Epigenetic Modeling as a Remedy for the Missing Heritability Problem

Stephen P. Smith
January 2024

Abstract: This paper argues that confounding between epigenetic and additive genetic effects explains the missing heritability problem, and offers two new epigenetic models describing transgenerational inheritance that may supplement the traditional mixed linear models used to make predictions of the additive genetic merit of animals. The new models utilize pedigree information and show the property of additivity, like the additive genetic models do. The combined predictions of both epigenetic and genetic effects may provide useful indexes that help farmers select replacement animals or breeding stock. The utility of these methods depends on data where both confounding and the parsimony of parameters come to bear on practical limits. The proposed epigenetic models are just a first step, and they may not win long-term favor when compared to future models that are still to be formulated.

Key Words: Animal Breeding, Epigenetics, Missing Heritability, Mixed Linear Model, Transgenerational Inheritance.

1. Introduction

The study of epigenetics has unfolded in a chronological sequence of discoveries, illuminating the intricate regulation of gene expression beyond the DNA sequence (cf., Peixoto, et al., 2020). Conrad Waddington coined the word *epigenetics* in the 1940s, highlighting the interaction between environment and genes. DNA methylation, the addition to DNA of methyl groups, was the first epigenetic modification identified in the 1970s, providing a significant insight into gene expression during embryonic development. Subsequent decades brought the discovery of histone modifications (acetalyzation), revealing the dynamic role of proteins associated with DNA in gene regulation. The advent of technologies like bisulfite sequencing and chromatin immunoprecipitation enhanced our ability to map epigenetic marks across the genome (cf., Hurbert and Demars 2022). More recently, the exploration of non-coding RNAs and three-dimensional chromatin architecture has expanded the epigenetic landscape. This chronological progression underscores the complexity of epigenetics, shaping our understanding of how environmental factors leave lasting impacts on the genome.

In recent years, groundbreaking research has unveiled the phenomenon of transgenerational inheritance through epigenetics; taking the subject beyond developmental biology. Initially, studies focusing on specific environmental exposures

demonstrated that epigenetic modifications could be passed from one generation to the next. The seminal work on agouti mice revealed that maternal diet could influence the coat color and health of subsequent generations (Waterland and Jirtle 2003). Further investigations across various species, including humans, illuminated the potential transmission of epigenetic marks through multiple generations. These findings redefine our understanding of heredity, indicating that acquired traits and environmental impacts can extend beyond a single individual, providing compelling evidence for the profound and lasting influence of epigenetic mechanisms on the inheritance of traits and diseases.

The deeper implication of transgenerational epigenetic inheritance goes to the foundation of biology, and leads to unsettling ramifications about evolution and the structure of life. For example, Shapiro (2011), offers an overpowering critique of the Modern Synthesis, the prevailing framework in evolutionary biology based on Neo-Darwinism. He refutes the conventional theory that evolution mainly occurs through gradual accumulation of genetic mutations, emphasizing the significance of non-random, adaptive processes that represent yet another layer of epigenetic regulation. Shapiro argues that the gene-centric view fails to capture the complexity of evolutionary mechanisms, particularly the role of natural genetic engineering. He asserts that organisms possess active mechanisms for restructuring their genomes, challenging the passive role ascribed to natural selection. In fact, this concept (that life is endowed by its own agency) is taken serious by well-respected scientists that made contributions to the scholarly book published in 2023, *Evolution "On Purpose" - Teleonomy in Living Systems*, and this concept now recast epigenetics in this broader context of a biology that actively regulates itself, of being its own agent. Shapiro and others urge a paradigm shift towards a new evolutionary framework, acknowledging the dynamic nature of genetic change and the active participation of organisms in their own evolution. By advocating a more inclusive perspective, Shapiro and others call for a reevaluation of established evolutionary principles to better incorporate the complexity inherent in the molecular processes shaping the course of evolution.

The Modern Synthesis is not alone in its demise. The Central Dogma of Biology and Weismann’s barrier have been assumed premises of theoretical biology, which depicts information flow out of DNA as a one-way affair. It is not supposed to be possible for information to flow from the soma to the germline, but in a review article Phillips and Nobel (2023) describe such a serious likelihood: they discuss evidence that such communication is mediated by somatic RNAs that travel inside extracellular vesicles (exosomes) to the gametes where they reprogram the offspring epigenome and phenotype. As an example of what might be transferred inside of exosomes, Rechavi, et al., (2011) show that the viral silencing agents, viRNAs, are transgenerationally transmitted and work to silence viral genomes present in C. elegans (nematodes). In a review article, Houri-Zeevi and Rechavi (2017) find that small RNAs are increasingly emerging as transgenerational carriers of epigenetic information, which can affect gene expression over several generations. In particular, Posner, et al., (2019) have

discovered that neuronal small RNAs can control behavior transgenerationally in C. elegans, contradicting the Central Dogma.

Bioelectricity is another layer of regulation (of agency directing its own morphogenesis and maintaining its homeostasis) that rest on top of the epigenetic regulation that controls the expression of genes, and Michael Levin describes these fascinating processes in Chapter 10 of the book, *Evolution "On Purpose" - Teleonomy in Living Systems*. On page 188, Levin writes: “bioelectric states are a medium that binds individual cells toward large-scale goals; it underlies scale-up and emergence of higher-level teleonomic individuals, much as it does to create brains with emergent unified mental content out of a collection of individual neuronal cells.” Individual cells show intelligence, organs show goal-seeking intelligence and organisms represent a collective intelligence (Levin 2023). Bioelectricity represents the membrane potential that is supported by gap junctions, ion channels and pumps, and is the substrate that unifies a communicating cognition in all cells, organs and the brain. Agency of this sort implies a memory that is well beyond the genome. It is now necessary to reissue the definition of epigenetics, to now represent all such layers of regulation that imply some level of agency and may also impact the germline, and represents a possible transgenerational communication. These layers may operate on different time scales, and they may be hard to distinguish individually by statistical means, but they are unlike the more static information encoded in the genome.

With the gene-centric view of biology brought into question, it is not surprising to find entrenched theories of genetics and evolution in crisis, and the missing heritability problem (see Manolio, et al., 2009) hints of such a crisis within genetic research where identified genetic variants cannot fully explain the heritability of complex traits or diseases. Despite the advancements in genome-wide association studies (GWAS), a significant portion of the heritability remains elusive. Nevertheless, *genomic selection* may offer an improved utilization of available information coming from GWAS (Goddard, et al., 2010). While genomic selection as a theory and practice has now integrated itself well into traditional breeding principles (Meuwissen, et al., 2016), this approach is, however, a reassertion of the old gene-centric view of biology. Because genetic studies may overlook many small genetic effects, rare variants, structural variations, and gene-gene interactions that collectively contribute to the missing heritability (see Manolio, et al., 2009), a defense of the Modern Synthesis can still be made. However, an alternative view is that non-genetic factors such as epigenetic modifications and environmental influences may explain the missing heritability (Banta and Richards 2018, Slatkin 2009), and that the Modern Synthesis is outdated as Shapiro warns.

Plant and animal breeding advanced substantially under the theoretical framework provided by the Modern Synthesis. However, what can be made of these advances if it is now discovered that the framework is invalid? In particular, what to make of the selection principles that have reached a high point with best linear unbiased prediction and the application of the mixed model equations (prior to GWAS and genomic selection)? These

practical tools of statistics (e.g., Henderson 1984) seemed to work just the same, with genetic advances described in dairy populations (Everett and Keown 1984; Van Vleck, et al., 1986), and in many other examples where artificial selection was found to make progress. However, if the Modern Synthesis is invalid because of an epigenetic inheritance that was not fully recognized, the linear prediction of breeding values probably found utility partly because of statistical confounding between genetic and epigenetic effects. Indeed, Banta and Richards (2018) argue that epigenetic effects can change how phenotypic variance is partitioned, making a serious issue. As a practical matter, the improvement of domestic animals can be accomplished even if the heritability is mischaracterized, because of this apparent confounding. Henderson's practical tools have already been adapted and changed dramatically with the introduction of genomic selection and this revolution in theory and practice is unlikely to be shifted because of the new revelations about the inheritance of acquired characteristics, unjustified or not. However, if the old tools worked anyway to help improve animal populations because of confounding, it is possible that adjusting the statistical models to include epigenetic effects may offer farmers valuable new information to select replacement animals. The purpose of this paper is to offer statistical models that are enlarged to include a possible epigenetic inheritance. These models may not be the perfected models that are universally adopted in the future once the Modern Synthesis finally expires, but they provide a start to that eventuality.

Section 2 reviews the additive genetic model, additive genetic inheritance, and the connection of this model to mixed model methodology. The additive genetic model is used as a framework to develop two new epigenetic models of inheritance, and these are presented in Section 3. Section 4 presents a discussion of the utility of these new models to aid in culling and selection.

2. Additive genetic effects

Additive genetic effects are those genetic effects that can be selected and result in a stable change in a population, and even in subsequent generations. The additive effects can accumulate with the addition of new additive genetic effects that may also be selected. In the statistical model, the additive effects are assumed to sum over loci, where each allele at a particular locus carries a small effect that contributes to the total variance. The model of inheritance is given by the mid-parent equation (1).

$$A_o = \tfrac{1}{2}A_p + \tfrac{1}{2}A_m + S_o \,, \tag{1}$$

where: A_o is the total additive genetic effect in the offspring, o, that resulted from mating parents p and m; A_p and A_m are the total additive genetic effects from the paternal and maternal parents, respectively; and S_o is the random effect resulting from segregation, or the genetic recombination during meiosis when parental gametes were made. The mean of S_o is taken as zero, and variance of S_o is given by the following equation.

(2) $$\mathrm{Var}\{S_o\} = \tfrac{1}{2}\,(1+F_o)\,\sigma^2_A\,,$$

where F_o is the inbreeding coefficient of offspring o, and σ^2_A is the additive genetic variance which is treated as a population parameter.

Equations (1) and (2) can be put in matrix notation to specify the additive genetic model completely for N animals. Let there be M<N animals without known genetic relationships represented by (1). These M animals are part of the base population. Indexing all N animals makes o, p and m indexes and functions of an index $k \leq N$, given by k=o[k], p[k] and m[k]. Without loss of generality, use an index ordering for all animals that is consistent with the partial ordering where k > p[k] and k > m[k] for all k=o[k] where (1) applies, and where the first M indexes represent the base population. Signify the column vector $\mathbf{a}_{N\times 1}$ that contains the additive genetic effects A_k, in the k-th location and for the k-th animal, $k \leq N$. Likewise, define the vector $\mathbf{r}_{N\times 1}$ where the k-th position equals A_k if $k \leq M$, otherwise it equals S_k. With these specifications, the variance of **r** is a diagonal matrix **D** of order N, with k-th diagonal elements equaling σ^2_A if $k \leq M$, or otherwise it is specified by (2) as $\tfrac{1}{2}\,(1+F_k)\,\sigma^2_A$; in shorthand, var{**r**}=**D**. Lastly, define the lower triangular matrix **P** that contains mostly zeros except that the k-th row contains ½ at column positions p[k] and m[k], for all k>M.

Equation (1) becomes the succinct matrix equation (3).

(3) $$\mathbf{a} = \mathbf{P}\mathbf{a} + \mathbf{r} \qquad \text{or} \qquad (\mathbf{I}-\mathbf{P})\mathbf{a} = \mathbf{r}$$

This completes the specification of the additive genetic model in animal breeding. Equation (3) was first formulated by Quaas (1988) and it leads directly to the rules presented by Henderson (1976) for inverting the relationship matrix. It's this inverse matrix that is plug directly into Henderson's mixed model equations. The mixed linear model is also needed to relate the additive genetic effects to phenotypes, and it comes as an extra model specification as given below.

(4) $$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

where **y** is a vector of observations (or phenotypes), **b** is a vector of fixed effects, **a** a vector of random additive genetic effects already described (including var[**a**] and its inverse) and **e** is a vector of random residuals. The matrices **X** and **Z** are incidence matrices that assign the various effects to observations and var[**e**]=**R**. Further details are beyond the scope of this paper.

This model, given by (3) and (4), can be generalized to accommodate diverse modifications. For example, it can be extended to multivariate traits, modified to include other random effects like common environmental influences and the additive genetic

effects representing the maternal environment. It can be reduced to represent a simple sire model. Attempts might also be made to represent some non-additive genetic effects like dominance and epistasis. Including non-additive genetic effects is based on the partitioning of total genetic variance following the work of Cockerham (1954), and others. However, there is a limit to the amount of non-additive genetic variance that can be realistically accommodated because of complexity, confounding and because not all parameters can be estimated well from data. Inbreeding makes the genetic partitioning even more complicated while inflating the number of genetic parameters that are needed. Much of the time these complexities are just ignored while preference is given to a simple additive genetic model. As already noted, the additive models are preferred because additive genetic effects are thought stable across generations and do not dissipate, and hence, these effects lend themselves to selection.

The additive genetic model has not been without critics, and at best this model should only be viewed as a useful approximation. With the model assumed true the heritability and other parameters must still be estimated, and parameters are sensitive to the choice of model. Sheridan (1987) offers criticism of model assumptions that typifies parameter estimation, and recommends estimating heritability from selection experiments and not from likelihood-based methods that use field records. Heritability and the additive genetic variance depend on gene frequencies and can change with selection. The maintenance of additive genetic variance in populations undergoing selection has been a subject of investigations (see Turelli 1988).

3. Possible Epigenetic Models

3.1 First epigenetic model

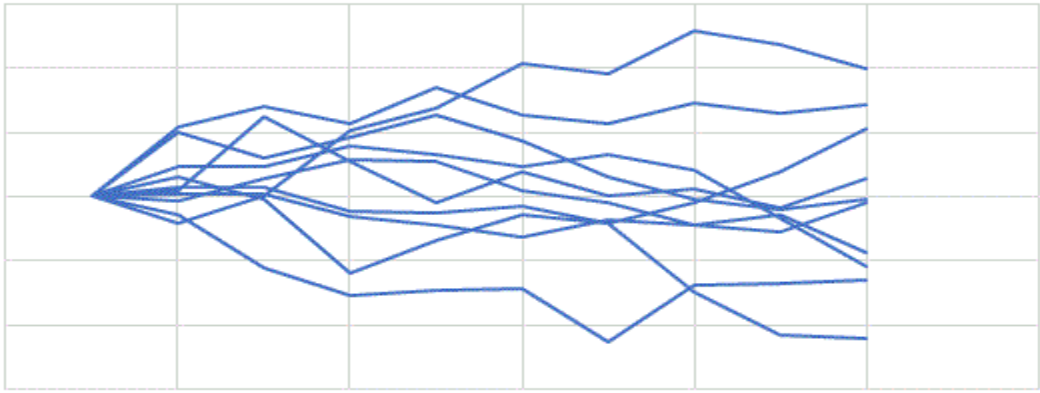
The additive model described in Section 2 is characterized by representing evolution on the slowest time scale, where the available genes were honed and fashioned by evolution. By analogy only, consider a simple linear regression on time: Observation=Intercept + Slope×Time. The first term (Observation) relates to the phenotype, and the constant term (Intercept) relates to the additive genetic model where additive effects are assumed to accumulate and lead to a stable selection. In this analogy, the variable term (Slope×Time) represents epigenetic inheritance that could come in the form of multiple levels that effect the phenotype and where each level may represent its own time scale. The appropriate way to model time sensitive effects is by (1) time series analysis, which is what will be done here.

The other requirement is that these epigenetic effects should also be (2) additive, so as to complement the additive genetic model and that may also partly confound with the additive model; hence, the plus sign in linear regression (Intercept + Slope×Time). Fortunately, the confounding will not be 100%, otherwise the situation would be

hopeless. Possible epigenetic models will be formulated to meet these two requirements.

The requirement of additivity eliminates stationary time series because these series return to their central mean value. What is needed is a non-stationary time series, and the simplest version is a first-order non-stationary time series, the random walk. Ten random walks that start from a common location are presented in Figure 1.

Figure 1. Random Walks Starting from Common Location



The variance diverges across the different walks, but the variation is centered around the common starting point. Hence, while the walk position at the start can expire with time passage, the central mean value remains despite the increase in variance. That is, additivity remains.

The variance becomes arbitrarily large as time increases. This condition is not artificial despite the associated limiting variance that approaches the unattainable infinity. Non-stationary time series may serve in non-parametric regression precisely because they are very malleable to data (e.g., Smith 2018). The differences between time steps necessarily show finite variance, and initial conditions may be taken as fixed effects. It is only the few fixed effects that can be treated with the Bayesian interpretation that lets very few variances drift to infinity without causing any challenges while introducing the non-informative prior. Moreover, biology has navigated the evolutionary landscape while happening on many diverse expressions of life, and a statistical model should be able to reflect the same malleability.

Define $V_k(t)$ as the additive epigenetic effect on the k-th individual at time t. For a non-stationary and first order time series, the following statistical model applies.

(5) $$V_k(t+d) = V_k(t) + \sqrt{d} \times \sigma \times \varepsilon_k \,,$$

where d is the change in time, σ is a dispersion parameter and ε_k a is a random deviate

with mean zero and variance 1. For discussion, the random deviate may be taken as normally distributed, as is typical of mixed linear models. The variance of the residual in (5) is $d\times\sigma^2$.

An individual will have various realized epigenetic effects in the course of a lifetime (a whole time series, in fact), but only an ordered short-list of these is required for modeling. As an example, let t_1 be the time the k-th individual was born, t_2 the time its phenotype was recorded, and t_3 the time this individual becomes a parent. In males, t_3 is the time sperm was produced to form a new zygote (not including the duration semen may have been frozen). In females, this is the time at birthing. While female mammals have all their egg cells formed during their embryonic development, epigenetic effects can still be passed across the placenta and so the gestation time is part of t_3. The case of embryo transfer is not considered in this paper. The short-list can have more than three time periods identified if the k-th animal has additional offspring, or more phenotypes recorded.

A model is also required to describe the blending of epigenetic effects from parents to offspring, and there are several alternatives available of which one is presented below.

(6) $$V_{o(k)}(t) = \rho\times V_{p(k)}(t) + (1-\rho)\times V_{m(k)}(t) ,$$

where $0\leq\rho\leq1$, and the time parameter is used transgenerationally in this instance and is only meant to depict the simultaneous joining of parents (p and m) to produce an offspring (o). Equation (6) differs from (1) in that ρ may not be ½, and there is no error term (corresponding to S_o) that represents recombination. Ideally, it is preferable to have known biology inform on the structure of (6), e.g., as much as meiosis informs on (1), but this specificity is lacking in the present treatment. Slatkin (2009) uses a population genetics framework to develop a transgenerational epigenetic model, but it is a different model. This leaves the simplified model (6) as one alternative. However, the parameter ρ is still required to be estimated from actual data. Model (6) permits an uneven distribution of epigenetic effects coming from paternal and maternal sources. This model confounds more with the additive genetic model when $\rho = ½$.

Combining models (5) and (6) leaves everything specified but the epigenetic effects, $V_k(t)$ where $k\leq M$, corresponding to the base population. The mean of these effects can be taken as null, as well as the covariances among these effects in the base population. However, the variance still needs to be specified. It is remarkable that these variances do not diverge to infinity, as typifies the random walk. This is because the blending of epigenetic effects by (6) acts to renormalize the variance. To show this in a hypothetical sense, consider a large population that mates randomly, with a fixed generation interval T, and without overlapping generations. Let σ^2_{Vj} be the epigenetic variance corresponding to the j-th generation, j =0, 2, 3, etc., and assume that paternal and maternal sources have the same variance within generations. By the time of mating, then

(5) implies that σ^2_{Vj} grows by $T \times \sigma^2$, and (6) shows how this change impacts the variances of generation j+1 compared to generation j:

(7) $$\sigma^2_{Vj+1} = [\rho^2 + (1-\rho)^2] [\sigma^2_{Vj} + T \times \sigma^2]$$

Starting with j=0, this recursion can be iterated indefinitely. The term representing σ^2_{V0} vanishes, and the iteration converges to a geometric series:

(8) $$\sigma^2_{V\infty} = \sum_{j=1}^{\infty} [\rho^2 + (1-\rho)^2]^j \times T\sigma^2 = \left[\frac{1}{2\rho(1-\rho)} - 1\right] \times T\sigma^2$$

Therefore, with T estimated crudely as the average generation interval, (8) provides a way to approximate the epigenetic variance in the base population for all animals at birth, and this approximation is found as a function of two unknown parameters of the model, ρ and σ^2. These unknown parameters can be estimated by a likelihood-based method.

The variance calculation given by (8) diverges to infinity when ρ equals 0 or 1, and in this case, there is no blending of parental epigenomes. While it is possible to treat this variance as infinite with the Bayesian interpretation, thereby treating the epigenetic effects in the base population as fixed effects, this might create an unacceptable loss of information. It may be preferable to estimate or nominate a separate variance to represent the base population, or otherwise use a ridge regression approach. It may also be preferable to use a grouping strategy in the base population where mean effects can be included in the model.

The epigenetic model is now completely specified by (5), (6) and (8) and this model can now be formulated in matrix notation. A representation similar to (3) is sought, and is presented by equation (9)

(9) $$\mathbf{q} = \mathbf{Q}\mathbf{q} + \mathbf{w} \quad \text{or} \quad (\mathbf{I}-\mathbf{Q})\mathbf{q} = \mathbf{w},$$

where the ordered short-lists of all epigenetic effects are contained in **q** (which is much longer than N), and **w** is a vector of residuals. The matrix **Q** is mostly populated with zeros, except for the number 1 coming from the equations given by (5), and except for ρ and 1- ρ that are implied by the equations given by (6). The variances of residuals (**w)**, are provided by (5) and (8). Define these variances collectively as the diagonal matrix, var[**w**]=**H**.

Because **q** is ordered to maintain the partial ordering of all the short-lists, and where parents are listed before offspring, the matrix **Q** is lower triangular as was the matrix **P**. The matrix **H** is also singular because the blending equation (6) does not come with a residual.

This completes the model specifications for the first epigenetic model, and the mixed model can be employed with a linear model similar to (4). If the mixed model equations are to be used, however, additional modifications are required because **H** is singular. It is possible to use the matrix **H** directly as part of a symmetric and indefinite matrix called the K-matrix. Numerical methods have been developed for treating such systems (Smith 2001 a & b, 2017), and examples of using the K-matrix are available (Smith, Lopez and Lam 2017, Smith 2018, Smith and Mäki-Tanila 2018).

According to this model, if a parent has a measure V of epigenetic merit, then its offspring will initially show ρV or (1-ρ)V, depending on the gender of the parent. Moreover, this gain (as an expectation conditional on V) will fall by a factor ½ for each new generation because an equal number of male and female descendants are expected. This additive model mimics the behavior of (1), despite the fact that it comes with an increase in variance during the life span of descendants and this also diminishes the impact of V with time passage.

3.2 Second epigenetic model

In the second model, epigenetic effects are to be tracked with two paths corresponding to males and females. Let the two paths be signified by XY for males, and XX for females, respectively and where these symbols are to be used as subscripts. Therefore, equation (5) reformulates into two equations depending on the gender of the k-th individual:

$$V_{XY.k}(t+d) = V_{XY.k}(t) + \sqrt{d} \times \sigma_{XY} \times \varepsilon_k \,, \quad \text{or} \quad V_{XX.k}(t+d) = V_{XX.k}(t) + \sqrt{d} \times \sigma_{XX} \times \varepsilon_k \,, \tag{10}$$

The only difference in the specifications given by (10) compared to (5), is the introduction of two new parameters (σ_{XY} and σ_{XX}) that replaces the one (σ).

The blending equation (6) also splits into two equations given by (11), with the introduction of two new parameters, $0 \le \rho_{XY} \le 1$ and $0 \le \rho_{XX} \le 1$:

$$V_{XY.o(k)}(t) = \rho_{XY} \times V_{XY.p(k)}(t) + (1-\rho_{XY}) \times V_{XX.m(k)}(t) \quad \text{or} \quad V_{XX.o(k)}(t) = \rho_{XX} \times V_{XX.m(k)}(t) + (1-\rho_{XX}) \times V_{XY.p(k)}(t) \tag{11}$$

It remains necessary to find approximate variances for epigenetic effects in the base

population. The recursion (7) based on non-overlapping generations and a large population generalizes to the new case, where T is again the generation interval:

$$\begin{bmatrix} \gamma_{XY}{}^2 \\ \gamma_{XX}{}^2 \end{bmatrix}_{j+1} = \begin{bmatrix} \rho_{XY}{}^2 & (1-\rho_{XY})^2 \\ (1-\rho_{XX})^2 & \rho_{XX}{}^2 \end{bmatrix} \left(\begin{bmatrix} \gamma_{XY}{}^2 \\ \gamma_{XX}{}^2 \end{bmatrix}_j + T \begin{bmatrix} \sigma_{XY}{}^2 \\ \sigma_{XX}{}^2 \end{bmatrix} \right)$$

This shows the generalized recursions in matrix form, where the subscript (j or j+1) indicates the generation number, and $\gamma_{XY}{}^2$ and $\gamma_{XX}{}^2$ are the epigenetic variances for males and females in a particular generation indicated by the subscript. This set of recursions again leads to a convergent series when it is iterated, the matrix form of the geometric series in fact, see:

$$(12) \quad \begin{bmatrix} \gamma_{XY}{}^2 \\ \gamma_{XX}{}^2 \end{bmatrix}_{\infty} = T \sum_{j=1}^{\infty} \begin{bmatrix} \rho_{XY}{}^2 & (1-\rho_{XY})^2 \\ (1-\rho_{XX})^2 & \rho_{XX}{}^2 \end{bmatrix}^j \begin{bmatrix} \sigma_{XY}{}^2 \\ \sigma_{XX}{}^2 \end{bmatrix} = T \left[\begin{bmatrix} 1-\rho_{XY}{}^2 & -(1-\rho_{XY})^2 \\ -(1-\rho_{XX})^2 & 1-\rho_{XX}{}^2 \end{bmatrix}^{-1} - \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \right] \begin{bmatrix} \sigma_{XY}{}^2 \\ \sigma_{XX}{}^2 \end{bmatrix}$$

With T estimated, equation (12) provides approximate epigenetic variances for males and females in the base population, and these depend on four parameters of the model. As with the previous model in Section 3.1, the specification given by (10), (11) and (12) can be represented in matrix form similar to (9), and statistical analysis may proceed.

According to this model, a male parent with epigenetic merit V will pass on to offspring either a merit of $\rho_{XY}V$ if the offspring is male, or $(1-\rho_{XX})V$ if the offspring is female. Therefore, the average merit (conditional on V) in the offspring is $\frac{1}{2}(1+\rho_{XY}-\rho_{XX})V$. In the next generation the conditional mean of the descendants is $\frac{1}{4}(1+\rho_{XY}{}^2-\rho_{XX}{}^2)V$, and it falls almost exponentially in every subsequent generation. Of course, these effects that are proportional to the historic measure V are diminished in the realized epigenetic values in descendants because of the accompanying increase in variance with time passage. A similar comparison holds for descendants of a female parent.

4. Discussion

Biological evolution could not have been a slow and haphazard process driven only by random mutations and an indifferent natural selection. It is far too proficient. Levin (2023) describes bioelectric networks and writes, “evolution exploits the generic computational properties of such networks (learning, generalization, counterfactual memories, representation, distributed control, etc.) at many scales, building flexible problem-solving engines instead of fixed solutions to specific environments.” Life is able to pull itself up from its bootstraps, to use a colloquialism, and it does this on all scales. This implies that

a type of greedy algorithm is enough to permit evolutionary and developmental progress where incremental gains are achieved even as the problem space changes in real time. This is the hallmark of an additivity that is realized on all scales. As an example, Hill, Goddard and Visscher (2008) describe how most genetic variance is additive genetic variance. It may be that additivity is a necessary condition for evolution, and this condition is satisfied when general problem-solving is possible. It's also the case that non-additive variation can turn into additive variation as situations change (e.g., Hill 2017), providing for the maintenance of heritable information on all scales.

The epigenetic models in Section 3 were specifically designed to show additivity, the type of variability that can support a proficient evolution and general problem-solving. However, the additive genetic effects in Section 2 should also be fitted in a common statistical model with the epigenetic effects, to make a better accounting of the confounding between the two. The two sources are confounded, but not necessarily in a fatal way that prevents their use together in a model. The possible limitation from confounding depends on data and is related to whether the unknown parameters show enough parsimony to be estimated. Nevertheless, a better statistical model may lead to improved indexes that can aid farmers with their selection and culling decisions, where both genetic and epigenetic gains are possible.

The models in Section 3 are not the only models that are possible, and other choices can be considered or recommended. This includes giving consideration to models that may be non-additive. One alternative is to put a residual in (6), and/or to replace coefficients (ρ, 1- ρ) with positive parameters that may add to a number less than 1. Another alternative is to treat the parameter ρ in Section 3.1 as random, where each blending of parental epigenomes realizes a new ρ to make the offspring, such that the selection of ρ is unique. This modification likely leads to a tractable analysis by way of Bayesian simulation, though it is probably computationally expensive. A third possibility is to put together two (or more) epigenetic models to make one phenotypic model, following Section 3.1 but representing fixed and different selections of ρ, such as ρ =1 and 0. This will make the confounding more of an issue, but limitations from confounding are data dependent.

New models based on quantum information theory, or something congruent with materials presented by Fields, et al., (2023), might can be developed. Another possibility is that a neural net can be developed that leaves the additive genetic and epigenetic effects confounded as part of a black box. Merit indexes that confound the two sources of information are likely to find utility with farmers anyway.

References

Banta, J.A., and C.L. Richards, 2018, Quantitative epigenetics and evolution, *Heredity*, 121 (3), 210-224.

Cockerham, C.C., 1954, An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis Is present, *Genetics*, 39(6), 859-882.

Everett, R.W., and J.F. Keown, 1984, Mixed model sire evaluation with dairy cattle - experience and genetic gain, *Journal of Animal Science*, 59(2), 529-554.

Fields, C., Aguirre, A., Friedman, D. A., Balzan, F., Berger, R., Draguta, D., Guénin-Carlut, A., Hurtado, A. M., Knight, V. B., Koles, K., Neophytou, H., Prior, C., Tickles, D., Tennant, I., Tolchinsky, A., & Active Inference Institute. (2023, December 31). Transcript of Chris Fields, "Physics as Information Processing" course in 2023. In Active Inference Journal. Active Inference Journal. https://doi.org/10.5281/zenodo.10447643.

Goddard, M.E., B.J. Hayes and T.H.E. Meuwissen, 2010, Genomic selection in livestock populations, *Genetics Research*, 92(5-6), 413-421.

Henderson, C. R., 1976, A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values, *Biometrics*, 32, 69-83.

Henderson, C.R., 1984, *Application of Linear Models in Animal Breeding*, University of Guelf.

Hill, W.G., M. Goddard and P.M. Visscher, 2008, Data and theory point to mainly additive genetic variance for complex traits, *PLOS Genetics*, 4 (2), 1-10.

Hill, W.G., 2017,“Conversion” of epistatic into additive genetic variance in finite populations and possible impact on long-term selection response, *Journal of Animal Breeding and Genetics*, 134(3), 196–201.

Houri-Zeevi, L., and O. Rechavi, 2017, A matter of time: small RNAs regulate the duration of epigenetic inheritance, *Trends in Genetics*, 33(1), 46-57.

Hurbert, J.-N., and J. Demars, 2022, Genomic Imprinting in the New Omics Era: A Model for Systems-Level Approaches, Frontiers in Genetics 13:838534, 1-9.

Levin, M., 2023, Bioelectric networks: the cognitive glue enabling evolutionary scaling from physiology to mind, *Animal Cognition*, 26, 1865-1891.

Meuwissen, T.H.E, B.J. Hayes, and M.E. Goddard, 2016, Genomic selection: a paradigm shift in animal breeding, *Animal Frontiers*, 6(1), 6-14.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, et al.,

2009, Finding the missing heritability of complex diseases, *Nature*, 461(7265), 747-753.

Peixoto, P., P.-F. Cartron, A.A. Serandour and E. Hervouet, 2020, From 1957 to nowadays: a brief history of epigenetics*, International Journal of Molecular Sciences*, 21, 7571, 1-18.

Phillips, D., and D. Noble, 2023, Bubbling beyond the barrier: exosomal RNA as a vehicle for soma-germline communication, The Journal of physiology, 11(7), 1-17.

Posner, R., I. A. Toker, O. Antonova, E. Star, S. Anava, E. Azmon, M. Hendricks, S. Bracha, H. Gingold, and O. Rechavi, 2019, Neuronal Small RNAs Control Behavior Transgenerationally, *Cell*, 177(7), 1814-1826.

Quaas, R. L., 1988, Additive genetic model with groups and relationships, *Journal of Dairy Science*, 71, 1338–1345.

Rechavi, O., G. Minevich and O. Hobert, 2011, Transgenerational Inheritance of an Acquired Small RNA-Based Antiviral Response in C.ϲelegans, *Cell*, 147, (6), 1248-1256.
Shapiro, J.A., 2011, *Evolution: a view from the 21st century*, FT Press Science.

Sheridan, A.K., 1987, Utility of genetic parameter estimates, Proceeding of the Sixth Conference, Australian Association of Animal Breeding and Genetics, Perth, 49-54.

Slatkin, M, 2009, Epigenetic Inheritance and the Missing Heritability Problem, *Genetics*, 182 (3), 845-850.

Smith, S.P., 2001a, The factorability of symmetric matrices and some implication for statistical linear models, *Linear Algebra and Its Applications*, 335, 63-80.

Smith, S.P., 2001b, Likelihood-based analysis of *linear* state-space models using the Cholesky decomposition, *Journal of Computational and Graphical Statistics*, 10 (2), 350-369.

Smith, S.P., 2017, The backward differentiation of the bordering algorithm for an indefinite Cholesky factorization, viXra:1707.0239.

Smith, S.P., C. Lopez and N. Lam, 2017, Combining genetic similarities among known relatives that connect to an unknown relative, viXra:1705.0295.

Smith, S.P., 2018, Non-parametric regression or smoothing on a two-dimensional lattice using the K-matrix, viXra:1812.0044.

Smith, S.P., and A. Mäki-Tanila. 2018, Estimating genetic parameters in a dominance

model that includes inbreeding, viXra: 1802.0057.

Turelli, M., 1988, Phenotypic evolution, constant covariances, and the maintenance of additive variance, *Evolution*, 42,1342-1347.

Van Vleck, L.D., R.A. Westall, and J.C. Scneider, 1986, Genetic change in milk yield estimated from simultaneous genetic evaluation of bulls and cows, *Journal of Dairy Science,* 69, 2963–2965.

Waterland, R.A., and R.L. Jirtle, 2003, Transposable elements: targets for early nutritional effects on epigenetic gene regulation, *Molecular and Cellular Biology*, 23(15), 5293–5300.