RESEARCH ARTICLE

AMERICAN JOURNAL OF
BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

WILEY

Formal models for the study of the relationship between fluctuating asymmetry and fitness in humans

Arodi Farrera [ORCID]

Mathematical Modeling of Social Systems
Department, Institute for Research on Applied
Mathematics and Systems, National
Autonomous University of Mexico, Mexico
City, Mexico

Correspondence
Arodi Farrera, Mathematical Modeling of
Social Systems Department, Institute for
Research on Applied Mathematics and
Systems, National Autonomous University of
Mexico, Mexico City, Mexico.
Email: arodi@comunidad.unam.mx

Abstract

Objectives: To evaluate three of the main verbal models that have been proposed to explain the relationship between fluctuating asymmetry and fitness in humans: the "good genes," the "good development," and the "growth" hypotheses.

Materials and Methods: A formal model was generated for each verbal model following three steps. First, based on the literature, a theoretical causal model and the theoretical object of inquiry were outlined. Second, an empirical causal model and the targets of inference were defined using observational data of facial asymmetries and life-history traits related to fitness. Third, generalized linear models and causal inference were used as the estimation strategy.

Results: The results suggest that the theoretical and empirical assumptions of the "good genes" hypothesis should be reformulated. The results were compatible with most of the empirical assumptions of "the good development" hypothesis but suggest that further discussion of its theoretical assumptions is needed. The results were less informative about the "growth" hypothesis, both theoretically and empirically. There was a positive association between facial fluctuating asymmetry and the number of offspring that was not compatible with any of the empirical causal models evaluated.

Conclusions: Although the three hypotheses focus on different aspects of the link between asymmetry and fitness, their overlap opens the possibility of a unified theory on the subject. The results of this study make explicit which assumptions need to be updated and discussed, facilitating the advancement of this area of research. Overall, this study elucidates the potential benefit of using formal models for theory revision and development.

KEYWORDS
causal inference, directed acyclic graphs, face, fitness, fluctuating asymmetry

1 | INTRODUCTION

Developmental stability (DS) is the ability of an organism to produce a consistent phenotype despite the environmental and genetic disturbances faced during development (Debat & David, 2001). This ability and its consequences are considered to have a fundamental role in the survival and/or reproduction of the individual and, therefore, to be an important part of fitness (Clarke, 1995). For decades,

74 | WILEY— AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

FARRERA

one of the most common measurements of DS has been the intra-individual variability of paired bilateral traits, also known as fluctuating asymmetry (FA; Graham, 2021a; Palmer & Strobeck, 1992). The main argument for using FA is that since the sides of symmetrical organisms develop under identical genotypes and environments, the differences between them are mainly due to developmental noise (DN) or stochastic variation occurring during development (Hallgrímsson et al., 2002). This definition posits FA as an indicator of developmental precision and as a useful, cheap proxy of other more direct fitness estimates (e.g., Clarke, 1995). However, even though several hypotheses have been developed to explain these links and much research has been done on this topic, most of the evidence is inconclusive.

Particularly in humans, the evidence on the relationship between asymmetry and the components of the individual's health and fitness is ambiguous, as some studies have found an association (meta-analysis: Møller & Thornhill, 1998; oxidative stress: Gangestad et al., 2010; attractiveness: Brown et al., 2008; sexual behavior: Kordsmeyer & Penke, 2017), while others have found little relationship between them (meta-analysis: Palmer, 2000; van Dongen & Gangestad, 2011; attractiveness: Jones & Jaeger, 2019) or no relationship (attractiveness: Kleisner et al., 2017; health: Foo et al., 2017). Although most of this conflicting evidence can be attributed to the fact that FA is a weak measure of DS, methodological flaws or selective reporting and publication of mostly statistically significant results (e.g., Graham & Özener, 2016; Palmer, 1999; van Dongen & Gangestad, 2011), I argue that the lack of formal models for the hypotheses that explain these phenomena could also be a contributing factor to this problem.

As with other complex phenomena (Smaldino, 2020), the relationship between FA and fitness has been usually explained in descriptive terms. Although descriptive explanations (i.e., verbal models) are useful for delimiting the topic of interest or triggering the development of new ideas, we now know that the ambiguous way in which they are expressed makes it difficult to, for example, clearly establish how hypotheses relate to observed data or to recognize whether a result constitutes evidence for or against a given hypothesis (Smaldino, 2017). Therefore, it is possible that the hypotheses that have been proposed thus far have been ill-defined (e.g., by confusing hypothesis with its predictions, Strode, 2015) or expressed in such an ambiguous way that they are obscuring the already weak evidence on FA as an indicator of developmental precision, thus hindering the interpretation of the link between FA and any health and fitness outcome. Furthermore, ambiguous explanations make it difficult to update hypotheses, as it is unclear how new ideas and assumptions connect to old ones, or how to use new results for theory revision and development. In this way, continuing to test outdated hypotheses could also be a contributing factor to the ambiguous evidence found in the study of FA and fitness.

Formal models address some of the problems related to descriptive explanations by specifying in precise terms which variables are relevant to a given topic and our assumptions about how they are related (Robinaugh et al., 2021; Smaldino, 2020). In humans, three main verbal models have been proposed to explain the relationship between FA and fitness: the "good genes," the "good development,"

and the "growth" hypotheses; but to my knowledge, no formal model has been developed on this relationship, nor have all three hypotheses been evaluated and compared simultaneously. As a first step in this direction, the aim of this contribution is twofold: to propose formal models for these common hypotheses and test these formal models in the particular case of facial asymmetries and reproductive success.

1.1 | A brief introduction to formal models

Both verbal and formal models articulate some aspect (e.g., components, relationship between components, consequences) of a complex phenomenon of interest (see Frigg & Hartmann, 2020), but while the former does so using descriptive explanations, the latter does so through graphical representations (e.g., Rohrer, 2018) and mathematical or computational modeling (Smaldino, 2020). By making the research question explicit, as well as the assumptions about what components are (or are not) relevant and how they connect to each other (see Smaldino, 2020; Robinaugh et al., 2021 for an introduction), formal models make it possible for research goals, methodology, and results to align. Although there is no unique procedure to generate these models, there are two approaches that help in this task: the estimand framework and the causal framework.

Traditionally, an estimand defines a target quantity to be estimated, while an estimator and an estimate refer, respectively, to the method used to obtain an approximation of this target and the specific value obtained when this method is applied to actual data (e.g., Little & Lewis, 2021). The estimand framework (see Lundberg et al., 2021 for an introduction) considers an additional distinction between theoretical and empirical estimands that improves the link between theory and evidence by clearly delineating the conceptual and empirical parts of the argument and accounting for cases where these estimands are not equivalent. This framework allows us to explicitly state what we try to know or describe (i.e., theoretical estimand: quantity of theoretical interest), what we can actually learn from available data and procedures (i.e., empirical estimand: quantity of practical interest), and how we can learn from data (estimation strategy).

Briefly, the theoretical estimand defines in precise terms the target of inquiry by formalizing the quantity most relevant to the theory and the target population over which to draw inferences. Since it is derived from theory, it can account for observable and unobservable variables (e.g., missing data). An empirical estimand, on the other hand, defines the quantity that can be recovered from observed data only, and thus informs us about the theoretical estimand under specific assumptions (e.g., convenience sample). The last component of this framework is the estimation strategy or the process that will be used to learn about the empirical estimand, which includes the estimator and estimate. Among other things, it is sought (see Wasserstein et al., 2019 for an introduction) that rather than estimation strategies that rely on making dichotomous inferences about the presence or absence of the effects of interest (e.g., using null hypothesis significance testing statistical or Bayes factor) and on reporting and

interpreting point estimates, estimation strategies focus on estimating the direction and size of these effects, and on embracing uncertainty, for example, by reporting frequentist confidence intervals or their Bayesian counterparts, credible intervals (Berner & Amrhein, 2022; Smith, 2018).

The causal framework (see Hernán & Robins, 2020; Pearl & Mackenzie, 2018, for an introduction) on the other hand, allows us to explicitly estate our assumptions about how the theoretical and empirical estimands connect to each other and to other variables, and it allows us to identify causal effects, rather than correlations, between these variables. One popular way of representing a causal structure is through directed acyclic graphs (DAGs). In these graphs, nodes represent variables and causal effects are represented by arrows pointing away from one variable to another (e.g., $X \rightarrow Y$, meaning X affects Y). The difference between DAGs and other ways of encoding the causal relationship, such as structural equation models (SEMs), is that while the former encodes the qualitative relationship between them, the latter specifically encodes the form (e.g., linear, additive relationships) of said relationship (Rohrer, 2018).

At a practical level, the assumptions made in a DAG can be used in observational studies for causal inference, that is, to identify causal effects between variables, rather than correlations (see Rohrer, 2018 for an introduction). Assuming that the DAG captures the true causal structure, a set of rules can be applied to determine the sufficient set of variables needed to estimate the actual causal effect of X on Y. These rules eliminate problems commonly present in observational studies that can bias this estimate or induce spurious associations, such as confounding (e.g., Westreich & Greenland, 2013) or collider bias (e.g., Schneider, 2020).

One of the advantages of the combined approach of estimands and causal framework is that the research question is no longer bound by statistical procedures. That is, rather than being used as an equivalent to, for example, scientific inference, importance, or decision making (Hubbard et al., 2019; Navarro, 2019; Wasserstein et al., 2019), statistical inference plays a limited role in this scientific process (i.e., a component of the estimation strategy that intends to learn about the empirical estimand). Other advantages are that methodological choices and conclusions are framed transparently within the corresponding empirical estimand (as opposed to theoretical ones) and that it provides a basis for cumulative knowledge on the subject (see Lundberg et al., 2021).

In the remainder of this article, I will use this combined framework to study the relationship between asymmetry and fitness. First, based on the literature, I outline, for each verbal model proposed to explain the relationship between FA and fitness (i.e., the "good genes," the "good development," and the "growth" hypotheses), a theoretical causal model of this relationship and establish what the theoretical object of inquiry (i.e., theoretical estimand) is. Then, from these general models, I outline the empirical causal model and the target of inference (i.e., empirical estimand) that is the focus of the present study, and the estimation strategy used.

1.2 | Theoretical assumptions: Fluctuating asymmetry and fitness in humans

The definition of FA entails two different interpretations of what an increase in FA means (Klingenberg, 2019; van Dongen & Gangestad, 2011). The first one considers that higher FA values result from the inability of an organism to buffer its development against stochastic variation. In contrast, the second considers that higher values are the result of prolonged or frequent exposure to non-genetic perturbations. For example, an individual may express greater FA because these perturbations occurred during a sensitive window of development (e.g., Oxilia et al., 2021), or because the symmetric structure was continuously exposed to asymmetric environments (chewing: Martinez-Gomis et al., 2009; motor tasks: Aune et al., 2016).

Both interpretations convey different assumptions. Specifically, while the first one focuses on the individual's attributes, the other highlights the environmental context. These different perspectives in turn result in three different descriptive explanations of the mechanisms that explain the relationship between FA and fitness. Figure 1 shows the theoretical causal model derived from each of these verbal models.

Based on the first explanation, the "good genes" hypothesis (Figure 1a) proposes that FA is a signal of good genes or genetic quality, that is, a biomarker with which individuals assess the efficiency of potential mates' buffering mechanisms (Jones et al., 2001; Thornhill & Gangestad, 1993). Because perfect symmetry imposes high physiological costs, it can only be achieved by individuals in excellent condition (Thornhill & Gangestad, 1993), especially if it is expressed in sexual traits that already convey a cost compared to nonsexual traits (Møller, 1991). According to this hypothesis, FA should be related to fitness components (e.g., attractiveness or mating success) because the selection of mates with symmetrical traits would enhance offspring viability (e.g., Møller et al., 1995; Thornhill & Gangestad, 1994).

The "good development" hypothesis (Figure 1b) relies on the second explanation to propose that FA is a biomarker of developmental plasticity under conditions of energetic stress. Specifically, from a life history perspective, the extra energy requirements associated with, for example, compensatory growth (Wells et al., 2006) or metabolically expensive tissues (Longman et al., 2021) would exert a cost in the development of symmetrical traits. In this case, FA would be indirectly related to fitness through the individual's health status because, similar to the previous hypothesis, only individuals in good conditions could afford a symmetric phenotype in harsh circumstances.

A third explanation (Figure 1c) is one in which the expression of FA during development does not require additional costs because it is tightly associated with the phenotype of the individual. Specifically, since traits that grow for longer periods and that are larger will have more opportunities for asymmetry (Leung, 1998; Palmer & Stobeck, 2003), FA in these phenotypes will be related to body size and will be amplified by environmental perturbations affecting growth. In these traits, the relationship between FA and fitness will be confounded if body size also influences the latter (e.g., Walker & Hamilton, 2008). Then, according to this "growth" hypothesis, size
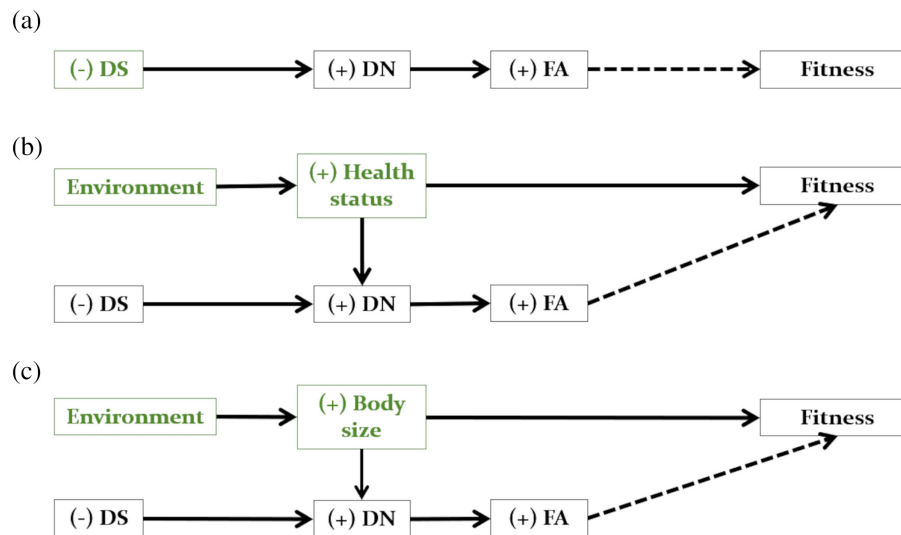
(a)

(b)

(c)

FIGURE 1 Theoretical causal models of the hypotheses on the relationship between fluctuating asymmetry and fitness. Arrows represent a causal effect from one variable (start) to another (end). The dashed arrows depict the theoretical object of inquiry: The relationship between FA and fitness. Following the literature, each model represents the corresponding set of causal assumptions. (a) *Good genes hypothesis*: Reduced DS leads to increased DN and asymmetry. FA is an indicator of the genetic quality (green) of an individual. (b) *Good development hypothesis*: Reduced DS leads to increased DN, but so does a poor health status. Both lead to increased asymmetry. The effect of FA on fitness is confounded (sensu causal inference: FA ← DN ← health status → fitness; Rohrer, 2018) by the effect of the internal and external environment in which the individual develops (green). (c) *Growth hypothesis*: Both reduced DS and larger size lead to an increase in DN. The effect of FA on fitness is also confounded (FA ← DN ← body size → fitness) by the effect of the internal and external environment (green) in which the individual develops. DS, developmental stability; DN, developmental noise; FA, fluctuating asymmetry. See main text for more details.

variation and FA may independently reflect meaningful information about the development of the individual and in those phenotypes in which both traits covary, FA and its consequences on fitness become a combination of the effects of DS, allometry, and environment (e.g., Palmer & Stobeck, 2003). In other words, this hypothesis assumes that FA is indirectly related to fitness through body size but does not exclude the possibility that FA directly affects fitness.

The variety of explanations represented by these hypotheses reflects the implicit complexity that exists in this topic; however, the theoretical estimand remains the same: the causal effect of FA on fitness. In the following, I test these hypotheses focusing on the widely studied topic of facial asymmetries.

2 | MATERIALS AND METHODS

To facilitate the comparison of the three hypotheses, I use observational data from a multigenerational pedigree sample that includes facial FA values and life-history traits related to fitness. Below, for each hypothesis, I detail the study population and the variables of interest and outline the relevant empirical estimands.

2.1 | Sample description

The sample included 314 subjects (207 females, 107 males; mean age = 38.47, SD = 18.02), distributed in 78 extended and nuclear families, with an average of three members per family (range = 0–11).

This dataset was collected in Chiapas province in Mexico. Volunteers who lived in the same geographic area throughout their lives, without previous facial surgery, craniofacial trauma, congenital anomalies, or orthodontic treatment were included in this study. Informed consent was signed by each participant before personal and phenotypic data were collected (Farrera, 2014).

2.2 | Data collection

2.2.1 | 3D photogrammetry imaging technique

The 3D facial shape was captured using photogrammetric methods applied to a series consisting of five separate digital photographs from different angles (left side, left angle: 45°, frontal, right angle: 135°, and right side). All photos were taken at a constant distance of 1.5 m with a standardized photographic protocol described in detail previously (see Quinto-Sánchez et al., 2015). Special care was taken with the hair and earrings. The 3D coordinates were obtained using the software Photomodeler (https://www.photomodeler.com/; Eos Systems, Vancouver, Canada), following the standard recommendations for quality and accuracy of the software. The distance chelion-chelion, measured directly on the individuals using a standard anthropometric caliper, was used as a scale factor.

Forty-two landmarks (17 bilateral and 8 sagittal landmarks) were placed on the photographs trying to avoid data redundancy and following standard terminology (see Figure 2). The points that needed to be located by palpation were marked in situ with a sticker (zygion,
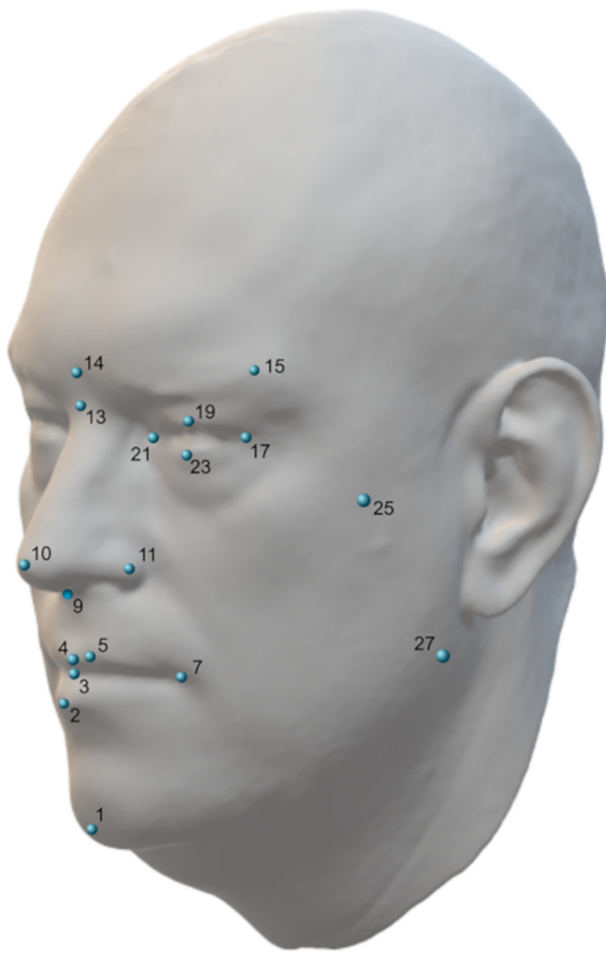
FARRERA

AMERICAN JOURNAL OF
BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists
WILEY 77

FIGURE 2 3D facial scan with digitized landmarks. (1) Gnathion, (2) Labrale inferius, (3) Stomion, (4) Labrale superius, (5, 6) crista philtri, (7, 8) Cheilion, (9) Subnasale, (10) Pronasale, (11, 12) Alare, (13) Nasion, (14) Glabella, (15, 16) Frontotemporale, (17, 18) Exocanthion, (19, 20) Palpebra superior, (21, 22) Endocanthion, (23, 24) Palpebra inferior, (25, 26) Zygion, (27, 28) Gonion. (Artec3D, 2021).

gonion, gnathion, frontotemporale). A subsample of 61 individuals was digitized a second time to evaluate intraobserver variation. A Procrustes ANOVA analysis was performed in this subsample to assess the repeatability of data acquisition in different sessions (Klingenberg & McIntyre, 1998). The results of this Procrustes ANOVA (Table 1) show that the measurement error is one order of magnitude smaller than fluctuating asymmetry. Given the overall goal of the manuscript, this measurement error was considered acceptable for subsequent analysis.

2.2.2 | Measurements

Fluctuating asymmetry

The 3D coordinates of all landmarks were superimposed using the generalized Procrustes analysis (GPA) in the MorphoJ software

(Klingenberg, 2011). This procedure standardizes the configurations of landmarks by eliminating differences in position, size, and orientation. For landmark configurations with object symmetry like the face, the Procrustes fit is performed on the original configurations and their mirror images. The individual scores of FA in units of Mahalanobis distance are obtained from this procedure as the variation of individual asymmetries around the mean asymmetries. These individual scores indicate the magnitude of FA (i.e., the higher the score, the higher the FA) independently of directional asymmetry (Klingenberg, 2015). These scores (mean = 7.49, SD = 0.91; range = 5.40–10.14) were used in subsequent analyses.

Life history traits

Data on life history traits were only available on a subsample of female individuals ($n = 207$; mean age = 39.6, SD = 17.4). The number of offspring (mean = 2.71, SD = 2.41; range = 0–11) was considered as a measure of fitness, and the adult height (mean = 153.2, SD = 5.90; range = 139–167) of the individual as a proxy of their health status (see empirical causal model below).

2.3 | Empirical assumptions: Facial fluctuating asymmetry and the number of offspring

H1. The "good genes" hypothesis: FA encoding genetic quality.

This hypothesis assumes that since facial asymmetries measure DS, FA can be used as a biomarker of genetic quality. Furthermore, it assumes that such an association will impact the reproductive success of the individual, through attractiveness (Møller et al., 1995). The empirical causal model shown in Figure 3a outlines these assumptions. Accordingly, the main empirical estimand is the direct causal effect of facial FA on reproductive success (i.e., number of offspring). This empirical causal model predicts a negative relationship between these variables (see Møller & Thornhill, 1998): individuals with higher facial FA values (i.e., less attractive) would have fewer offspring.

Additionally, this hypothesis assumes that DS, and therefore, FA, must be partly under genetic control for it to respond to selection (i.e., mate choice; Leamy & Klingenberg, 2005). The idea is that, if FA has genetic variability, closely related individuals will have FA values more similar than those who are not. Therefore, a secondary empirical estimand associated with this hypothesis is the heritability (h^2) of FA, expecting values greater than 0 (Figure 3a). I estimated the heritability of FA using a generalized linear mixed model via Markov chain Monte Carlo methods using the MCMCglmm, R package (Hadfield, 2010). The additive genetic variance and the residual environmental variance of FA were estimated using the pedigree information of the total sample. Weak priors were set for both variance components and, based on the empirical causal model (Figure 3a) of this hypothesis, no fixed effects were included. The MCMCglmm was run for 10,000,000 iterations with a burn-in period of 2000 iterations and a thinning interval of 1000.

Effect	SS	MS	df	F	p (Param.)
Individual	0.50767632	0.0001239747	4095	6.20	<0.0001
Directional asymmetry	0.00863661	0.0001542253	56	7.72	<0.0001
Fluctuating asymmetry	0.07273774	0.0000199829	3640	5.12	<0.0001
Measurement error	0.03065837	0.0000039035	7854		

TABLE 1 Results of Procrustes ANOVA displaying the measurement error in shape variables
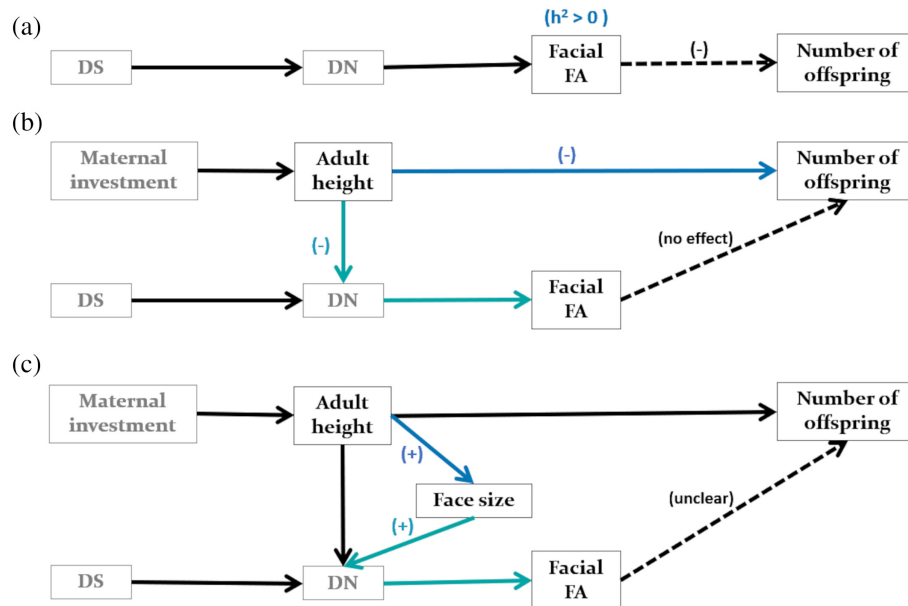

FIGURE 3 Empirical causal models of the relationship between facial fluctuating asymmetry and reproductive success (i.e., number of offspring) according to the set of assumptions derived from the (a) "good genes," (b) "good development" and (c) "growth" hypotheses. Arrows represent a causal effect from one variable (start) to another (end). Dashed arrows depict the main empirical estimand: The causal effect of facial FA on the number of offspring. Secondary empirical estimands are shown in blue. Predictions for each empirical estimand are shown in parentheses. Variables in black correspond to the dataset analyzed, while those in gray represent unmeasured variables. DS, developmental stability; DN, developmental noise; FA, fluctuating asymmetry. See main text for more details.

H2. The "good development" hypothesis: FA encoding developmental plasticity.

The empirical causal model for this hypothesis was generated based on the work of Wells (2018, 2019), which incorporates the maternal phenotype as the main environmental factor that impacts the development of the offspring. According to this verbal model greater maternal investment in both pregnancy and lactation would favor in the offspring the allocation of energy to somatic growth and maintenance and, therefore, would result in large adult size. In this scenario, offspring quality is favored over quantity. In contrast, a reduced maternal investment would favor the allocation of energy to reproduction, which would result in small adult size. If the post-natal energy supply improves in the latter scenario, the small size will also be accompanied by elevated fat stores and an increased risk of disease. In both scenarios, offspring quantity would be favored over quality.

Based on these intergenerational maternal effects, Figure 3b shows the empirical causal model generated for the "good development" hypothesis. In contrast to the previous hypothesis, the model does not predict a direct causal effect, instead, the main empirical estimand in this model is the indirect causal effect of facial FA on the

number of offspring through adult size. A secondary empirical estimand derived from this model is the causal effect of adult height on the reproductive success of the individual (e.g., Wells, 2018). Therefore, a negative relationship between variables is expected: individuals who developed in poor conditions (i.e., short stature) would have more offspring and vice versa. Another secondary empirical estimand derived is the causal effect of adult height on facial FA through developmental noise. Because individuals who developed in suboptimal conditions (i.e., short stature) could not afford symmetric traits (e.g., Longman et al., 2021), a negative relationship between these variables is expected: shorter individuals would have greater facial FA values (e.g., Özener & Ertuğrul, 2011).

H3. The "growth" hypothesis: FA as a by-product of body size.

This hypothesis is tested using the previous empirical causal model with two additional assumptions (Figure 3c): that body height and face size are positively allometrically related in adults (e.g., Mitteroecker et al., 2013) and that face size has a causal effect on FA because more growth gives more opportunities for asymmetry (e.g., Palmer & Stobeck, 2003). The literature is not clear on whether the effect of

FARRERA

AMERICAN JOURNAL OF
BIOLOGICAL ANTHROPOLOGY
The Official Journal of the American Association of Biological Anthropologists

WILEY 79

facial FA on the number of offspring is direct or indirect, so this model has two main empirical estimands: a direct causal effect as in the first model or an indirect causal effect as in the second one. The first assumption predicts an association between these variables, while the second predicts no direct causal effect. There are also two secondary empirical estimands: the causal effect of height on face size, and the effect of face size on facial FA. In both cases, a positive relationship is expected: taller individuals will exhibit bigger faces and individuals with larger faces will exhibit larger facial FA values, respectively.

2.4 | Estimation strategy

The main empirical estimand is the same across hypotheses: the causal effect of facial FA on the number of offspring, conditional on the empirical causal structure assumed (empirical DAG). Given that the estimation of this effect involves life-history traits, all three hypotheses were evaluated using the subsample of females.

The strategy to estimate the causal effect of facial FA on the number of offspring given the corresponding empirical causal model was, first, to use an automated causal modeling tool (R package: *Dagitty*, Textor et al., 2016) to identify the right covariates. This tool allows one to test whether the assumptions encoded in the DAGs are consistent with the data and to test whether the inferences made in the original DAGs are valid for a range of different but statistically similar DAGs (Textor et al., 2016). The DAG that resulted from this process was consistent with the data and robust against different causal specifications, and therefore, used in the following analysis. Then this DAG was used to identify, for each empirical estimand, the set of covariates to be adjusted for in the statistical models. Poisson regression was used for the analysis of count data, whereas linear regression was used for continuous data.

For H1, conditional on the empirical DAG, the Poisson regression model estimated the effect of facial FA on the number of offspring without adjusting for any variable.

For H2, conditional on the empirical DAG, the causal effect of facial FA on the number of offspring was estimated using a Poisson regression model adjusted for height. On the other hand, the effect of height on the number of offspring was estimated using a Poisson regression but given the DAG, no adjustment was necessary. To estimate the effect of height on facial FA assumed in this hypothesis, a linear model was fit without adjusting for any covariate.

For H3, conditional on the empirical DAG, the Poisson regression model between facial FA and the number of offspring was fit both without adjustment and adjusting for height. Additionally, the effect of height on facial size was estimated using a linear model without adjusting for any covariate. Finally, the effect of face size on facial FA was estimated using a linear model adjusted for height.

Continuous variables were standardized before the analysis. Data analysis was conducted using the R language and environment (R Core Team, 2020), version 1.2.5033 via the RStudio integrated development environment.

3 | RESULTS

The results for H1 show that there was a positive effect ($\beta = 0.323$) of facial FA on the number of offspring, with possible values for this parameter that were most compatible with the data ranging from 0.244 to 0.402 (95% confidence interval). In other words, conditional on the DAG, individuals with higher asymmetries have more children. Additionally, the generalized linear mixed model provided a mean heritability of facial FA close to zero ($h^2 = 0.046$) with possible values ranging from 0.0002 to 0.174 (95% confidence interval), which means that almost none or little of the phenotypic variation in FA is explained by genetic variation.

The results for H2 show that there was a positive effect ($\beta = 0.294$) of facial FA on the number of offspring conditional on the individual's adult height, with possible values for this parameter ranging from 0.213 to 0.374 (95% confidence interval). This means that individuals with higher FA values for their height have more offspring. In addition, there was a negative effect ($\beta = -0.275$) of the adult height on the number of offspring, with possible values ranging from −0.359 to −0.191 (95% confidence interval). In other words, those individuals who are taller have fewer offspring. Finally, there was a negative effect ($\beta = -0.13$) of height on facial FA, with possible values ranging from −0.266 to 0.006 (95% confidence interval). This means that taller individuals exhibit lower facial FA.

The results for H3 show that the direct and indirect effects of facial FA on the number of offspring calculated under this model are the same as those effects obtained in the first and second hypotheses, respectively. Moreover, there was a very low positive effect of height on face size ($\beta = 0.01$) with possible values for this parameter that were most compatible with the data in a relatively wide range from −0.127 to 0.148 (95% confidence interval). Finally, the results show a very low negative effect ($\beta = -0.04$) of face size on facial FA with possible values in a relatively wide range from −0.179 to 0.093 (95% confidence interval).

4 | DISCUSSION

In this contribution, I evaluated three of the most common verbal models used to understand the relationship between FA and fitness in humans: the "good genes," the "good development," and the "growth" hypotheses. For this purpose, I generated formal models (i.e., estimands and causal frameworks) for each hypothesis and tested them in the particular case of facial asymmetries and reproductive success.

4.1 | Theoretical assumptions

The present study shows that even if the approaches are different, some of the theoretical assumptions overlap across hypotheses (Figure 1), opening the opportunity for a unified formal model. Nonetheless, they show differences in two key assumptions. First, these

hypotheses differ in whether they consider that FA reflects some cost to the individual, distinguishing between FA as a reliable signal of DS that reflects the quality of the individual (H1 and H2: symmetrical traits are costly) and as a reliable signal that requires no additional cost because it is tightly associated with some attribute of the individual (H3: allometric constraints that link body size and FA). This distinction has been discussed mainly in the framework of signaling theory (Barker et al., 2019), but in the context of human asymmetries and fitness, this discussion is currently problematic primarily because the way these concepts have been applied overlooks recent conceptual advances.

From the framework of signaling theory, attributes other than physiological information are recognized as signals (e.g., embodied capital or noetic attributes, Barker et al., 2019). A broader concept like this would allow for more comprehensive verbal models of the relationship between FA and fitness in humans, in which cultural practices such as the use of makeup (Killian et al., 2018), and social norms like standards of beauty (Kleisner et al., 2017) are also included in the interpretation and scope of the research. Signaling theory also recognizes that the way multiple signals are integrated with each other and with socioecological factors is an important source of information (Patricelli & Hebets, 2016). This would promote studying asymmetry along with other types of signals, as has been done during the last decade on topics such as mate choice (Jones & Jaeger, 2019; van Dogen et al., 2020) or the individual's health status (Foo et al., 2017; Mogilski & Welling, 2017). Addressing multiple signals as an integrated signaling phenotype or explaining how they are theoretically related to each other (e.g., Luoto et al., 2021) could improve and extend our understanding of the topic. Moreover, instead of being considered a static measurement (i.e., values computed at one point in time), individual asymmetries could be explored over different timescales. In the dynamic context of face-to-face interaction, for instance, asymmetric facial movement can be perceived as unattractive, regardless of the static asymmetry score of the individual (Hughes & Aung, 2018) because, for example, it conveys information about the sender's age (Kamachi et al., 2019). Taking into account that the causes and effects of asymmetry can be different in static and dynamic contexts could also clarify some of the contradictory evidence on the subject.

Another theoretical assumption in which the hypotheses evaluated differ is whether they highlight the role of developmental plasticity (i.e., phenotypic adjustments in response to the environment) on the expression of phenotypic variation and, particularly, on the production of asymmetric traits. Specifically, this assumption differentiates between research on FA variation that focuses on its genetic basis (H1: symmetry reflects good genes) and research that focuses on the development pathways that lead to such within-individual variation (H2 and H3: symmetry reflects the interplay between the organism and its circumstances). Although the former ignores the idea that has been present since the 1980s in the field of evolutionary developmental biology (Müller, 2007) that the influence of genotype on the phenotype is structured by developmental processes, the role of development in the latter is not entirely clear either. New verbal and formal models with a different set of theoretical assumptions are needed to get a better, refined representation of the role of development in the relationship between FA and fitness in humans.

4.2 | Empirical assumptions

This study also shows some similarities and differences between hypotheses when the results are compared with the expectations derived from the empirical assumptions. In the case of the "good genes" hypothesis, the corresponding empirical causal model expects that individuals with greater facial FA values have less reproductive success than individuals with less asymmetry. Moreover, it expects that part of the phenotypic variance of facial FA is explained by genetic variation. In contrast, I found that individuals with greater facial FA values have more offspring and a heritability close to zero (i.e., almost none or little facial FA variation is explained by genetic variation). The latter result is the first report of heritability of FA in the human face and is consistent with previous research showing very low or no heritability of other traits in humans and other species (Johnson et al., 2008; Leamy & Klingenberg, 2005). These results suggest that the empirical causal model for this hypothesis needs to be revised and refined.

The empirical causal model derived from the "good development" hypothesis posits that because FA is the result of poor health, no direct link should be found between facial FA and the number of offspring. In contrast, the results showed a positive effect of facial FA on the number of offspring. On the other hand, based on the intergenerational maternal effect (Wells, 2018), this empirical causal model assumes that the short stature of some individuals is the result of a suboptimal maternal niche and that individuals who develop under these conditions may favor quantity over quality of offspring, and vice versa. The results of this study were compatible with this assumption. Specifically, it was found that, regardless of asymmetry, individuals with poor health status (measured as adult height) had more children, an effect reported in some previous studies (e.g., Krzyzanowska et al., 2015), but not in others (e.g., Helle, 2008). The results were also compatible with the expected negative effect of height on facial FA in this hypothesis, an effect reported in previous studies (Kirchengast, 2019; Özener & Ertuğrul, 2011) using FA measurements of non-facial traits. In other words, these results are compatible with most of the assumptions derived from the empirical causal model for this hypothesis, except for the assumption of no direct link between facial FA and the number of offspring, which should be refined to include potential mechanisms that may explain the relationship between these variables.

The empirical assumptions of the "growth" hypothesis are unclear as to whether facial FA directly and/or indirectly influences the number of offspring, what would be the expected direction of this effect, or what mechanism would be responsible. Therefore, it is currently not possible to interpret the results obtained on this assumption. Nonetheless, this empirical causal model posits two additional assumptions. First, that body and face size are allometrically related in adults and that facial FA is a by-product of individual growth. In contrast to previous studies (e.g., Gateño et al., 2018; Mitteroecker

et al., 2013), the results were compatible, with high uncertainty, with an effect close to zero. These results suggest that more discussion is needed on the empirical causal model derived for this hypothesis.

In all hypotheses, I found a positive association between facial FA and the number of offspring, which is not consistent with any of the three empirical causal models evaluated. This result suggests that additional explanatory variables should be formally included in these models to further understand and test this relationship. One candidate variable could be the age-dependent pattern of FA expression (e.g., Wilson & Manning, 1996). Since facial FA can be a by-product of soft tissue aging, older individuals may express higher values. Further, this link could be related to the number of offspring in two ways. First, in line with the "good development" hypothesis, since reproduction takes time and considerable metabolic demands, individuals who have reproduced more and are older may also be more asymmetric. Second, in line with the "growth" hypothesis, fully developed (older and bigger) and therefore more asymmetric individuals could be those who have also had more opportunities to reproduce. Datasets collected specifically for testing these verbal models and updated formal models are needed to confirm the role of aging or any other variable outside those proposed in this work.

There are at least two factors related to the estimation strategy that limit the interpretation of these results (section 4.2). One of them is the sample over which inferences were drawn. The dataset used in this study was not explicitly collected to answer the theoretical object of inquiry (i.e., the relationship between FA and fitness), and thus, the empirical causal models were designed after data collection, instead of before as required to warrant causal claims (Rohrer, 2018). Other potential factors are related to bias in the computation of FA values, which have been extensively reviewed elsewhere (Graham, 2021b; Graham et al., 2010), including the presence of other forms of asymmetry, measurement error, or mixtures of additive and multiplicative errors. These limitations suggest that these results (section 4.1) must be replicated using more rigorous estimation strategies and other databases that allow comparing the three hypotheses.

Future studies could further benefit from revising, in light of theory development, the statistical practice associated with FA. For instance, rethinking isolated FA values as a target of inquiry when evidence suggests that in some contexts it is common to find different forms of asymmetry together (e.g., human face: Farrera et al., 2015; Quinto-Sánchez et al., 2015). Formal models of descriptive explanations that instead address the dynamics that could give rise to patterns of asymmetric mixtures (e.g., Graham et al., 1993; Hallgrímsson, 1998) could shed new light on the topic or clarify existent evidence.

5 | CONCLUSIONS

The overlap between theoretical and empirical assumptions across hypotheses supports the idea that the relationship between asymmetric variation and fitness cannot be understood using only one of them, but rather requires a general model that integrates different explanations on this topic. However, before a unified framework can be developed, several theoretical and empirical assumptions of the three most common hypotheses on this subject need to be revised and updated.

Altogether, the results of this study suggest that the "good genes" hypothesis needs to be reformulated for several reasons. The first of which is because its theoretical assumptions have not been revised and updated since its development in the 1990s, despite conceptual advances in signaling theory and evolutionary development to understand phenotypic variation. Another reason is that, although early studies showed support for this hypothesis (e.g., Møller & Thornhill, 1998; Perret et al., 1999), the present study and accumulated evidence show little or no support (e.g., Foo et al., 2017; Kleisner et al., 2017; Kočnara et al., 2019; Kruuk et al., 2003; Palmer, 1999; Zheng et al., 2021), particularly under naturalistic settings (Jones & Jaeger, 2019).

The results of this study were compatible with most of the empirical assumptions made for the "good development" hypothesis. In contrast, it is not yet clear whether its theoretical assumptions hold, and to what extent they need to be extended or refined, mainly because despite being formulated in the early 2000s (Wells et al., 2006), only a few studies to date have addressed it directly (Kirchengast, 2019; Longman et al., 2021; Özener & Ertuğrul, 2011). Finally, the present study is less informative about the "growth" hypothesis because the literature is not clear about the theoretical assumptions on the relationship between FA and fitness. This makes it impossible to further assess the empirical assumptions derived from them. Additionally, the results of this study were not compatible with the secondary empirical assumptions derived from this hypothesis.

AUTHOR CONTRIBUTIONS
Arodi Farrera: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); writing – original draft (lead).

CONFLICT OF INTEREST
The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT
The source code for generating all the analyses in section 2.4 is available here: https://github.com/arodifr/FA_fitness.git. Due to the nature of the pedigree data, the data used for section 2.3. H1 are not available.

82 | WILEY | AMERICAN JOURNAL OF **BIOLOGICAL ANTHROPOLOGY** The Official Journal of the American Association of Biological Anthropologists

FARRERA

ORCID

Arodi Farrera https://orcid.org/0000-0001-9751-9906

REFERENCES

Artec 3D. (2021). 3D face scan. Professional 3D scanning solutions. https://www.artec3d.com/3d-models/face.

Aune, T., Ettema, G., & Vereijken, B. (2016). Bilateral asymmetry in upper extremities is more pronounced in distal compared to proximal joints. *Journal of Motor Behavior*, 48(2), 143–152. https://doi.org/10.1080/00222895.2015.1056766

Barker, J., Power, E., Heap, S., Puurtinen, M., & Sosis, R. (2019). Content, cost, and context: A framework for understanding human signaling systems. *Evolutionary Anthropology*, 28(2), 86–99. https://doi.org/10.1002/evan.21768

Berner, D., & Amrhein, V. (2022). Why and how we should join the shift from significance testing to estimation. *Journal of Evolutionary Biology.*, 35, 777–787. https://doi.org/10.1111/jeb.14009

Brown, W., Price, M., Kang, J., Pound, N., Zhao, Y., & Yu, H. (2008). Fluctuating asymmetry and preferences for sex-typical bodily characteristics. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35), 12938–12943. https://doi.org/10.1073/pnas.0710420105

Clarke, G. (1995). Relationship between developmental stability and fitness: Applications for conservation biology. *Coservation Biology*, 9(1), 18–24.

Debat, V., & David, P. (2001). Mapping phenotypes: Canalization, plasticity and developmental stability. *Trends in Ecology and Evolution*, 16(10), 555–561. https://doi.org/10.1016/S0169-5347[01]02266-2

Farrera, A. (2014). *Heritability of the facial traits in a sample of Mexican population. A quantitative genetics study* (Master's thesis). Universidad Nacional Autónoma de México (in Spanish). https://ru.dgb.unam.mx/handle/DGB_UNAM/TES01000715270

Farrera, A., Villanueva, M., Quinto-Sánchez, M., & González-José, R. (2015). The relationship between FA and attractiveness in Mexican students. *American Journal of Human Biology*, 27(3), 387–396. https://doi.org/10.1002/ajhb.22657

Foo, Y., Simmons, L., & Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Scientific Reports*, 7, 39731. https://doi.org/10.1038/srep39731

Frigg, R., & Hartmann, S. (2020). Models in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. (Spring 2020 Edition). Metaphysics Research Lab, Standford University. https://plato.stanford.edu/archives/spr2020/entries/models-science/

Gangestad, S., Merriman, L., & Thompson, M. (2010). Men's oxidative stress, fluctuating asymmetry and physical attractiveness. *Animal Behavior*, 80, 1005–1013. https://doi.org/10.1016/j.anbehav.2010.09.003

Gateño, J., Jones, T., Shen, S., Chen, K., Jajoo, A., Kuang, T., English, J., Nicol, M., Teichgraeber, J., & Xia, J. (2018). Fluctuating asymmetry of the normal facial skeleton. *International Journal of Oral and Maxillofacial Surgery*, 47(4), 534–540. https://doi.org/10.1016/j.ijom.2017.10.011

Graham, J. (2021a). Nature, nurture, and noise: Developmental instability, fluctuating asymmetry, and the causes of phenotypic variation. *Symmetry*, 13, 1204. https://doi.org/10.3390/sym13071204

Graham, J. (2021b). Fluctuating asymmetry and developmental instability, a guide to best practice. *Symmetry*, 13, 9. https://doi.org/10.3390/sym13010009

Graham, J., Freeman, D., & Emlen, J. (1993). Antisymmetry, directional asymmetry, and dynamic morphogenesis. *Genetica*, 89, 121–137. https://doi.org/10.1007/BF02424509

Graham, J., & Özener, B. (2016). Fluctuating asymmetry of human populations: A review. *Symmetry*, 8, 154. https://doi.org/10.3390/sym8120154

Graham, J., Raz, S., Hel-Or, H., & Nevo, E. (2010). Fluctuating asymmetry: Methods, theory, and applications. *Symmetry*, 2, 466–540. https://doi.org/10.3390/sym2020466

Hadfield, J. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22. https://doi.org/10.18637/jss.v033.i02

Hallgrímsson, B. (1998). Fluctuating asymmetry in the mammalian skeleton. *Evolutionary and developmental implications. Evolutionary Biology*, 30, 187–251. https://doi.org/10.1007/978-1-4899-1751-5_6

Hallgrímsson, B., Willmore, K., & Hall, B. (2002). Canalization, developmental stability, and morphological integration in primate limbs. *Yearbook of Physical Anthropology*, 119(S35), 131–158. https://doi.org/10.1002/ajpa.10182

Helle, S. (2008). A tradeoff between reproduction and growth in contemporary Finnish women. *Evolution and Human Behavior*, 29(3), 189–195. https://doi.org/10.1016/j.evolhumbehav.2007.11.009

Hernán, M., & Robins, J. (2020). *Causal inference: What if*. Chapman & Hall.

Hubbard, R., Haig, B., & Parsa, R. (2019). The limited role of formal statistical inference in scientific inference. *The American Statistician*, 73, 91–98. https://doi.org/10.1080/00031305.2018.1464947

Hughes, S., & Aung, T. (2018). Symmetry in motion: Perception of attractiveness changes with facial movement. *Journal of Nonverbal Behavior*, 42, 267–283. https://doi.org/10.1007/s10919-018-0277-4

Johnson, W., Gangestad, S., Segal, N., & Bouchard, T., Jr. (2008). Heritability of fluctuating asymmetry in a human twin sample: The effect of trait aggregation. *American Journal of Human Biology*, 20(6), 651–658.

Jones, A., & Jaeger, B. (2019). Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry*, 11, 279. https://doi.org/10.3390/sym11020279

Jones, B., Little, A., Penton-Voak, I., Tiddeman, B., Burt, D., & Perret, D. (2001). Facial symmetry and judgements of apparent health. Support for a "good genes" explanation of the attractiveness-symmetry relationship. *Evolution and Human Behavior*, 22(6), 417–429. https://doi.org/10.1016/S1090-5138(01)00083-6

Kamachi, M., Chiba, T., Kurosumi, M., & Mizukoshi, K. (2019). Perception of human age from faces: Symmetric versus asymmetric movement. *Symmetry*, 11, 650. https://doi.org/10.3390/sym11050650

Killian, A., Mitra, S., & Peissing, J. (2018). The role of regional contrast changes and asymmetry in facial attractiveness related to cosmetic use. *Frontiers in Psychology*, 9, 2440. https://doi.org/10.3389/fpsyg.2018.02448

Kirchengast, S. (2019). Asymmetry patterns are associated with body size and somatic robustness among adult !Kung San and Kavango people. *Anthropological Review*, 82(1), 43–63. https://doi.org/10.2478/anre-2019-0004

Kleisner, K., Kočnara, T., Tureček, P., Stella, D., Akoko, R., Třebický, V., & Havlíček, J. (2017). African and European perception of African female attractiveness. *Evolution and Human Behavior*, 38(6), 744–755. https://doi.org/10.1016/j.evolhumbehav.2017.07.002

Klingenberg, C. (2011). MorphoJ: An integrated software package for geometric morhpmetrics. *Molecular Ecology Resources*, 11(2), 353–357. https://doi.org/10.1111/j.1755-0998.2010.02924.x

Klingenberg, C. (2015). Analyzing fluctuating asymmetry with geometric morphometrics: Concepts, methods and applications. *Symmetry*, 7, 843–934. https://doi.org/10.3390/sym702084

Klingenberg, C. (2019). Phenotypic plasticity, developmental instability, and robustness: The concepts and how they are connected. *Frontiers in Ecology and Evolution*, 7, 56. https://doi.org/10.3389/fevo.2019.00056

Klingenberg, C., & McIntyre, G. (1998). Geometric morphometrics of developmental instability: Analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution*, 52(5), 1363–1375.

Kočnara, T., Saribay, A., & Kleisner, K. (2019). Perceived attractiveness of Czech faces across 10 cultures: Associations with sexual shape dimorphism, averageness, fluctuating asymmetry, and eye color. *PLoS One*, 14(11), e0225549. https://doi.org/10.1371/journal.pone.0225549

Kordsmeyer, T., & Penke, L. (2017). The association of three indicators of developmental instability with mating success in humans. *Evolution and Human Behavior*, *38*, 704–713. https://doi.org/10.1016/j.evolhumbehav.2017.08.002

Kruuk, L., Slate, J., Pemberton, J., & Clutton-Brock, T. (2003). Fluctuating asymmetry in a secondary sexual trait: No associations with individual fitness, environmental stress or inbreeding, and no heritability. *Journal of Evolutionary Biology*, *16*, 101–113. https://doi.org/10.1046/j.1420-9101.2003.00503.x

Krzyzanowska, M., Mascie-Taylor, N., & Thalabard, J.-C. (2015). Is human mating for height associated with fertility? Results from a British national cohort study. *American Journal of Human Biology*, *27*, 553–563. https://doi.org/10.1002/ajhb.22684

Leamy, L., & Klingenberg, C. (2005). The genetics and evolution of fluctuating asymmetry. *Annual Review of Ecology, Evolution, and Systematics*, *36*, 1–21. https://doi.org/10.1146/annurev.ecolsys.36.102003.152640

Leung, B. (1998). Correcting for allometry in studies of fluctuating asymmetry and quality within samples. *Proceedings of the Royal Society of London B*, *265*, 1623–1629. https://doi.org/10.1098/rspb.1998.0480

Little, R. J., & Lewis, R. J. (2021). Estimands, estimators, and estimates. *Journal of the American Medical Association*, *326*(10), 967–968. https://doi.org/10.1001/jama.2021.2886

Longman, D., Oyama, S., Cracknell, J., Thompson, N., Gordon, D., Stock, J., & Wells, J. (2021). Fluctuating asymmetry, a marker of poor growth quality, is associated with adult male metabolic rate. *American Journal of Physical Anthropology*, *175*, 646–655. https://doi.org/10.1002/ajpa.24276

Lundberg, I., Johnson, R., & Stweart, B. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *American Sociological Review*, *86*(3), 532–565. https://doi.org/10.1177/00031224211004187

Luoto, S., Krama, T., Rubika, A., Borráz-León, J., Trakimas, G., Elferts, D., Skrinda, I., Krams, R., et al. (2021). Socioeconomic position, immune function, and its physiological markers. *Psychoneuroendocrinology*, *127*, 105202. https://doi.org/10.1016/j.psyneuen.2021.105202

Martinez-Gomis, J., Lujan-Climent, M., Palau, S., Bizar, J., Salsench, J., & Peraire, M. (2009). Relationship between chewing side preference and handedness and lateral asymmetry of peripheral factors. *Archives of Oral Biology*, *54*, 101–107. https://doi.org/10.1016/j.archoralbio.2008.09.006

Mitteroecker, P., Gunz, P., Windhager, S., & Schaefer, K. (2013). A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix the Italian Journal of Mammalogy*, *24*(1), 59–66. https://doi.org/10.4404/hystrix-24.1-6369

Mogilski, J., & Welling, L. (2017). The relative importance of sexual dimorphism, fluctuating asymmetry, and color cues to health during evaluation of potential partners' facial photographs. A conjoint analysis study. *Human Nature*, *28*, 53–75. https://doi.org/10.1007/s12110-016-9277-4

Møller, A. (1991). Sexual ornament size and the cost of fluctuating asymmetry. *Proceedings: Biological Sciences*, *243*(1306), 59–62. https://doi.org/10.1098/rspb.1991.0010

Møller, A., Soler, M., & Thornhill, R. (1995). Breast asymmetry, sexual selection and human reproductive success. *Ethology and Sociobiology*, *16*(3), 207–219. https://doi.org/10.1016/0162-3095[95]00002-3

Møller, A., & Thornhill, R. (1998). Bilateral symmetry and sexual selection: A meta-analysis. *The American Naturalist*, *151*(2), 174–192. https://doi.org/10.1086/286110

Müller, G. (2007). Evo-devo: Extending the evolutionary synthesis. *Nature Reviews Genetics*, *8*, 943–949. https://doi.org/10.1038/nrg2219

Navarro, D. (2019). Between the devil and the deep blue sea: Tensions between scientific judgement and statistical model selection. *Computational Brain and Behavior*, *2*, 28–34. https://doi.org/10.1007/s42113-018-0019-z

Oxilia, G., Sartorio, J., Bortolini, E., Zampirolo, G., Papini, A., Boggioni, M., Martini, S., Marciani, F., et al. (2021). Exploring direction and fluctuating asymmetry in the human palate during growth. *American Journal of Physical Anthropology*, *175*, 847–864. https://doi.org/10.1002/ajpa.24293

Özener, B., & Ertuğrul, B. (2011). Relationship between shortness of final body height and fluctuating asymmetry in Turkish young males. *Annals of Human Biology*, *38*(1), 34–38. https://doi.org/10.3109/03014460.2010.486770

Palmer, R. (1999). Detecting publication bias in meta-analyses: A case study of fluctuating asymmetry and sexual selection. *The American Naturalist*, *154*(2), 220–233. https://doi.org/10.1086/303223

Palmer, R. (2000). Quasireplication and the contract of error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics*, *31*, 441–480. https://doi.org/10.1146/annurev.ecolsys.31.1.441

Palmer, R., & Stobeck, C. (2003). Fluctuating asymmetry analyses revisited. In M. Polak (Ed.), *Developmental instability: Causes and consequences* (pp. 279–319). Oxford University Presss.

Palmer, R., & Strobeck, C. (1992). Fluctuating asymmetry as a measure of developmental stability: Implications of non-normal distributions and power of statistical tests. *Acta Zool Fennica*, *191*, 57–72.

Patricelli, G., & Hebets, E. (2016). New dimensions in animal communication: The case for complexity. *Current Opinion in Behavioral Sciences*, *12*, 80–89. https://doi.org/10.1016/j.cobeha.2016.09.011

Pearl, J., & Mackenzie, D. (2018). *The book of why. The new science of cause and effect*. Basic Books.

Perret, D., Burt, M., Penton-Voak, I., Lee, K., Rowland, D., & Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior*, *20*, 295–307. https://doi.org/10.1016/S1090-5138(99)00014-8

Quinto-Sánchez, M., Adhikari, K., Acuña-Alonzo, V., Cintas, C., Silva de Cerqueira, C., Ramallo, V., Castillo, L., Farrera, A., et al. (2015). Facial asymmetry and genetic ancestry in Latin American admixed populations. *American Journal of Physical Anthropology*, *157*, 58–70. https://doi.org/10.1002/ajpa.22688

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.R-project.org/

Robinaugh, D., Haslbeck, J., Ryan, O., Fried, E., & Waldorp, L. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*(4), 725–743. https://doi.org/10.1177/1745691620974697

Rohrer, J. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Schneider, E. (2020). Collider bias in economic history research. *Explorations in Economic History*, *78*, 101356. https://doi.org/10.1016/j.eeh.2020.101356

Smaldino, P. (2017). Models are stupid, and we need more of them. In R. Vallacher, S. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 311–331). Routledge.

Smaldino, P. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, *51*(4), 207–218. https://doi.org/10.1027/1864-9335/a000425

Smith, R. (2018). The continuing misuse of null hypothesis significance testing in biological anthropology. *American Journal of Physical Anthropology*, *166*(1), 236–245. https://doi.org/10.1002/ajpa.23399

Strode, P. (2015). Hypothesis generation in biology: A science teaching challenge and potential solution. *The American Biology Teacher*, *77*(7), 500–506. https://doi.org/10.1525/abt.2015.77.7.4

Textor, J., van der Zander, B., Gilthorpe, M., Liśkiewicz, M., & Ellison, G. (2016). Robust causal inference using directed acyclic graphs: The R

package 'dagitty'. *International Journal of Epidemiology*, *45*(6), 1887–1894. https://doi.org/10.1093/ije/dyw341

Thornhill, R., & Gangestad, S. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature*, *4*(3), 237–269. https://doi.org/10.1007/BF02692201

Thornhill, R., & Gangestad, S. (1994). Human fluctuating asymmetry and sexual behavior. *Psychological Science*, *5*(5), 297–302.

van Dogen, S., Ekrami, O., & Claes, P. (2020). Lack of correlation between facial sexual dimorphism, fluctuating asymmetry and self-perceived attractiveness in men and women. *Symmetry*, *12*, 236. https://doi.org/10.3390/sym12020236

van Dongen, S., & Gangestad, S. (2011). Human fluctuating asymmetry in relation to health and quality: A meta-analysis. *Evolution and Human Behavior*, *32*, 380–398. https://doi.org/10.1016/j.evolhumbehav.2011.03.002

Walker, R., & Hamilton, M. (2008). Life-history consequences of density dependence and the evolution of human body size. *Current Anthropology*, *49*(1), 115–122. https://doi.org/10.1086/524763

Wasserstein, R., Schirm, A., & Lazar, N. (2019). Moving to a world beyond "p < 0.05". *The American Statistician*, *73*, 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wells, J. (2018). Life history trade-offs and the partitioning of maternal investment. Implications for health of mothers and offspring. *Evolution, Medicine, and Public Health*, *1*, 153–166. https://doi.org/10.1093/emph/eoy014

Wells, J. (2019). Developmental plasticity as adaptation: Adjusting to the external environment under the imprint of maternal capital. *Philosophical Transactions of the Royal Society B*, *374*, 20180122. https://doi.org/10.1098/rstb.2018.0122

Wells, J., Hallal, P., Manning, J., & Victora, C. (2006). A trade-off between early growth rate and fluctuating asymmetry in Brazilian boys. *Annals of Human Biology*, *33*(1), 112–124. https://doi.org/10.1080/03014460500480391

Westreich, D., & Greenland, S. (2013). The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology*, *177*(4), 292–298. https://doi.org/10.1093/aje/kws412

Wilson, J., & Manning, J. (1996). Fluctuating asymmetry and age in children: Evolutionary implications for the control of developmental stability. *Journal of Human Evolution*, *30*, 529–537. https://doi.org/10.1006/jhev.1996.0041

Zheng, R., Ren, D., Xie, C., Pan, J., & Zhou, G. (2021). Normality mediates the effect of symmetry on facial attractiveness. *Acta Psychologica*, *217*, 103311. https://doi.org/10.1016/j.actpsy.2021.103311