect, $\Delta z = 0$ because z and z' are equal (for more details on the derivation of the Price equation see Okasha [2006], pp. 18-39). But for our purpose we will start, following Okasha ([2006]), with another formulation of the Price equation by dividing both sides of equation (1) by $\overline{w}$. This yields:

$$\Delta\overline{z} \;=\; Cov\,(\omega, z) \;+\; E_w(\Delta z) \qquad (2)$$

ω is the relative fitness ($\omega = w/\overline{w}$) and $E_w(\Delta z)$ is the fitness-weighted average of the quantity $\Delta\overline{z}$.

As can be seen, there is no heritability term h^2 in (2). But Okasha proposes a formulation of the Price equation starting from equation (2) that will make h^2 explicit. After some mathematical manipulations, which are detailed in the Appendix, Okasha gets the following formulation:

$$\Delta\overline{z} \;=\; h^2 Cov\,(\omega, z) \;+\; E(\Delta z) \qquad (3)$$

where h^2 is heritability defined as the linear regression of average offspring trait z' on parental trait z. According to Okasha ([2006]), the term $h^2 Cov\,(\omega, z)$ can be interpreted as the change on character z due to selection, and the term $E(\Delta z)$ as the change on character z that would have occurred anyway without selection. This is the interpretation I follow here. If the term h^2 is nil, then no evolutionary change due to natural selection will result. The term $Cov\,(\omega, z)$ embeds the two other ingredients of recipes for ENS namely difference in fitness and variation (Okasha [2006], p. 36), so that in accordance with the recipes if $Cov\,(\omega, z) = 0$, then no ENS is observed.

Finally, one difference between (2) and (3) is that in (3) something must be said about Δz in the second term $E(\Delta z)$ of equation (3) in the case a given individual has no offspring (fitness nil). In such a case, it is not immediately clear what value z' should be given in the second term of the RHS. Equation (2) is immune to this problem because the second term is weighted by fitness. By convention, in equation (3), when an individual does not have any offspring, the value given to the character z' will be the value of the average character of the offspring population ($\bar{z}'$). A possible justification of this convention is that whether an individual has some offspring with, on average, the same value on the character than the average value on this character in the whole offspring population or whether it has no offspring, the evolutionary consequences at the population level are perfectly equivalent. Leaving aside equation (3) for the moment, I turn now to what Earnshaw-Whyte (2012) has claimed to be a major flaw for the recipe account of ENS.

3. A problem for the recipes?

Let us suppose a population of entities of two types A and B that reproduce synchronically with the particularity that each entity whether it is of type A or type B has a 50% chance to produce an

offspring of type A and 50% chance to produce an offspring of type B, so that there is no heritability on the type as it has been defined so far under the regression approach,[1] i.e. the slope of the linear regression of average offspring type on parent type is 0. Let us postulate that both types have the same fertility of 1 offspring every unit of time if they survive until reproduction, but that type A has on average a viability of 0.70 every unit of time whereas B has only a viability of 0.30. This is the model proposed by Earnshaw-Whyte ([2012], p. 398).

Under an intuitive reading of the recipes, no ENS should be observed because there is no heritability on the type. However, that is not what is observed. Figure 1 represents a simulation of this model (Model 1) using the software NETLOGO 5.02 (Wilensky [1999]) classically used for individual-based modeling. At t=0, 5000 individuals (2500 of type A and 2500 of type B) are created. Then each individual is asked its type. If the type is A, the individual has a 70% chance of survival until the next unit of time. If the type is B, the chance is only 30%. This represents the *selection* stage of the model. From there, all the surviving individuals are asked to reproduce 1 individual with 50% chances of producing an individual of type A and 50% chances of producing an individual of type B. This represents the *reproduction* stage of the model. Finally, after reproduction, if the population size is larger than 5000, individuals are killed randomly until the population reaches its initial size. I call this stage *check-for-overcrowding*. This last stage represents the only difference with Earnshaw-Whyte's verbal model in which the population size is unlimited. Computational limitations do not allow for simulation with unlimited population size. That has no important consequence for purpose since unlimited population size is an unrealistic assumption and the large size of the population makes drift extremely weak. Once this stage is over, one unit of time is counted and the sequence *selection* → *reproduction* → *check-for-overcrowding* restarts indefinitely for each surviving entity unless the population goes extinct.

---

[1] Some complex inheritance mechanism is supposed to produce this pattern

The graph on figure 1 shows that contrary to what is expected upon an intuitive interpretation of the recipes (i.e. that the frequency of A and B remain stable around 0.5), the frequency of type A increases up to an equilibrium which is reached after t=7, roughly equals to 0.62 while the frequency of type B decreases to 0.38.
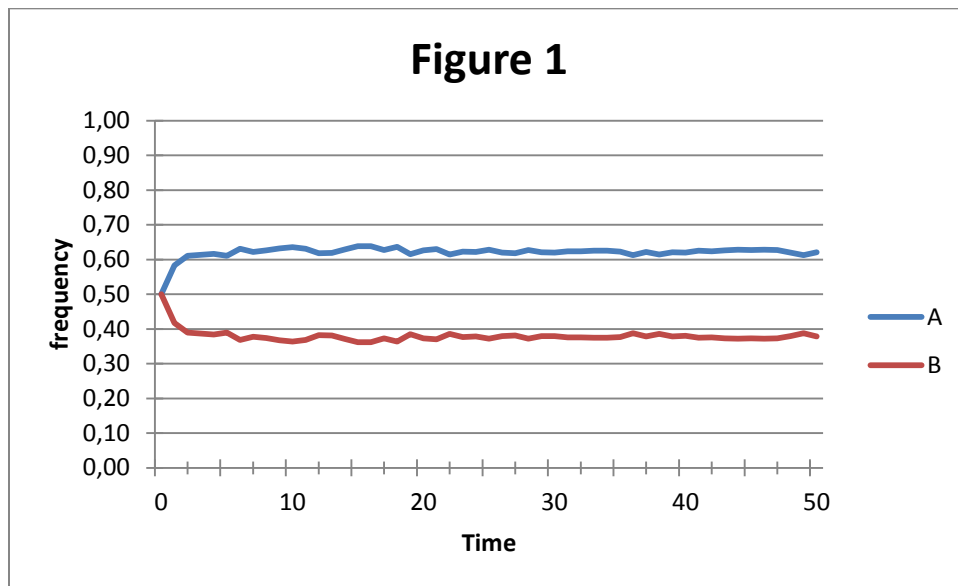

Figure 1: simulation of Model 1 over 50 units of time

The evolution observed is limited, yet we can confidently assume that it is a genuine case of ENS since the only difference between the two types is a systematic difference in fitness (more precisely in viability) within the same environment and thus attributable only to natural selection. Let us recall that a textbook definition of evolution in population genetics is a change of frequency of a given variant in the population. This very simple model shows therefore that there can be ENS in spite of no heritability as classically interpreted under the regression approach. I need to emphasize here that the phenomenon observed in the model does not in itself represent an oddity. Indeed, as it

will become clearer in the next section, the same phenomenon can be observed in populations of organisms that do not reproduce such as those described by Bouchard ([2008], [2011]).

If ENS is observed in this model, does that mean therefore that the classical formulation of ENS is flawed? That is what Earnshaw-Whyte claims. In his own words, "th[e] classical account of ENS is flawed and should be abandoned" ([2012], p. 97). In my view his position is far too strong, since as I will demonstrate, it is the interpretation of the recipes for ENS given by Earnshaw-Whythe that is faulty, not the recipes.

For one thing, Earnshaw-Whyte's model can be formalized in two different ways. These two ways depend entirely on what is understood by "parental" and "offspring" entities. Yet, from an evolutionary perspective they are perfectly equivalent. Under a charitable interpretation of his model, every individual produces one offspring if it survives until the reproduction stage (See figure 2a). Under such interpretation, there is a clear parent, a clear offspring and no heritability defined as the slope of regression of average offspring character on parental character. Importantly under this interpretation, there is an overlap of generations. Under a less charitable interpretation for Earnshaw-Whyte's agenda, the fertility of one offspring with survival of the parent and subsequent same probability of selection is understood as an entity reproducing asexually such as a bacterium for instance (See figure 2b). The parental entity ceases to exist when the two offspring are produced. There is therefore no overlap of generations. From one bacterium of type A, we can get either two bacteria of type A or one bacterium of type A and one bacterium of type B with equal probability. With this second interpretation, there is heritability in the population as classically interpreted since from one parent of a given type, the two offspring resulting from reproduction are on average more similar to their parent than two individuals chosen randomly at different generations in the population. This demonstrates clearly that the classical concept of ENS is not flawed in light of

Earnshaw-Whyte's model. Rather it suggests that the concept of heritability is not as straightforward as typically assumed in the recipes and that the notion of parent and offspring can be problematic for the recipe approach if not interpreted carefully.
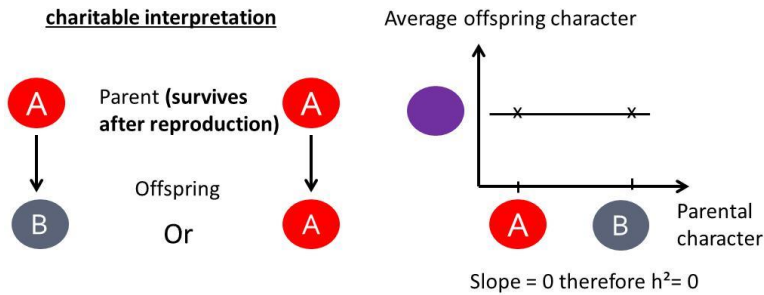

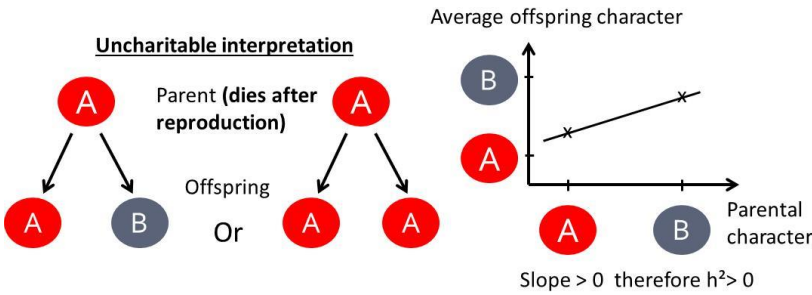*Figure 2a: illustration of Earnshaw-Whyte's model charitably interpreted*

Figure 2b: illustration of Earnshaw-Whyte's model uncharitably interpreted

A more convincing and realistic model for the point Earnwhaw-Whyte is attempting to make would be a case of age-structured populations with individuals of different ages with different viabilities and fertilities (Charlesworth [1994]; see also Godfrey-Smith [2007] for a short introduction). Thus, let us suppose, for the sake of argument, that there is an obvious or objective distinction between parent and offspring such as in parthenogenetic organisms or those that reproduce by budding as in the case of yeasts so that we therefore are in a clear-cut case of a population with overlapping generations and can interpret charitably Earnshaw-Whyte's model. In

such cases the notion of heritability used so far does not seem to capture, what is indubitably a case of ENS. Yet, I do not think that the tripartite description of ENS should be abandoned even in these cases. Instead, I will argue that in light of the equation (3) given earlier, we should reinterpret the concept of heritability in recipes for ENS to make it more rigorous.

4. Reinterpreting Heritability for Recipes

The Price equation, as I said in Section 2, predicts the evolutionary changes of a given character over time. One of its important features is that although one generation is usually the pace of time over which evolutionary change is measured, that is not at all necessary. In fact, any arbitrary period of time can be used (Rice [2004], p. 169). This remark is extremely important because it means that evolutionary change can be measured over a period of time smaller than one generation, or even with entities that do not reproduce. Thus when viability selection is the only type of selection present in a population, a *temporal* interpretation of the Price equation can be used (Okasha [2006], p. 24).

The claim of the possibility of a temporal interpretation of the Price equation when the entities do not reproduce immediately raises the question of what account of heritability to give in these cases[2] in equation (3). Recall that we defined heritability as the linear regression of average *offspring* character on *parental* character. But in these cases, we do not have any parent or offspring. We only have the same individuals over time, minus those which, if any, died at a later time. There seems thus to be some inconstancy between the view that the existence of heritability, as classically understood under the regression approach, depends on the existence of generations and the view

---

[2] This is also obviously true for the concept of fitness, but I said earlier that I leave this problem for further work.

gained from the Price equation and true by definition, that heritability can be defined without the existence of generations, as displayed in equation (3).

At that point, one could either deny that the Price equation provides a valid description of evolutionary change or that the concept of heritability used classically is not general enough or inexact. The first option seems unlikely. In fact, making the claim that the Price equation does not represent a valid description of evolutionary change would fly in the face of some of the most recent developments in evolutionary theory. I will thus adopt in this section the second option and aim at developing a more inclusive notion of heritability that will accommodate Earnshaw-Whyte's model as well as the notion of ENS in populations of non-reproducing entities (Bouchard [2008], [2011]; Bourrat [in press]; Doolittle [in press]) and ENS in populations with overlap of generations.

Starting from the Price equation, to predict the change of a given character over time in a population that does not reproduce between t_1 and t_2, we must consider for each individual at t_1 that survives until t_2, that they are their own offspring. Once this is done, the slope of linear regression between the character z of all the surviving entities between t_1 and t_2 obtained becomes positive (provided that individuals retain their character z for their whole life). From an evolutionary perspective, this means that persisting can be seen as formally equivalent to reproducing. A correct interpretation of heritability to be used for recipes for ENS should thus include a component of persistence. My use of the term *persistence* should be distinguished from that of Bouchard ([2008], [2011]). For Bouchard, persistence means persistence of lineages through time. Here persistence means persistence of the entity between two times (for example between two generations). For biological organisms, persistence means survival. But I prefer to use the more general term of persistence to be able to account for cases of ENS in domains other than the biological one.

Up to this point, I have only shown that persistence should be taken into account in our definition of heritability in populations that do not reproduce. Yet, some readers might remain unconvinced that a component of persistence should be incorporated to our interpretation of the concept of heritability since they consider any population of entities that does not reproduce as an odd or negligible case of ENS and that the classical interpretation of heritability as the slope of the regression line of the average offspring on parental character is perfectly valid when entities reproduce. Although, I disagree that non-reproducing populations should be dismissed as negligible or unimportant, to make my account work, I need nonetheless to show that persistence is also an important component of heritability in populations that do reproduce.

If our goal is to describe accurately the evolutionary change resulting from the reproduction of the member of a population, persistence will only be relevant if there is overlap of generations. In fact, some parents need to survive after they reproduce if one wants to talk about persistence after reproduction. Interestingly, Okasha makes the assumption of no overlap of generations in most of his treatment of the Price equation, but he clearly states that it is made for simplification purposes only and that the Price equation can be derived even when this assumption is relaxed ([2006], p. 20 footnote 11). Making this assumption along with the assumptions of synchronicity in reproduction, asexuality and perfect transmission is a common move made in the philosophy of biology and more generally in evolutionary theory. Yet, one should be aware of the limitations they impose on the interpretation of the concept of heritability (but also fitness) in an evolutionary context. The demonstrations I provide below can be regarded both as an extension of Okasha's work when the assumption of non-overlap of generations is relaxed as well as an insight on what our concept of heritability should look like once less simplifying assumptions are made about the population. To visualize the problem more clearly, let us turn back to Earnshaw-Whyte's model in which entities do reproduce and in which there is overlap of generations (under the charitable interpretation).

16

As we have already seen, in Earnshaw-Whyte's model, if heritability is interpreted as a case of non-overlapping generations, the slope of the regression between the parent character and the offspring character is nil. This is because both types produce 50% of offspring of each type. But this leaves out parental entities which have survived one or many events of reproduction and which 1) are still present in the population and 2) have the same chance of survival and reproduction than the offspring entities produced at the previous generation.[3] These surviving entities must therefore be incorporated in the linear regression in the same way that offspring are since if that is not done, we cannot pretend that the predictions made at the population level are correct. This is because one part of the population is not taken into account after evolution has occurred. Once the surviving individuals are taken into account in our measure of heritability between the population measured before and after reproduction (assuming there is synchronicity in reproduction between all the members of the population and that the surviving individuals do not mutate), the slope of the linear regression becomes positive, and the correct predictions of evolutionary change between the two times can be made. This shows that the classical interpretation of heritability is insufficient to accurately predict evolutionary change even in populations of reproducing entities.

Let us sum up what has been argued thus far. ENS can be observed in a number of cases. One of them is described accurately by the intuitive interpretation of heritability in the recipes. This is the case when there is heredity and variation in fitness in a population of reproducing entities with non-overlapping generations. In this case the classical and intuitive concept of heritability defined as the slope of the linear regression between the average offspring character on the parental one will yield the correct prediction of ENS. I relabel the classical and intuitive concept of heritability as *generational* heritability. In other cases, there is ENS but the intuitive interpretation of recipes either

---

[3] The fact that the surviving parents and offspring have the same chance of survival and reproduction at the next generation is a peculiarity of Earnshaw-Whyte's model. Different viability assumptions could be made without invalidating the points I am making.

misses that there is ENS at all or makes inexact predictions about the extent to which there is ENS. The intuitive interpretation of the recipes will miss ENS when there is no generational heritability in the population in spite of reproduction[4] but there is persistence of the trait in the surviving entities (this is Earnshaw-Whyte's model). It will also miss it when there is no reproduction. Finally it will make inexact predictions when there is both persistence of the parental trait between generations (via surviving parents) and generational heritability in the population. In those cases heritability must be interpreted with a component of persistence instead of only reproduction. I call this component *persistence heritability*[5] which I define as the linear regression of average parental character in the offspring generation – or at a later stage if entities do not reproduce – on parental character in the parental generation. While the value of generational heritability between two times depends on mechanisms of heredity, the value of persistence heritability depends on mechanisms preventing the mutation of entities into entities of other types.

To give an example of how inexact predictions can be made by using generational heritability and neglecting persistence heritability, we can imagine a population of two types A and B reproducing with perfect inheritance. Under the intuitive interpretation of heritability there is a heritability of 1 on the character 'type' in this population. As in Model 1 presented in Section 3, there is the same difference in viability between A (0.7) and B (0.3). However in the present case, both A and B, if they survive after reproduction, mutate with a probability of 0.5. If that happens, the entity changes its type (i.e. becomes A if it was B and B if it was A) and takes all the properties of its new type. The prevalence value of 0.5 for the mutations is deliberately unrealistic, since my

---

[4] Assuming for simplicity that reproduction is synchronic between all the entities of the population
[5] Persistence and persistence heritability are two different concepts. The first one refers to an individual while second one refers to a population in the same way heredity refers to an individual and heritability to a population.

goal here is only to make a general point on the role of persistence for ENS, not to match a specific target system.

An intuitive interpretation of recipes for ENS predicts that because there is perfect heritability, the type A should invade the population. However that is not what we observe if this model is run. Figure 3 represents a simulation of this model (Model 2).[6] When compared to model 1, there are two differences. First, the transmission of types between parent and offspring is perfect. Second, a *mutation* stage is added just after the reproduction stage. During that stage each parental entity has a 50% chance to change its type and thus become a member of the other type, taking all the properties of its new type.[7] Thus the sequence undergone by each entity every unit of time is the following: *selection* → *reproduction* → *mutation* → *check-for-overcrowding*. At t=0, 2500 individuals of each type are created.
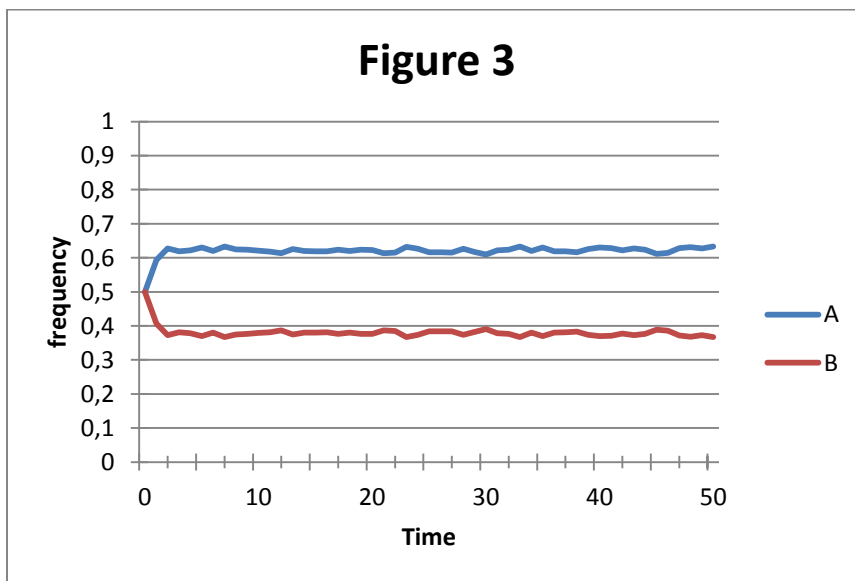

Figure 3: simulation of Model 2 over 50 units of time with heritability persistence = 0.5

---

[6] NETLOGO 5.02 has been used to run this simulation
[7] Importantly newly produced offspring are not concerned by the mutation stage. They will only be concerned by it if they survive the next selection stage

As can be seen on the graph in figure 2, the prediction made by the intuitive interpretation of the recipes that relies on generational heritability is wrong simply because type A does not invade the population. In fact, the pattern of evolution observed is strictly the same as the one observed in model 1. Why is that the case? This is because persistence heritability and generational heritability play the same role for ENS. Simply put, in Model 1 persistence heritability is maximal while generational heritability is nil. In Model 2 the contrary holds: persistence heritability is nil while generational heritability is maximal.

This demonstrates that a correct interpretation of the concept of heritability for measuring evolutionary changes must take into account the transmission over time of a trait in a given population without discrimination of the origin of the transmission i.e. either coming from persistence or/and from reproduction. I propose to label this concept: *temporal heritability*. Thus (temporal) heritability should not be defined as "the linear regression of average offspring character *z′* on parental character *z*" but as "the linear regression of the average character *in the offspring population z′* on the character in the parental population *z*" or again as "the linear regression of the average character *z′* at time t_2 subsequent to t_1 on the character *z* at time t_1." In order to formally ensure this claim and anchor it in the existing literature, I provide in the next paragraph a formulation of the Price equation in which I decompose *h²* presented in equation (3) in two components that makes the links between temporal, generational and persistence heritability apparent.

We start with the form (3) of the Price equation given by Okasha ([2006], p. 35)[8] and presented in section 2:

---

[8] For the purpose of the demonstration I have added the indices which were dropped in Section 2

$$\Delta \bar{z} = h^2 \, Cov \, (\omega_i, z_i) \; + E(\Delta z_i) \qquad (3)$$

By standard least-squares theory, it follows that:

$$h^2 = \frac{cov \, (z_i{}', z_i)}{Var(z_i)} \qquad (4) \text{ (Okasha [2006], p. 34)}$$

For Okasha, h^2 is defined as generational heritability, i.e. "the linear regression of average offspring character z' on parental character z". As I have argued, if there is overlap of generations or no reproduction, this represents an error of interpretation of what heritability means in the context of the Price equation. Rather the definition to be used should correspond to temporal heritability, i.e. "the linear regression of the average character *in the offspring population* z' on the character in the parental population z". At generation $n+1$, the offspring population is composed of both the offspring of the parent(s) of the generation n and the surviving individual(s) of the generation n at the generation $n+1$.

We can decompose z' which represents the character value for each individual i in the offspring population as follows:

$$z'_i = \frac{(s_i z_i + n_i z'_{i_o})}{(s_i + n_i)} \quad (5)$$

$s_i z_i$ is the part of the average offspring-population character value of individual i due to the possible survival of this individual, where s_i is an indicator variable for surviving/dying.[9] $n_i z'_{i_o}$ is the part of the average offspring-population character value of individual i due to the offspring produced by i, where n_i is the number of offspring produced by i and z_{i_o} the average character value of the offspring.[10] If an individual i has no offspring and does not survives between the two times $(s_i + n_i = 0)$, by convention the value of the character z' will be the value of the average character of the

---

[9] The value of the character is assumed here to remain constant in one given individual, but that is not a necessary assumption

[10] Reproduction is supposed to be asexual

offspring population ($\bar{z}'$) comprising both the surviving individual of the parental population and/or the offspring produced by the parental population.[11] The justification of this convention is the same as the one given at the end of section 2: whether the contribution of an individual at a later time on character z' is nil or amounts exactly to the average value of character at the (offspring) population level, does not make any evolutionary difference.

If we replace (5) in (4), dropping the indices i for convenience, we find that:

$$h^2 = \frac{cov\left(\frac{sz + nz'_o}{s+n}, z\right)}{Var(z)} = \frac{cov\left(\frac{sz}{s+n}, z\right) + cov\left(\frac{nz'_o}{s+n}, z\right)}{Var(z)} = \frac{cov\left(\frac{sz}{s+n}, z\right)}{Var(z)} + \frac{cov\left(\frac{nz'_o}{s+n}, z\right)}{Var(z)} \tag{6}$$

$\frac{cov\left(\frac{sz}{s+n}, z\right)}{Var(z)}$ is the regression coefficient of the parental character in the offspring population sz on parental character z in the parental population, which is the definition of persistence heritability.

$\frac{cov\left(\frac{nz'_o}{s+n}, z\right)}{Var(z)}$ is the regression coefficient of the average offspring character z'_o in the offspring population on parental character z, which is the definition of generational heritability.

Each of these two coefficients can be nil. When we replace (6) in (3) and we get:

$$\Delta\bar{z} = \left(\frac{cov\left(\frac{sz}{s+n}, z\right)}{Var(z)} + \frac{cov\left(\frac{nz'_o}{s+n}, z\right)}{Var(z)}\right) Cov\left(\omega, z\right) + E(\Delta z) \tag{7}$$

This is a formulation of the Price equation separating h^2 into generational and persistence heritability. The demonstration provided here shows how generational heritability and persistence heritability are commensurable. $E(\Delta z)$ is still interpreted as the change on character z that would have occurred anyway without selection, but this change can occur both by a transmission bias of

---

[11] The same convention will be used in equation (7) for the term $E(\Delta z)$

the character from parent to offspring or a transmission bias of the character from one and the same individual over time by some form of mutation.

5. Conclusion: Why Should We Care?

In this paper, I have presented a different, less intuitive concept of heritability. Both persistence heritability and generational heritability (i.e. the intuitive conception of heritability) are entailed in this concept that I have called temporal heritability. I have argued that temporal heritability is the right interpretation of heritability to use when talking the most generally about ENS and thus the one to be used in the recipe approach to ENS.

One might object at this point, that the benefits of interpreting heritability as temporal heritability instead of generational heritability in recipes do not out-weigh the costs of doing so on the grounds that the difference between generational and temporal heritability is not important in most cases of ENS. I disagree. On the one hand, persistence alone does not straightforwardly allow the same kind of complex adaptations observed in Darwinian population in which there is genuine reproduction. This point is intuitively obvious but for a detailed account of the reasons why that is the case, see Bourrat ([in press]). On the other hand, persistence is certainly of importance in organisms that do not reproduce or for which reproduction and growth are hard to distinguish (see Godfrey-Smith [2009] and Bouchard [2008], [2011] for examples). For all these cases persistence heritability will be of importance. Evolutionary biology and philosophy of biology have for too long focused solely on organisms that resemble us and marginalized cases that do not fit them. But there are many species for which a persistence account of heritability can be given. This is usually not done because although, in practice, the concept I defined as temporal heritability is the one used in evolutionary theory (the formal equations are exact), everything is done *as if* the concept of generational

heritability was the one used. Earnshaw-Whythe and to some extent Okasha are not the only ones to use a generational interpretation of the concept of heritability when they should use a temporal one. I have found at least four other instances of this in the literature (Gillespie [2004]; Godfrey-Smith [2009]; Hamilton [2009]; Jacquard [1983])

To this can be added that the concept of temporal heritability might be useful to explore in a number of domains, such as the major transitions in evolution or more generally with the multilevel approach to selection, since persistence at one level can be seen as a birth-death process at another. For example, when a multicellular organism survives, almost all its cells are replaced over time. One might thus think that commensurability between persistence and reproduction is possible. Furthermore, in a context in which the concept of ENS is explored by domains aside from biology, it can be interesting to have at hand a general interpretation of the concept of heritability that does not impede the use of less general ones when necessary. Finally, using the concept of temporal heritability rather than generational heritability has important consequences for the concept of fitness. As can be seen in equation (3) or (7), when there is no reproduction in the population or less extremely, when there is overlap of generations, the fitness of an individual cannot straightforwardly be the number of offspring produced, because in some case there is no offspring, but only an individual surviving or both offspring and surviving parents. This calls for a different and more general concept of fitness. Van Valen ([1976]), Earnshaw-Whyte ([2012]), Bouchard ([2008], [2011]) and Bouchard and Rosenberg ([2004]) as well as Matthen & Ariew ([2002]), although in different terms, have made some steps towards that direction, but a lot remains to be done.

Pierrick Bourrat
The University of Sydney,
Faculty of Arts,
School of Philosophical and Historical enquiries

Department of Philosophy
Main Quadrangle A14
Sydney, NSW 2006, Australia
Email: p.bourrat@gmail.com

Reference

Bouchard, F. [2004]. *Evolution, fitness and the struggle for persistence*. Unpublished PhD diss., Duke University.

Bouchard, F. [2008]. Causal processes, fitness, and the differential persistence of lineages. *Philosophy of Science, 75*(5), 560-70.

Bouchard, F. [2011]. Darwinism without populations: a more inclusive understanding of the "Survival of the Fittest". *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 42*(1), 106-14.

Bouchard, F., & Rosenberg, A. [2004]. Fitness, probability and the principles of natural selection. *The British Journal for the Philosophy of Science, 55*(4), 693-712.

Bourrat, P. [in press]. From survivors to replicators: evolution by natural selection revisited. *Biology & Philosophy*, 1-22.

Brandon, R. N. [1990]. *Adaptation and environment*. Princeton: Princeton University Press.

Charlesworth, B. [1994]. *Evolution in age-structured populations (2nd edition)* (Vol. 2): Cambridge University Press Cambridge, UK.

Darwin, C. [1859]. *On the origin of species by means of natural selection*. London: J. Murray.

Doolittle, W. F. [in press]. Natural selection through survival alone, and the possibility of Gaia. *Biology & Philosophy*, 1-9.

Downes, S. M. [2009a]. Heritability.

Downes, S. M. [2009b]. Moving past the levels of selection debates. *Biology and Philosophy, 24*(5), 703-9.

Earnshaw-Whyte, E. [2012]. Increasingly Radical Claims about Heredity and Fitness. *Philosophy of Science, 79*(3), 396-412.

Endler, J. A. [1986]. *Natural selection in the wild*. Princeton: Princeton University Press.

Frank, S. A. [1998]. *Foundations of Social Evolution*. Princeton: Princeton University Press.

Gardner, A., & Grafen, A. [2009]. Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology, 22*(4), 659-71.

Gillespie, J. H. [2004]. *Population genetics: a concise guide*: Johns Hopkins University Press.

Godfrey-Smith, P. [2007]. Conditions for evolution by natural selection. *Journal of Philosophy, 104*(10), 489.

Godfrey-Smith, P. [2009]. *Darwinian populations and natural selection*: Oxford University Press, USA.

Hamilton, M. [2009]. *Population genetics*. Oxford: Wiley-Blackwell.

Jacquard, A. [1983]. Heritability: one word, three concepts. *Biometrics*, 465-77.

Lewontin, R. C. [1970]. The units of selection. *Annual Review of Ecology and Systematics, 1*(1), 1-18.

Lewontin, R. C. [1978]. Adaptation. *Scientific American,* (293), 212-28.

Lewontin, R. C. [1985]. Adaptation In R. Levins & R. Lewontin (Eds.), *Dialectics and reductionism in ecology* (pp. 65–84). Cambridge, MA: Harvard University Press.

Matthen, M., & Ariew, A. [2002]. Two ways of thinking about fitness and natural selection. *The Journal of Philosophy, 99*(2), 55-83.

Okasha, S. [2006]. *Evolution and the Levels of Selection*: Oxford University Press, USA.

Okasha, S. [2010]. Replies to my critics. *Biology and Philosophy, 25*(3), 425-31.

Price, G. R. [1970]. Selection and covariance. *Nature, 227*, 520-21.

Price, G. R. [1972]. Extension of covariance selection mathematics. *Annals of human genetics, 35*(4), 485-90.

Rice, S. H. [2004]. *Evolutionary theory: Mathematical and conceptual foundations*. Sunderland MA: Sinauer.

Ridley, M. [1996]. *Evolution (2nd edition)*. Oxford: Blackwell

Roughgarden, J. [1979]. *Theory of population genetics and evolutionary ecology*: Macmillan New York, New York, USA.

Sesardic, N. [2005]. *Making sense of heritability*: Cambridge University Press.

Van Valen, L. M. [1976]. Energy and evolution. *Evolutionary theory, 1*(1), l79-229.

Visscher, P. M., Hill, W. G., & Wray, N. R. [2008]. Heritability in the genomics era—concepts and misconceptions. *Nature Reviews Genetics, 9*(4), 255-66.

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/ (Version 5.0). Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Appendix: Okasha's derivation of the Price equation making explicit the heritability term

Okasha ([2006]) derives a formulation of the Price equation that makes explicit the term h^2 of heritability. In this appendix I detail the different steps to get to this formulation.

The formulation of the Price equation Okasha starts with is the following:

$$\Delta\bar{z} = Cov\,(\omega, z) + E_w(\Delta z) \qquad (2)$$

ω is the relative fitness ($\omega = w/\bar{w}$) and $E_w(\Delta z)$ is the fitness-weighted average of the quantity $\Delta\bar{z}$.

From there, he notes that the term $E_w(\Delta z)$ in (2) can be decomposed as follows (Okasha 2006, 26):

$$E_w(\Delta z) = E(\Delta z) + Cov\,(\omega, \Delta z)$$

Thus, (2) can be rewritten as follows

$$\Delta\bar{z} = Cov\,(\omega, z) + E(\Delta z) + Cov\,(\omega, \Delta z)$$

Noting that $\Delta z = z' - z$, this equation can be simplified. This yield:

$$\Delta\bar{z} = Cov\,(\omega, z) + E(\Delta z) + Cov\,(\omega, z') - Cov\,(\omega, z) = Cov\,(\omega, z') + E(\Delta z)$$

$Cov\,(\omega, z')$ is the covariance between an entity's relative fitness and the average trait value of its offspring.

From this formulation of equation (2), we can now make heritability h^2 apparent. In fact, we know that with a regression line in place on a graph representing the average offsrping values of character z on parental values (or mid-parental in the case of sexual reproduction) of each individual of the initial population, for each point z' on the figure we have:

$$z' = a + h^2 z + e$$

where a is the intercept and e is what remains unexplained by the regression.

Replacing this decomposition of z' in the Price equation we obtain:

$$\Delta \bar{z} \; = \; Cov\,(\omega, a + h^2 z + e) + E(\Delta z)$$

$$= \; Cov\,(\omega, a) \; + h^2 Cov\,(\omega, z) \; + Cov\,(\omega, e) \; + E(\Delta z)$$

Because $Cov\,(\omega, a) = 0$, since a is a constant and that the assumption $Cov\,(\omega, e) = 0$ is a

reasonable one (Okasha [2006], p. 35) we can simplify this result into:

$$\Delta \bar{z} \; = \; h^2 Cov\,(\omega, z) \; + E(\Delta z) \qquad (3)$$