

BIOLOGICAL ESSENTIALISM, GENDER, TRUE BELIEF, CONFIRMATION BIASES, AND SKEPTICISM

Patricia Adair Gowaty

In this chapter, I describe a damning case study demonstrating that true beliefs—a type of common self-deception promoting theory tenacity (Loehle, 1987) and commitment to essential sex differences—have hijacked the scientific study of the evolutionary origins of sex differences and similarities. Theory tenacity is the “persistent belief in a theory despite contrary evidence” (Loehle, 1987, p. 397), and essentialism is the idea of fundamental, intrinsic, necessary, determinative differences between entities, such as females and males. Essentialism is a prominent and usually unconsciously held conceptual frame (i.e., a perception bias) for evolutionary biologists studying the origins of sex differences. It is a philosophical trap closing off attention to alternative hypotheses and experimental designs that could be used to reject false hypotheses. Here, I describe observations suggesting that the belief in a theory despite contrary evidence (Loehle, 1987) is strongly associated with intuitive gender schemas and gendered conceptual frames (see Chapter 7, this volume). I argue that intuitive gender schemas have fueled theory tenacity and inhibited the healthy development of habitual skepticism, thus retarding progress in the evolutionary study of sex differences and similarities.

I suspect that typical gender schemas, consistent with the double standard in human behavioral, social, and sexual norms, may also promote the use of confirmatory designs in the empirical practice

of some sex differences research. Furthermore, I suspect that typical gender schemas have obscured the visionary works of skeptics in anthropology and evolutionary biology (Anderson, 1974; Hrdy, 1981, 1990, 1999, 2006; Parish, 1994; Parish, de Waal, & Haig, 2000) about female enthusiasm for mating and multiple mating. In addition, the implications of female multiple mating are often overlooked and the imperative is ignored to study a range of potential fitness benefits, not just costs, for females who mate multiply. Furthermore, it is clear from hindsight that something has inhibited the advance and recognition of alternative hypotheses about the origins and extent of sex differences and similarities in behavior, such as in choosy or indiscriminate reproductive decision making. As I explain in subsequent sections, many investigators assume choosy or indiscriminate reproduction is the key sex difference fueling sex-differentiated opportunities for fitness and selection.

To some readers, many of my points will seem like old news (Gowaty, 2003; Schiebinger, 1991, 1993, 2001). Also, I most certainly am not arguing from the perspective that there is anything novel about bias in science or that scientists are immune from bias. Far from it: I consider the ongoing, ever-developing—albeit slowly—practice of science to be bias busting (Hilborn & Mangel, 1997; Platt, 1964; Popper, 1945). Some will consider my arguments narrow, but other practicing scientists will find my

I thank the editors for inviting me to contribute to this handbook and asking me to write this chapter. Adriana Maldonado Chaparro, Justin Garcia, Root Gorelick, Sarah Blaffer Hrdy, Stephen P. Hubbell, Zuleyma Tang Martinez, Pablo Marquet, James Moore, Yi Zhai (a recent postgraduate from Peking University), Matthew Gillum (an undergraduate at the University of California, Los Angeles), Cheryl Travis, Jackie White, and Alexandra Rutherford each read a previous draft and made useful suggestions, for which I am extremely grateful.

<http://dx.doi.org/10.1037/XXXXXXX-008>

APA Handbook of the Psychology of Women: Vol. 1. History, Theory, and Battlegrounds, C. B. Travis and J. W. White (Editors-in-Chief)
Copyright © 2018 by the American Psychological Association. All rights reserved.

concerns thoroughly contemporary, not too foreign from recent scientific discussions of critical practice in social psychology and ecology (Aarts et al., 2015; Brandt et al., 2014; Chin, 2014; Hesselmann & Moors, 2015; Klein et al., 2014; López-López, 2014; Makel, 2014; Marquet et al., 2014; Miguel et al., 2014; Murayama, Pekrun, & Fiedler, 2014; Nakagawa & Parker, 2015; Wicherts & Bakker, 2014). Moreover, I hasten to emphasize that my goal here is not a thorough critique of evolutionary-based sex differences research. I am attempting only to make a singular point providing evidence that many evolutionary scientists saw not what was on the page but rather what they believed. What is novel in my studies of Bateman's work (Gowaty, Kim, & Anderson, 2012, 2013; Snyder & Gowaty, 2007), and what I emphasize here, is the evidence demonstrating that the errors of inference and method in a foundational study were easy to spot and elementary, demanding only a familiarity with simple manipulations of fractions and modest appreciation of the syllogisms (Dennett, 1995; Gowaty, 2014) of Darwinian selection (Darwin, 1871).

In response to critics who say I do not go far enough, I emphasize that my appeal is not to those already aware of the power of gender biases to affect the practice of science. Rather, I hope to capture the interest of and inspire skepticism among those who have retained confidence in the canon of the evolutionary origins of sex differences. The evidence implies that true believers suffer in the glare of their beliefs, which obscures vision just as profoundly as sun glare hides the details of backlit faces.

To illustrate how gender schemas and stereotypical gendered cognitive frames may impinge on the evolutionary study of sex differences and similarities, I (a) briefly review the elements (the flow of hypotheses) of the classical theory of the evolution of sex differences and the nature of common empirical tests; (b) describe the discovery of flaws in Bateman's (1948) linchpin study of sex differences in fitness (i.e., in mating and reproductive success); and (c) link the cognitive frames of essentialism and gender schemas to belief in the

still-influential texts of the classical theory. At the same time, I consider the types of logic on which the classical claims rest in an effort to identify why it has taken so long for data inconsistent with the canonical predictions to be acknowledged as inconsistent or for testable alternative hypotheses for the evolutionary and developmental origins of sex differences and sex similarities to gain traction and be tested. Last, I speculate on whether evolutionary studies of sex differences are peculiarly subject to bias or whether the collective failure to understand the elementary but substantial errors in Bateman's study is really just normal science (Kuhn, 1962).

CORNERSTONES OF THE CANONICAL THEORY OF SEX DIFFERENCES

Adaptive scenarios for the evolution of sex differences are often attempts to explain why males are aggressive, sexually profligate, and indiscriminate in mating behavior, whereas females are dependent, coy, and choosy. These conceptual notions of sex differences are where many biologists, evolutionary psychologists, behavioral ecologists, and sociobiologists begin to cast further hypotheses about sex differences, whether or not these ideas capture the range of variation in the sexes (Hyde, 2005; Martin, 1991). To my eyes, most evolutionary hypotheses about sex differences flow from arguments with the form "sex differences predict sex differences."

The usual justificatory thread begins with Darwin's (1871) hypothesis of sexual selection. It then jumps to the ideas of Parker, Baker, and Smith (1972) and Trivers (1972) about the selective forces of sex-differential energetic costs of reproduction, with both arguments predicting that in most species, females are the choosy sex and males the indiscriminate sex. Trivers and Parker et al. tied their ideas to Bateman's (1948) claims that greater among-male variance¹ in number of mates (V_{NM}) produced higher variances in reproductive success (V_{RS}) in males than in females, thereby "proving" that evolutionary potential is greater in males than in females. The threads of logic and evidence from these sources

¹Fitness variances, such as in number of mates (V_{NM}) or in reproductive success (V_{RS}), are ideas important for understanding how Darwinian selection operates. Variances in components of fitness refer to the variation among individuals expressed as a population-level metric. In the absence of other information, higher variances in components of fitness indicate a greater potential for selection within a given population.

have produced an evolutionary guide, a scenario for why the sexes are as they are (see Figure 8.1).

Taking the scenario depicted in Figure 8.1 apart to understand its power requires evaluation of the structure of the hypotheses of Darwin (1871), Trivers (1972), and Parker et al. (1972) and of the fidelity of the data and conclusions from Bateman's (1948) study. Thus, in the next sections, I (a) describe sexual selection hypotheses, emphasizing their deductive form; (b) discuss the qualitative sources of the predictions about behavior from anisogamy (sex-differentiated morphological variation in gametes) and parental investment theories; and (c) describe how Bateman's study seemed to fulfill those predictions when in fact it was riddled with errors, methodological flaws invalidating the results, and methods later misrepresented by those authors who made false claims about the methods of the study.

Darwin's Sexual Selection

In 1859, Darwin said, "Sexual selection . . . depends, not on a struggle for existence, but on a struggle between the males for possession of the females; the result is not death to the unsuccessful competitor,

but few or no offspring" (p. 88). By 1871, he had a more inclusive—and to my eyes, more interesting—definition: "Sexual selection . . . depends on the advantage which certain individuals have over other individuals of the same sex and species, in exclusive relation to reproduction" (p. 256). There is a world of difference between the 1859 and 1871 definitions of sexual selection, which is hardly surprising as Darwin's 1871 book is about the many ways that he could see sexual selection operating as well as the ways he put aside for later exploration. The 1871 argument is broader than the 1859 one, because it includes the idea that females, not just males, can experience within-sex selection in relation to reproduction (Gowaty, 2011, 2014). Reproduction is more than "possession of the females" and more than the number of mates that a male has. Even though the usual interpretation of male–male sexual selection is directed toward understanding the evolution of dramatic, showy male traits (see Figure 8.1), sexual selection can affect other heritable traits too, as Darwin (1871) emphasized when he said "the manner in which the individuals of either sex or of both sexes are affected through

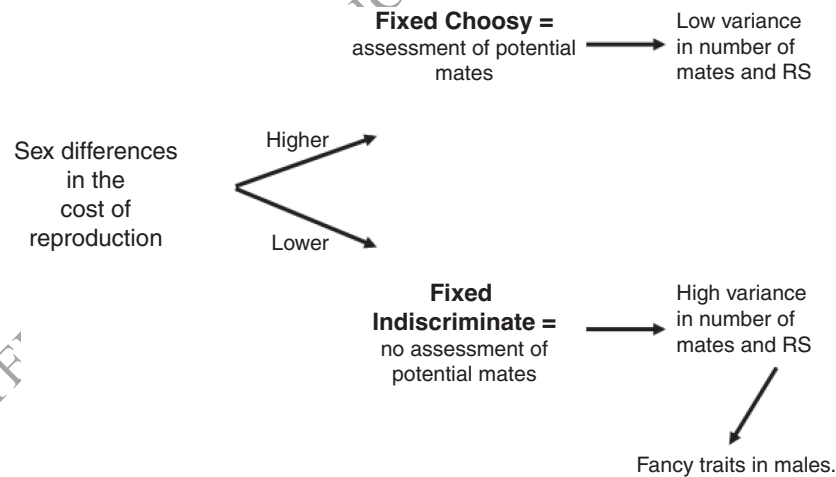


FIGURE 8.1. The canonical model for the evolution of sex differences in behavior and fitness. The logic of these relationships is discussed in detail in the text. Sex differences in the cost of reproduction (Williams, 1966) supposedly operate to produce sex differences in Darwin's (1871) mechanisms of sexual selection, with females being choosy and males indiscriminate about mates. This logic flow implies what Bateman (1948) claimed to prove: that the variances in reproductive success (RS) within a population are less for females and that the higher variances in males are due to choosy females preferring to mate with some favored males more than others, usually because of certain male traits likely to provoke female attention (e.g., fancy tails in birds).

sexual selection cannot fail to be complex in the highest degree” (p. 296).

What modern evolutionary scholars realized, perhaps more slowly than Darwin intended, is that the rules—arithmetic-like assumptions²—required by any selection hypothesis, whether natural or sexual selection, can result in many iterations of selection hypotheses (Dennett, 1995). The required laws or assumptions are about the trait variation in the units of selection, the circumstances that affect the components of fitness so that differential reproduction or survival of individuals occurs. When they are variable and heritable, any selection hypothesis becomes one of evolution via selection (Gowaty, 2014). To be excessively clear about sexual selection,

1. it is always a process of sorting among individuals of a sex because of the trait variations of the individuals of that sex within a population;
2. it must happen within a population—not between populations—because the mechanisms by which the selective sorting occurs are through opposite-sex mate choice and within-sex rivalry (either behavioral or physiological), which means that potential mates and rivals must interact; and
3. it may have selective force via the correlated effects of mate choice and same-sex rivalry on many possible components of fitness, including survival, number of offspring, or the viability of offspring.

Hypotheses of selection—whether natural or sexual selection—are hypothetico-deductive syllogisms (Gowaty, 2014) that have admirable clarity, making the assumptions and predictions readily challenged with empirical tests, which is a hallmark of efficient science (Marquet et al., 2014). Thank you, Mr. Darwin!

Clearly, in any population, sexual selection may act among females and separately among males. Yet selection may be different in terms of the trait variation it affects, the mechanisms by which mate choice or same-sex rivalries sort among the individuals subject to selection, and the components of reproductive success that the mechanisms affect. To imagine that sexual selection proceeds in only one

way, as shown in Figure 8.1, is a common, essentialist conceptual yoke, a cognitive straitjacket that is not only short sighted but also conceptually at odds with Darwin’s expectations about the complexities of sexual selection, not to mention at odds with experimental tests of mate preferences (as described in the following paragraphs). The assumption that sexual selection must work identically in every instance in both sexes puts all the emphasis on males and thereby (a) obscures the existence of heritable trait variation among females, (b) hides from imagination and study the mechanisms of selection that act on heritable trait variation in females, and (c) conceals information about the components of fitness that female–female reproductive competition affects. In other words, the interpretation of sexual selection promoted by the cost-of-reproduction arguments is like a burka, hiding females from view theoretically and empirically because the canonical view says that all, or most, of the action of sexual selection is on males.

Cost-of-Reproduction Arguments are Essentialist Assertions

In *Adaptation and Natural Selection*, Williams (1966) anticipated sociobiology’s fervor for the bottom line of posited selection pressures from the costs of reproduction acting to produce coy and careful females and profligate, competitive males:

It is commonly observed that males show a greater readiness for reproduction than females. This is understandable as a consequence of the greater physiological sacrifice made by females for the production of each surviving offspring. A male mammal’s essential role may end with copulation, which involves a negligible expenditure of energy and materials on his part, and only a momentary lapse of attention from matters of direct concern to his safety and well-being. The situation is markedly different for the female, for which copulation may mean a commitment to a prolonged burden, in

²Throughout, I use the term *assumption* to refer to parts of a hypothesis producing deductions, as in hypothetico-deductive hypotheses.

both the mechanical and physiological sense, and its many attendant stresses and dangers. Consequently the male, having little to lose in his primary reproductive role, shows an aggressive and immediate willingness to mate with as many females as may be available. If he undertakes his reproductive role and fails, he has lost very little. If he succeeds, he can be just as successful for a very minor effort as a female could be only after a major somatic sacrifice. Failure for a female mammal may mean weeks or months of wasted time... Natural selection should regulate her reproductive behavior in such a way that she will assume the burdens of reproduction only when the probability of success is at some peak value that is not likely to be exceeded.

The traditional coyness of the female is thus easily attributed to adaptive mechanisms by which she can discriminate the ideal moment and circumstances for assuming the burdens of motherhood. One of the most important circumstances is the inseminating male. It is to the female's advantage to be able to pick the most fit male available for fathering her brood. (pp. 182–184)

The questions about sex differences that evolutionary biologists might ask from a female perspective are emblematic of the obfuscatory power of essentialist belief. Answers to many important questions are still opaque, hardly examined, and seldom asked. Obviously, there is no doubt that mammalian females gestate and lactate and males do not, but does that mean that females suffer a greater physiological sacrifice? Or does it mean something altogether different, such as that mammalian mothers enjoy greater control than fathers over reproductive outcomes and over subsequent offspring phenotypes and life-history traits? Selection must have acted particularly strongly among females to produce their extraordinary influence over offspring, and yet, what is really known about the selective dynamics yielding such effects? What sorts

of contemporary social environments for females are associated with increased and decreased variances in their components of reproductive success or their survival? Are females really traditionally coy and males aggressively willing to mate? What would it mean if males are choosy too, and if females behave indiscriminately? The answers to such questions (Drickamer, Gowaty, & Holmes, 2000; Drickamer, Gowaty, & Wagner, 2003; Gowaty et al., 2007; Gowaty, Drickamer, & Schmid-Holmes, 2003; Gowaty, Kim, Rawlings, & Anderson, 2010; Gowaty, Steinichen, & Anderson, 2002, 2003) are often ignored even when they are answered, becoming grist for a historian's mill. When tests of mate choice seek only evidence of female choice while ignoring the possibility of male choice, the tests are obviously incapable of producing evidence about sex differences or sex similarities.

Perhaps an even more important question than the alternative ways to characterize differences in reproductive roles is how to test Williams's (1966) version of the cost-of-reproduction arguments. From an evolutionary perspective, any extraordinary costs of reproduction for females would have exerted strong selection on them that likely would reduce or minimize ongoing costs, lowering variances in some components of female fitness. But what about any remaining variances in the many components of female fitness that are not usually tested? Variances in certain components of fitness, such as the ratio of adult offspring to the number of eggs laid or offspring born, might be important to females (Altmann, 1997). Perhaps that is the bit that ought to be the focus of attention: What were the selective pressures that produced these seemingly fixed female traits? Are they at work currently? What variations in females have researchers failed to see given that the "Williams–Bateman bias" so often suggests that there is little or no important variation in female fitness that would fuel selection on females' traits?

Similarly curious are the many studies that show clear inconsistencies with the Williams–Bateman predictions, such as a recent conceptual review of empirical mate choice studies (Ah-King & Gowaty, 2016). If the argument is about the effects of ancient selection pressures producing fixed trait variation, such as "all females are choosy" and "all males

are indiscriminate,” a single observation (e.g., in a beetle species) of indiscriminate females or of choosy males would call the theory into question (Johnson, 1982; Johnson & Hubbell, 1984), just as an observation of a black swan challenges the statement that “all swans are white.”

Along these lines, consider that the vast majority of mate preference studies test only for female mate preferences, leaving male mate preferences unexamined, by and large. A defense of this practice is that most female mate preference studies pursue explanations for the evolutionary origins of exotic (i.e., bizarre, elaborate, or fancy) traits in males (think of peacock tails). There are thus two problems: one is the failure to consider the evolution of traits—fancy or not—in females. The other is the question of the fitness consequences for choosers of potential mates, independent of the expectations investigators have about the inducing traits—the mediators of preference—in the chosen.

After Williams’s (1966) informal but influential arguments about the selective consequences of the cost of reproduction, additional interest came from Parker et al.’s (1972) mathematical anisogamy theory and Trivers’s (1972) qualitative parental investment theory. Anisogamy and parental investment theories made the same predictions as Williams: Males are active, indiscriminate, and competitive over mating access to females, and females are passive, always choosy about with whom they mate, and competitive, not over access to mates but over access to essential resources.

Anisogamy theory came from an analytical—meaning algebraically expressed—model of how encounters with potential mates in a population of normally distributed gametes could result in disruptive selection to produce two gamete sizes (Parker et al., 1972; Trivers, 1972; Williams, 1966). It is a theory of selection pressures on the morphology of sex cells. The math assumes that individuals must meet to mate and that zygotes resulting from gamete fusion must have resources to develop, which the authors assumed come from big, resource-accruing

gametes (the eggs). The dynamics of mate encounter, under differential costs and benefits of gamete size and relative vagility, favored the evolution of two sizes of gametes. The model of disruptive selection on gamete sizes is a first-principles deductive idea assuming that encounters with potential mates matter, as does the ability to sequester resources necessary for zygote development and survival. Sometimes in nature these two roles are simultaneously met by a single gamete size. Nevertheless, the anisogamy model described a reasonable pathway to the evolution of two sizes of gametes, which seems to be the most common pattern in nature.

From there, the authors made additional, non-mathematical caveats associating the evolution of morphological size variation with choosy and indiscriminate mating behavior. The add-on discussion was similar to Williams’s (1966) logic: There are sex differences in the cost of reproduction so that selection favors females as the choosy sex and males as the indiscriminate sex. Predictions derived from anisogamy theory about behavioral sex differences were not mathematically deduced but were ad hoc additions, intuitively expected correlates of morphological variation in gametes. Thus, the facts of big, sessile, resource-rich eggs and small, mobile sperm do not necessarily imply either that the bearers of sperm compete and mate indiscriminately or that the bearers of the larger gametes are choosy and relatively noncompetitive over mating opportunities. Two generations of scholars have taken as law the intuitively derived association of big gametes with choosy behavior and small gametes with indiscriminate behavior despite data challenging the theory’s predictions and the existence of alternative theoretical arguments (Dewsbury, 1982). Is it because common gender schemas inhibit imagination of alternative hypotheses, thereby hijacking the scientific enterprise and promoting seeing what one believes?

Besides the problem of believing is seeing, a first-principles deductive and mathematically proven theorem³ predicts that adaptive, fitness-enhancing

³I am using the term *mathematically proven theorem* in the sense that mathematical theorists use the term. If the mathematically expressed assumptions of the theorem are met in nature, the predictions must follow from the theorem. Tests of a mathematically proven theorem therefore focus on whether the assumptions are met in nature or in a particular system. An example of such a theorem is the Hardy–Weinberg theorem, which proved mathematically that the genotype frequencies in a population remain constant from generation to generation unless mutation, selection, drift, or other evolutionary forces are at work. Mathematically proven theorems force the attention of empiricists onto the fit of the assumptions in a particular empirical setting.

behavioral flexibility (being choosy and indiscriminate) of individuals—regardless of their sex—can evolve without anisogamy or other claims about sex differences in morphology (Gowaty & Hubbell, 2005, 2009, 2010, 2013a, 2013b; Hubbell & Johnson, 1987). Moreover, mathematical deductive theorems have proven that selection will be against individuals fixed for choosy or fixed for indiscriminate behavior whenever the assumptions of the theory are met (Gowaty & Hubbell, 2005). Thus, again, it is not necessarily true that the sexes are what they are because of the size of the gametes they carry (Parker, 1979).

Moreover, experimental tests of anisogamy's predictions of fixed sex differences exist and allow rejection of its predictions of sex-differentiated, fixed reproductive decision making (Altmann, 1997; Anderson, Kim, & Gowaty, 2007; Drickamer et al., 2003; Gowaty et al., 2002, 2007; Gowaty, Drickamer, & Schmid-Holmes, 2003; Gowaty, Steinichen, & Anderson, 2003; Johnson, 1982; Johnson & Hubbell, 1984; Moore, Gowaty, Wallin, & Moore, 2001; Ryan & Altmann, 2001). *Drosophila pseudoobscura*, the fruitfly species with the tiniest known sperm (at least among flies) and “regular-sized” ova, is a model species for the study of anisogamy theory's predictions: *D. pseudoobscura* has the biggest known asymmetry in gamete size of any fruitfly species. The empirical tests (Anderson et al., 2007) show that both females and males are choosy, and both sexes are indiscriminate. What may be more interesting is that in both sexes, choosy subjects experimentally bred with individuals they preferred had offspring of higher viability than when they were paired with individuals they did not prefer. The bottom line is that both sexes were sometimes choosy and other times indiscriminate, indicating flexibility, and regardless of their sex, their mate preferences predicted the viability (health) of the offspring they produced when they mated with potential mates they preferred versus those they did not prefer.

Results such as the one just described imply another epistemological question: How many times must researchers observe inconsistency with the anisogamy predictions before they reject the hypothesis? Must investigators perform experimental tests

in all known extant species? Or is a single rejection, such as the single observation of a black swan, enough to put the anisogamy hypothesis off the table as the only, or even the favored, explanation for sex differences in behavior? This is a question that goes to the heart of sex differences research: Is it efficient science? Claiming consistency via confirmatory results only (Janicke, Häderer, Lajeunesse, & Anthes, 2016; Jones, 2003; Jones, Arguello, & Arnold, 2002) is hardly good enough. Indeed, claiming consistency via confirmatory results is a violation of the hypothetico-deductive method and, as Popper (1945, 1959) warned, risks turning a scientific issue into a metaphysical one. The quest for consistency with favored hypotheses is a scientific pitfall and certainly an inefficient way to advance scientific knowledge (Marquet et al., 2014) in that it propels faith, not reason.

Trivers's (1972) qualitative parental investment hypothesis stated that patterns of parental investment were the causes of sex differences in behavior. His argument was an elaboration of Williams's (1966) story justifying choosy female mammals because of the relative costs to them of reproduction. The argument goes that the energetic cost of any copulation for females compared with males would potentially be greater, because females must also gestate, lactate, and sometimes care for offspring for extended periods. All of these duties of female energetic investment were said to create a series of selection pressures that favored choosy females. Because males generally do not experience the same obligate costs as females, males would not experience selection favoring choosy behavior and so would be indiscriminate. The parental investment hypothesis went further, however, to predict that when males invested more energy than females, the males would be the choosy sex and the females the indiscriminate sex, and so began **ad hoc** modifications to first-canon texts, including derivative theories predicting the “direction” of sexual selection, as though when one sex experiences strong sexual selection the other sex cannot. In fact, the direction-of-sexual-selection language is a dead giveaway that users misunderstand the basics of sexual selection, or perhaps they make unstated restrictive modifications to sexual selection syllogisms, because theoretically whether sexual selection acts strongly on

females says nothing about whether sexual selection acts strongly on males and vice versa. This point is obvious when one looks carefully at Darwin's ideas and uses his syllogisms to construct sexual selection hypotheses (Gowaty, 2014).

The empirical question related to Trivers's (1972) parental investment predictions is whether experimental tests can reject the hypothesis. Tests of female and male mate choice behavior in a typical polygynous mammal species with greater female than male investment—feral house mice (*Mus musculus*)—showed that both females and males have mate preferences. These tests also demonstrated that both sexes gain the fitness benefit of enhanced offspring health when they reproduce with individuals they prefer compared with individuals they do not. The results of experiments in feral house mice rejected the parental investment hypothesis because parental investment predictions were not met (Drickamer et al., 2000, 2003; Gowaty, Drickamer, & Schmid-Holmes, 2003). Had the experiment only focused on female house mice, its results would have been consistent with parental investment predictions and would have contributed to “consistency science.” However, because the investigators tested both sexes simultaneously, controlling ages of subjects, their environments, the form of the preferences tests, and so forth, the tests were able to produce data that rejected the parental investment hypothesis.

Even more fascinating are the large number of empirical articles showing that in species with seemingly greater investment in reproduction by females than males, females can and do switch between choosy and indiscriminate behavior in the presence of predators; females also switch when they are parasitized, sick, or experience differences in their encounter rates with potential mates. Ah-King and Gowaty's (2016) review indicated that the many articles showing variation in females' reproductive decision-making produced data inconsistent with the parental investment hypothesis. However, authors of the reviewed articles seldom made that connection in their reports, as though belief in the parental investment hypothesis stands despite their own data.

Both Trivers (1972) and Parker et al. (1972) cited Bateman (1948), who remains the key player

in the flow of arguments for the evolutionary origins of sex differences (see Figure 8.1). What Bateman's study did was connect essentialist sex differences in behavior to the essentialist ideas about sex differential costs of reproduction and then to observable sex differences in variances in reproductive success (V_{RS}). For some, such a linkage fastened the knot of essentialist answers to why the sexes are as they are.

In the next section, I discuss Bateman's (1948) study, designed to demonstrate experimentally the differences in fitness accrual of males and females. As my collaborators and I have shown, Bateman's published study was complex and difficult to understand partly because it was not organized in the standardized ways of modern studies that most of us are used to reading. The study had important errors, including analytical missteps, but also fatal flaws in the methodology that rendered more than a third of his 63 populations incapable of fairly answering his questions. Many errors were so straightforwardly obvious that any thoughtful reader should have seen them on first reading. Why they did not is a question that dogs me. The fact that readers saw what they believed emerges as a compelling explanation.

Bateman's Hypotheses Became Conclusions That Became “Bateman's Principles”


In the decade before Bateman's (1948) experiment, Fisher's (1930) fundamental theorem of natural selection focused attention on variances in fitness as central to understanding the potential for evolutionary response: Without trait variation and without fitness variation, there is no opportunity for evolution by selection. Fisher also originated the statistical models of analysis of variance, which was the main statistical tool that Bateman used to analyze his data. Although Bateman did not cite Fisher's work, it is not hard to imagine that Fisher's ideas about variance in fitness inspired Bateman's hypotheses and the remarkably ambitious experiment he did to test them.

Bateman began his experiment sometime in the mid-1940s and he published it in the first issue of the new journal *Heredity*. In keeping with Darwin's ideas about the evolution of exaggerated male traits, he hypothesized that female choice

and male–male behavioral rivalry over access to females were the real crux of mating matters among males (whether ornamented or not). His subjects were fruitflies, *Drosophila melanogaster*, in small, equal sex-ratio populations. Bateman's hypothesis of female choice—not male choice—as a driver of sexual selection was a hypothesis that contrasted with previous experiments using fruitflies. In male-choice experimental designs, investigators placed a male in a vial with two females, in those days usually of different species. Investigators then noted which female the male mated first and took this mating behavior as an indication of a cognitive process called *male preference*. Bateman's alternative hypothesis was that if sexual selection on males was happening the following predictions would be met: (a) Female mate choice and male–male rivalry over mating would produce greater variances in numbers of mates (V_{NM}) for males than females;⁴ (b) the greater V_{NM} for males than females would in turn produce greater variances in reproductive success (V_{RS}) for males than females; and (c) there would be a stronger dependence of V_{RS} on V_{NM} for males than females, indicating the cause of sexual selection acting among males. The repercussions of Bateman's hypotheses that became his conclusions are legend.



Nearly 50 years after his initial publication, Bateman's (1948) hypotheses and conclusions became Bateman's principles (Arnold, 1994). In other words, the hypotheses and conclusions became facts to count on. Bateman's principles propped up belief in the idea that there are no fitness benefits for females of mating with more than one male and that females are indeed coy and males indeed profligate and competitive. Thus, Bateman's principles provided an evolutionary explanation for the double standard in sexual behavior of women and men, something that prescient biologists, such as Hrdy (1981), noted nearly 40 years ago and which they also challenged with observations of humans' closest relatives. In terms of most studies that were done, it was as though female subjects were just not interesting in relation to the evolutionary project. Notable exceptions were in disciplines with an unusual number of female professionals

(e.g., anthropology and psychology), who did focus on females. As historians have shown, studies of nonprimate species that put females in the spotlight were often criticized and more often than not simply ignored. Only in the past few years has one seen articles that begin with language such as “Despite Bateman's principles, females do mate with multiple males and do gain fitness rewards.” It is as though the idea has gained support only recently, despite the fact that modern genetic parentage studies of wild living flies proving that females mate with multiple males go back a long time (Anderson, 1974; Anderson et al., 2007; Turner & Anderson, 1983).

Given the association of Bateman's (1948) conclusions with contemporary double-standard expectations for humans, I wonder why so few asked the crucial questions that Bateman's obvious errors later stimulated. For example, Snyder and Gowaty (2007) showed that many people who promoted Bateman's hypotheses, results, conclusions, and principles obviously did not know what Bateman had done. For example, they did not know how he inferred parentage; they did not attend to the structure of his statistical tests, or the justifications for the graphs he used to illustrate his conclusions. How did the graphs (see Figure 8.2) make it into textbooks? What one sees in the history of Bateman citations is not an immediate critical engagement with his methods or his results; rather, it took decades before the obvious errors were noticed and alternative hypotheses in the form of mathematical theorems were proposed that were also able to explain Bateman's data (Hubbell & Johnson, 1987; Sutherland, 1985). Observations of female multiple mating began in the late 1970s, accelerated in the 1980s, and kept accumulating, results that  facie challenged Bateman's principles. Yet Bateman's principles continued to stand: Something akin to “true belief” must have been afoot, rather than the kind of engaged criticism that is usually a hallmark of science. Alternatively, perhaps many believers just did not read Bateman. Even today the traces of skepticism in citations to Bateman (1948) are as rare as hens' teeth. Why? Could this belief, contrary to the facts, be sustained for more than half a century

⁴Note that variances are metrics of population-level variation. V_{NM} and V_{RS} indicate variation in measures of average NM or average RS. So, for example “higher V_{NM} in females than males” means that females in that population are less like each other in their numbers of mates than males are like each other.

TABLE 4
Sample result of one mating. From series 6 $B \times A$, sixth, bottle
(CyL, Cy, Mc) \times (Pm, H, Sb)

 	CyL	Cy	Mc	+	Total	No. of mates
Sb • •	• 13	0	0	16	29	1
Pm • •	10	12	15	68	105	3
H • •	7	29	0	41	77	2
+ • •	60	38	40	110	248	•••
Total • •	90	79	55	235	459	•••
No. of mates	3	2	1	•••	•••	•••

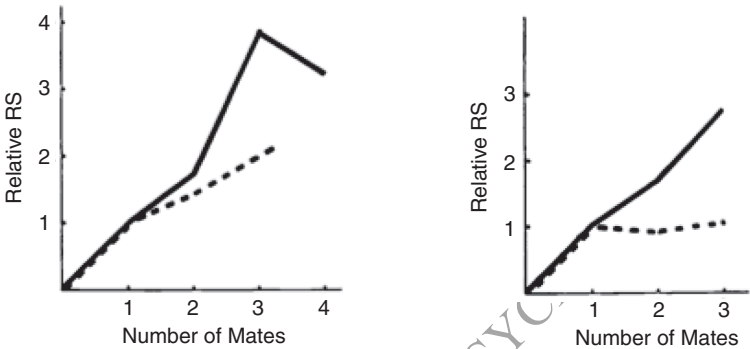


FIGURE 8.2. Key illustrations from Bateman (1948). Panel a shows Bateman's (1948) original Table 4. Panels b and c show Bateman's original Figure 1a and Figure 1b, respectively. See text for a detailed discussion of these items. RS = reproductive success. From "Intra-Sexual Selection in *Drosophila*," by A. Bateman, 1948, *Heredity*, 2, pp. 357, 362. Copyright 1948 by Nature Publishing Group. Reprinted with permission from Macmillan Publishers Ltd.

merely on the basis of careless science? I suspect that historians will conclude, as I have, that scientists saw what they believed.

In the next paragraphs, I describe Bateman's (1948) errors (i.e., failure to fully test Mendel's expectation of types of offspring given parental types; lack of control experiments; inappropriate degrees of freedom in his analyses of variance; haphazard graphical displays). And I ask again why they were not noted in the 1970s or earlier, because the errors were easy to spot, requiring no advanced skill sets or numeracy beyond an understanding of fractions.

To study sex differences in V_{NM} and V_{RS} , one needs to know who mates, who mates with whom, and how many offspring each subject has. Bateman's (1948) sole source of data was a genetic parentage assignment study, which involved adult subjects (the possible parents) who had unique genes coding

for unique observable phenotypes; thus, when offspring inherited the nametag gene of a parent, they would express it phenotypically and thereby identify the parent. All of Bateman's inferences depended on the phenotypes of offspring produced in each population. Given the limitations of molecular genetics for parentage studies at the time (they had not yet been invented), Bateman's experiment used a very clever method.

Bateman (1948) cultured his subject adults so that each was heterozygote dominant at a unique locus. The culturing of the subject flies—the adults whose number of mates and reproductive success he set out to measure—was by itself an ambitious undertaking. For the experiments, he needed individuals with different phenotypically expressed genetic mutations—nametags—that could be culled from cultures in such a way as to guarantee that his subjects were each heterozygote dominants at their

unique nametag locus (i.e., that a single identifying phenotype was coded by a single mutant allele at a different locus for each subject adult). Thus, even before he put subjects to breed in populations, he needed several months of careful fly culturing to create his experimental populations of interacting females and males. For populations with three females and three males, six different nametags identified the adults. For populations with five males and five females, 10 different nametags identified the adults. In sexual species, each offspring has both a mother and a father. Parental gametes each contribute a single set of chromosomes to each offspring. Thus, when all possible parents are heterozygote dominants at different loci, Mendel's law reduces to its simplest form (see Figure 8.2): 25% of offspring will inherit a nametag from each parent, *DD offspring*; 25%, a nametag from their mother and a wildtype allele from their father, *D+ offspring*; 25% a wildtype allele from their mother and a nametag allele from their father, *+D offspring*; and 25% wildtype alleles from both parents, *++ offspring*.

For Mendel's law to hold, a critical stipulation is that, when expressed in offspring, the nametag genes are not deleterious whether inherited in combination or singly. When this assumption is met, one can estimate with minimal error the number of mates per subject and how many offspring each subject produced. However, one must first test the validity of this crucial risk-of-lethality-in-offspring assumption, which can be done in at least two ways. A control experiment using a series of monogamous pairings (one female and one male in a vial) would have tipped Bateman (1948) off, as it did in our studies (Gowaty et al., 2012), that when offspring inherited nametag mutations from both parents simultaneously, they often died before they emerged as adults. Dead babies obviously were invisible, making it impossible to know who their parents were, and these missing offspring necessarily indicate that some subjects who mated and reproduced (producing progeny that did not survive to an age when nametags expressed) would be overlooked (Anderson, 1969), rendering the overall results about number of mates potentially and seriously biased. Another easier way of validating the assumption, done on each of Bateman's populations, is a

robust test against Mendel's expectations, which were all the rage in genetic studies during the first half of the 20th century (see Figure 8.2). That is what Gowaty et al. (2012) did in their strict replication, using the nametag mutations of the populations Bateman graphed (reproduced as Figure 8.2c), and it is what Hoquet and Gowaty (2016) did when analyzing the populations Bateman recorded in his lab notebooks.

After Bateman (1948) placed heterozygote-dominant subjects in populations, he allowed them to interact and mate for 3 or sometimes 4 days, after which he removed the subjects holding the vials until offspring emerged from pupae as adults. Then, he sorted offspring into types, as shown in the table reproduced as Figure 8.2a, noting the number of mates for each subject and the number of offspring.

The mismatch in Bateman (1948; see Figure 8.2a) between the observed offspring types of his only fully reported population and Mendel's rules are obvious and should immediately have troubled readers, as it did Snyder and Gowaty (2007). The reason is that the deficit in *DD* offspring necessarily implies that the estimate of the number of mates for each subject was inaccurate to some unknown degree, so that to some unknown degree the estimates of V_{NM} would be wrong, both overestimating the number of individuals with no mates and underestimating those individuals with more than one mate (see Figure 1 in the supporting information for Gowaty et al., 2012).

The Mendelian expectations may sound complex, but they are not (see Figure 8.2): It is a simple problem at a high school level of inference as to what the frequency of alleles will be in offspring, given parental genotypes. Had Mendel's rules been met, Bateman (1948) could have reliably estimated who mated with whom as well as the relative reproductive success of subjects. If the experiment had worked, it would have been elegant, but it did not work, and it was not elegant. Gowaty et al.'s (2012) repetition proved that there was a fatal flaw in methodology for the populations Bateman used to illustrate his findings (as reproduced in Figure 8.2c). When Hoquet and Gowaty (2016) analyzed all the populations in Bateman's original lab notes, a third of the populations failed to meet Mendel's

rules. Bateman should have removed these populations from further consideration because failure to match Mendel's expectations proved that the data were incomplete relative to Bateman's questions and therefore incapable of providing reliable answers. Why did Bateman not do appropriate tests of Mendel's expectations? A third grader can easily do the calculations by hand. Anyone with minimal exposure to genetic inference should have noticed.

Snyder and Gowaty (2007) were also the first to publish that Bateman (1948) had pseudo-replicated his analysis of variance: When he tested for sex differences among male and female subjects in variances in number of mates and variances in reproductive success, he inflated the degrees of freedom for individuals by pooling data across populations, yielding apparent statistically significant results when in fact there were none. How did Fisher, the originator of analysis of variance, let Bateman's analysis of variance pass peer review? He was the editor of *Heredity* at that time, but instead of critiquing the inflated degrees of freedom, Fisher wrote to Bateman that the results of the experiment did not surprise him. Perhaps even Fisher was vulnerable to seeing what he believed. Why did the army of fierce graduate students who read Bateman from the 1970s to today also not find the error? The "lie in believe" may be a general problem.

So, how did Bateman's (1948) article catch on and spread? How could it have been read so uncritically for 60-some years? For example, Bateman never watched the behavior of his subjects: No surprise, as it would have been an impossible workload for him to watch the individuals in his populations—despite the fact that each subject had a distinctive and phenotypically observable genetic mutation—because each of the populations lasted for 3 or 4 days. Even with today's technology, watching six or 10 flies interact would be challenging. Dewsbury's (2005) reading of Bateman, and others who cited Bateman so prominently, revealed that some influential authors later claimed that Bateman had reported that he observed females choosing among males. However, there is no evidence in Bateman that he or anyone else might have watched behavior. Nor is there such evidence in Bateman's recently discovered handwritten lab notes (personal

observation, T. Hoquet and P. A. Gowaty, 2013). Dewsbury's digging showed that the later authors did not read Bateman, failed to read it carefully, or just saw what they believed. Whether these authors were truly self-deceived, seeing what they believed or not, they compounded Bateman's error, misleading generations of gullible readers. The enthusiasm felt in the heady days of the 1970s now seems akin to what happens to gullible consumers encountering a con artist on a carnival midway: Readers saw what they believed, suspended critical judgment, perhaps bowed to those more powerful, and took as true that for which there was no evidence.

Other problems, as clear as the nose on one's face, include the curious case of Bateman's (1948) graphs (reproduced in Figure 8.2). Bateman haphazardly pooled his data across populations, then said that the data from the populations he pooled into the second graph "looked different." In plain language, he cherry picked and emphasized the second graph that showed an apparent dramatic sex difference (via slope) in the relationship between number of mates and reproductive success. Anyone who had read Bateman's article would have seen this suspect justification for two graphs: "Series 5 and 6 differed somewhat from the rest." One of Bateman's graphs (see Figure 8.2c) contains pooled data from 24 or 25 populations. The other graph (see Figure 8.2b) pooled more than 40 populations and thus had considerably more data, including the populations that had 10 adult subjects, not just six. Pooling data over populations is inappropriate for analyses of sexual selection because sexual selection is a process that occurs within populations, not between them. Putting aside the error of pooling across populations, it is worth emphasizing that the graph with the closest similarity to the gender schema of the double standard (see Figure 8.2c) was the one that got so much press. Again, I take the failure of others to note the errors as evidence that they were seeing what they believed, without skeptical engagement with the ideas, the data, the results, or the implications of the study.

Graphs such as those in Figure 8.2, which are digital copies of those from the original article, have become known as Bateman gradients, a supposed telltale sign of sexual selection acting on males.

A pooled graph over multiple populations obviates the opportunity to test sexual selection predictions, as I have already explained. Quantitative geneticists, by definition numerate scholars, from Fisher onward failed to notice that the graphs of pooled data were misrepresentations—in fact, they were false representations of the claimed outcome of sexual selection. Why did quantitative geneticists make that mistake? They surely knew that sexual selection is a within-population, within-sex process. The data for a relationship between number of mates and reproductive success must represent a within-population process, because for mate choice or within-sex rivalry to have fitness effects, individuals must interact with each other. To be more than perfectly plain, flies in separate jars could not interact with each other. The rivalry of individuals in one jar could not affect the fitness of same-sex individuals in another jar. The mate choice of individuals in one jar could not affect the mating success of individuals in other jars. Uncritical readers looking at Bateman's (1948) graphs made inferential mistakes, thinking that some males interacted with four other rivals when they could actually interact with only two rivals or that some individuals chose among five possible mates when they could choose among only three possible mates. In fact, Bateman's published graphs (see Figure 8.2) imply absolutely nothing about sexual selection.

Gowaty et al. (2012) repeated a key part of Bateman's original study: the populations that Bateman (1948) included in his graphs (see Figure 8.2). To do so, they constituted 46 populations, each with three males and three females, and tested in each population the distribution of offspring types. They found that "none of the 46 populations had more than 24.3% DD offspring, 44 had fewer than 20% (range from 6.9%)" DD offspring. The binomial probability that all 46 populations would have DD frequencies under 25% was 1.42×10^{-14} (p. 11741). None of these replicated populations could be used to reliably test Bateman's hypotheses. It was a simple control analysis to test if some of the combinations of subjects' nametag alleles would be lethal in offspring. Bateman (1948, p. 355) contained a table with a description of the dominant marker genes that he used to identify subjects in his experiment,

and it gave hints that offspring lethality was possible. The table showed that seven of 10 markers were homozygote lethal, meaning that the mutations had strong effects on viability, especially if a fly inherited the mutation at both alleles. It would have seemed completely reasonable, then, for Bateman (as it was for Gowaty et al.) to imagine that a DD offspring, heterozygote dominant at two loci, might experience lethality effects from inheriting two different dramatic mutations even if they were at different loci, because both mutations would have been expressed. For example, it would be no surprise if an offspring in Bateman's experiment that simultaneously inherited two dramatic nametag mutations (e.g., curly wings and microcephaly) died before emerging as adult (when their inherited parental nametags would have been visible to observers): Bateman's clever method was not so clever without a check on the required frequencies of offspring phenotypes given parental types. What is also befuddling is that the simple Mendelian expectations require no special numeracy beyond fractions to figure out. Why did Bateman not do it? Why did later readers not ask the following simple question: How reliable were the data for the questions Bateman asked? Did those armies of readers find Bateman's hypotheses—observations—conclusions—principles so compelling because they were seeing what they believed?

So, what explains the math mistakes in Bateman's (1948) analyses? He obviously did not have an electronic calculator, much less a computer; he probably used a mechanically cranked calculator, if that. Arithmetic errors are not really important, because with the data in hand, one can always redo the analysis, which is what Thierry Hoquet and I recently did (Hoquet & Gowaty, 2016). Hoquet unearthed and transcribed Bateman's original, handwritten lab notes; we computerized the data and then used SAS statistical software to analyze the data, population by population. First, we asked whether the data in each population met Mendel's rules. Thirty-eight of 65 populations passed Mendel's test; however, more than 40% failed. Thirty-eight fair populations is a reasonably large sample of populations, much larger than the vast majority of modern studies of sexual selection, so we continued to evaluate Bateman's hypotheses using his data. Our report showed that

when using the appropriate level of analysis—that is, within populations that met Mendel’s rules—(a) there were ~~no~~ statistically significant sex differences ~~in any population~~ in the number of mates or the variance in number of mates (V_{NM}); (b) there ~~were~~ ~~no~~ statistically significant sex differences in ~~any~~ population in the reproductive success or the variance in reproductive success (V_{RS}); and (c) the dependence of V_{RS} on V_{NM} was not statistically greater for males than for females. We concluded that Bateman’s data reject Bateman’s hypotheses. Do the Hoquet and Gowaty (2016) results mean that all other tests of Bateman’s hypotheses are also wrong? Obviously not, and we do not claim so. However, Bateman’s original data are inconsistent with his hypotheses. And that fact gives pause. What I conclude is that true believers allowed a flawed study to guide further studies as though Bateman’s conclusions were rock solid, principles one could count on.

If the classic study was flawed and no one noticed until we did, how many other subsequent studies of Bateman’s principles are flawed? Is it reasonable to claim that all are correct? Given that so much about Bateman (1948) was unexamined and errors were overlooked for so long, and given that the errors in Bateman were elementary, it is hard to understand why so many people remain so committed to a set of principles that may also be wrong, particularly when their implications are likely partial at best. After Gowaty et al. (2012) appeared, an article in *Heredity* (Wade & Shuster, 2005) urged readers not to throw the baby out with the bath water. Presumably the baby not to be thrown out was Bateman’s principles. Perhaps it is time to think hard about how much sense Bateman’s principles really make. Questioning the solidity of those principles is an act the field of evolutionary sex science should embrace.

Bateman’s (1948) conclusions, which became *Bateman’s principles* (Arnold, 1994), were that greater variance in reproductive success (V_{RS}) for males than females is a sign of “intramascu- line selection” and that the higher variance in number of

mates (V_{NM}) for males is also a sign of intramascu- line selection, but the stronger correlation in males than females of V_{NM} with V_{RS} is the cause of intra- masculine selection.⁵

How does all that make sense of intramascu- line selection or, as it is called today, “sexual selection acting among males”? First, higher V_{RS} may some- times indicate greater opportunity for selection, but high variances in reproductive success do not necessarily indicate selection without other informa- tion. For example, one needs information about the mechanisms by which the fitness variances arose, as well as the trait variation that mate choice or same- sex rivalry might have sorted. None of that informa- tion was available from Bateman’s (1948) study. If the inference hangs on the sex differences in variance in reproductive success (V_{RS}), it is even harder to infer that it is due to sexual selection acting on males, because logically—following Darwin’s syllogisms of sexual selection—the among-female variance in reproductive success says nothing about sexual selection on males, and vice versa. Again, more infor- mation is needed. The pattern of variation in repro- ductive success of males and females might more readily indicate a sign of intrinsic limits on egg pro- duction in females, itself a reliable indicator of past strong selection acting on females, not just a sign of opportunity for selection. Just as greater male vari- ance in reproductive success is a potential sign of an opportunity for selection, the lower female variance in reproductive success is not a sign that sexual selec- tion is not acting among females. Nor is it a sign that in females there is little opportunity for selection, because sexual selection on females may act through other components of fitness, such as the number of offspring that actually go on themselves to reproduce (i.e., females may compete over the quality of their offspring). There was never enough information in analyses of sex differences in V_{RS} or V_{NM} to infer a sign of sexual selection acting on either sex.

Second, sex differences in variances in number of mates (V_{NM}) could be—as Bateman (1948) argued—because females are inherently choosy,

⁵In Bateman’s (1948) own words, “It can now be seen that the sex difference in variance of fertility, which is itself a sign of intra-masculine selection, is due to the effect of number of mates per fly on fertility. This takes effect in two ways: (a) The higher variance, in males, of the number of mates per fly. This is a sign of intra-masculine selection. (b) The stronger correlation, in males, between number of mates and fertility. This is the cause of intra-masculine selection” (p. 362).

finding no benefit in repeat or multiple mating, or sex differences in V_{NM} could be because postmating latencies (or the handling time of a previous copulation) of females are naturally longer than for males, reducing the time females have available for remating, suggesting as it did to Sutherland (1985) that chance can explain sex differences in the number of mates and their variances.

Another alternative occurred to Snyder and Gowaty (2007), namely, that a type of sexual conflict could explain Bateman's (1948) data: If ejaculate components from a previous mate physiologically manipulate females' remating behavior, females may be delayed in returning to receptivity to further mating, which might reduce their mating rate. Yet another alternative is that female-female sexual selection may not be over number of mates but over the quality of mates, or the quality of offspring a female can produce given with whom she shares gametes. If this alternative stands, the silent claim in Bateman's principles that sexual selection proceeds the same way in males and females is violated, so that any inference of sexual selection acting on males using data on sex differences in variances in number of mates (V_{NM}) would be an uninformative exercise. The point is that explanations other than choosy females and competitive males can explain Bateman's data. Obviously, the existence of alternative hypotheses challenges claims that sex differences in V_{NM} are signs of sexual selection acting on males because they are also justifiable signs of chance effects, perhaps signs of sexual conflict, signs of an otherwise flawed methodology, or simply signs of a bad graph.

Why, if the question is about intramasculine selection, is a comparison of sex differences in V_{NM} relevant? While standing on Darwin's shoulders, to my eyes there seems nothing to be learned about sexual selection on either sex from a graph of sex differences in slopes of reproductive success on number of mates. To me, using Darwin's logic of sexual selection, it makes more sense to seek evidence of within-sex selection with comparisons between populations of fitness variances of one sex. One might manipulate male trait variation between populations to study how the mechanisms of mate choice and same-sex rivalry play out under

differences in experimentally manipulated male trait variation or demographic or social circumstances and then observe treatment effects on fitness variances. High mean reproductive success and lower variances among males in some treatments relative to males in other treatments would in fact be able to yield signs of selection between treatments, if there were associations between treatments and male fitness variances, either in their survival or their reproductive success.

As I have emphasized throughout, Darwin's syllogisms of sexual selection make clear that selection within one sex might tell one nothing about sexual selection in the other sex. My collaborators and I proposed such a study to the National Science Foundation. We proposed evaluating parentage across a series of experiments varying the demographic circumstances of individuals, so that it would be possible to make strong inferences about the effects of individual survival probabilities, the probabilities of encounter of potential mates, and other demographic variables on fitness variances of same-sex individuals in different populations experimentally manipulated to increase or decrease opportunities for mate choice and same-sex rivalry. Our proposal was declined because reviewers said Bateman's principles have been repeatedly validated, thus implying that no more testing is necessary.

THE CANONICAL TAPESTRY UNRAVELS

From observations of sex differences, investigators predict the existence of other sex differences, which affirms the consequent. A first-principles approach it is not, but rather one that may encourage confirmation bias, defined as "the tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs or hypotheses, while giving disproportionately less consideration to alternative possibilities" (see "Confirmation Bias," 2017, para. 1). Confirmatory associations between reproductive decisions and theories of anisogamy or parental investment tell one nothing about causes of selection, because even formal correlations are not causation. The closest of the foundational texts to a first-principles theory is the anisogamy math that deduced from an analytical solution the origin

of variation in gamete morphology. The common conclusion from anisogamy theory that males are competitive and females are choosy is, however, nothing more than an intuitive association about presumptive selection pressures that from an essentialist worldview seem to be linked with gamete size and mobility. What is empirically most often left undone is strong inference testing of the hypotheses of sex differences in behavior: Are the suppositions correct? Are there alternatives to the ideas that past selection for morphological variation simultaneously produced fixed sex differences in behavior? In the literature of sex differences, some investigators (e.g., Schärer, Rowe, & Arnqvist, 2012) have claimed anisogamy theory as the self-evident origins of sex differences in behavior and fitness. The kinds of tests that would allow rejection of the predictions, if they are indeed false, are seldom recognized and almost never done; if done, inconsistent results are characterized as curious exceptions to the rule and almost never claimed as inconsistent with the predictions of anisogamy or parental investment theories, even when they are. Consistency science (Janicke et al., 2016) inappropriately fights back with confirmatory claims.

As for Bateman's (1948) principles, they were built on quicksand: The errors of method, inference, and later the associated myths (e.g., that Bateman watched behavior) are a catalogue of unconscious hype (a nicer word than *lies*) from true believers, who are perhaps unconsciously promoting false findings. Given that so many overlooked the elementary errors in Bateman's article, I wonder whether contemporary readers and reviewers have also failed to catch the errors of inference that may exist in the modern studies reporting Bateman gradients. How many of those articles failed to test the genetic assumptions of their methods to test for reliability, given their questions? How many have tested their results under chance expectations? How many of those studies pooled data across populations, rather than limiting their analyses within populations within breeding seasons, wherein individuals interact to affect outcomes associated with mate choice and same-sex rivalry over access to more or better mates? How many of those studies documented the existence and character of same-sex

rivalries? How many of those studies evaluated whether individuals of either sex made assessments of mate quality before allowing subjects to mate? How many evaluated the basis of mate choice, in terms of either quality of potential mates or access to resources? How many studies controlled for the effects of potential male manipulation of female reproduction? For that matter, how many controlled for possible effects of female manipulation of male reproduction? If male manipulation of females' reproductive decisions occurs as it does in *D. melanogaster*, female remating rate may be inhibited by behavioral sexual conflict, as in mate guarding, or components in ejaculates that increase the latency of females to remating, increasing the likelihood that the variance in number of mates is smaller in females not because of evolved female tendencies but because of males' manipulations (Angier, 1999). How many modern studies test for the effects of chance on mean number of mates (V_{NM}) and mean reproductive success (V_{RS})? Why is it taking so long to produce credible, testable alternative hypotheses? Is it because many investigators are satisfied with "the answer," because it reflects what they believe anyway? Failure to pursue the few extant alternative hypotheses seems antiscientific and surely inhibits efficient science.

A HYPOTHESIS ABOUT THE PSYCHOLOGY OF SCIENTISTS STUDYING SEX DIFFERENCES

In a 1974 commencement speech about science given at the California Institute of Technology, Richard P. Feynman said, "The first principle is that you must not fool yourself and you are the easiest person to fool." He went on to imply that a type of integrity in science is "taking care not to fool yourself." I argue that we should self-consciously take care not to fool ourselves: We should use the scientific method that demands we repeat studies, honestly report what we did and why we did it, and ferret out what is wrong with our own studies, not just what is right or expected. We need to use the best experimental controls—double-blind observations, blinded analyses, sharing our data, and honest attempts to understand the work of others—in order

to constructively criticize each other. And we must control investigator-specific biases, that is, our own individual and particular brands of sex-gendered expectations. To do that, of course, implies and requires a type of self-awareness that may be difficult to come by. However, I believe it is essential in evolutionary studies of sex differences and similarities that we at least attempt to confront our own biases with appropriate controls. No doubt, this is hard to do. As Root Gorelick (personal communication, May 31, 2016) reminded me in an email, “I don’t mind people being biased, so long as they are willing to eventually let go of it. Bias is knowledge, providing a framework that allows either progress or regress to be possible. The problem is when we all have the same biases.” Indeed, common gender schemas are very nearly universal biasing beliefs. It may be time for evolution-minded sex scientists to consistently challenge studies, demanding explicit attention to controls against common gender schemas. This is something that will be hard because self-deception may inhibit our views of ourselves.

That gender schemas are common and shared sources of internalized bias predicts that it may be easier to fool ourselves while studying the evolution of sex differences than, say, while seeking evidence of gravitational waves. Common conceptual frames about the nature of women and men are like an ouroboros: They prop up essentialist visions of the sexes and help produce further essentialist scientific visions of the sexes.

References

- Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., & Axt, J. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716. <http://dx.doi.org/10.1126/science.aac4716>
- Ah-King, M., & Gowaty, P. A. (2016). A conceptual review of mate choice: Stochastic demography, within-sex phenotypic plasticity, and individual flexibility. *Ecology and Evolution*, 6, 4607–4642. <http://dx.doi.org/10.1002/ece3.2197>
- Altmann, J. (1997). Mate choice and intrasexual reproductive competition: Contributions to reproduction that go beyond acquiring more mates. In P. A. Gowaty (Ed.), *Feminism and evolutionary biology: Boundaries, intersections and frontiers* (pp. 320–333). http://dx.doi.org/10.1007/978-1-4615-5985-6_13
- Anderson, W. W. (1969). Selection in experimental populations: I. Lethal genes. *Genetics*, 62, 653–672.
- Anderson, W. W. (1974). Frequent multiple insemination in a natural population of *Drosophila pseudoobscura*. *American Naturalist*, 108, 709–711. <http://dx.doi.org/10.1086/282949>
- Anderson, W. W., Kim, Y. K., & Gowaty, P. A. (2007). Experimental constraints on mate preferences in *Drosophila pseudoobscura* decrease offspring viability and fitness of mated pairs. *Proceedings of the National Academy of Sciences, USA*, 104, 4484–4488. <http://dx.doi.org/10.1073/pnas.0611152104>
- Angier, N. (1999). *Woman: An intimate geography*. New York, NY: Houghton Mifflin Harcourt.
- Arnold, S. J. (1994). Bateman’s principles and the measurement of sexual selection in plants and animals. *American Naturalist*, 144, S126–S149. <http://dx.doi.org/10.1086/285656>
- Bateman, A. J. (1948). Intra-sexual selection in *Drosophila*. *Heredity*, 2, 349–368. <http://dx.doi.org/10.1038/hdy.1948.21>
- Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., . . . van’t Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217–224. <http://dx.doi.org/10.1016/j.jesp.2013.10.005>
- Chin, J. M. (2014). Psychological science’s replicability crisis and what it means for science in the courtroom. *Psychology, Public Policy, and Law*, 20, 225–238. <http://dx.doi.org/10.1037/law0000012>
- Confirmation bias. (2017). Retrieved from https://en.wikipedia.org/wiki/Confirmation_bias
- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, England: John Murray.
- Darwin, C. (1871). *The descent of man, and selection in relation to sex* (Vol. 1). London, England: John Murray.
- Dennett, D. C. (1995). *Darwin’s dangerous idea: Evolution and the meanings of life*. New York, NY: Simon & Schuster.
- Dewsbury, D. A. (1982). Ejaculate cost and male choice. *American Naturalist*, 119, 601–610. <http://dx.doi.org/10.1086/283938>
- Dewsbury, D. A. (2005). The Darwin-Bateman paradigm in historical context. *Integrative and Comparative Biology*, 45, 831–837. <http://dx.doi.org/10.1093/icb/45.5.831>
- Drickamer, L. C., Gowaty, P. A., & Holmes, C. M. (2000). Free female mate choice in house mice affects reproductive success and offspring viability and performance. *Animal Behaviour*, 59, 371–378. <http://dx.doi.org/10.1006/anbe.1999.1316>

- Drickamer, L. C., Gowaty, P. A., & Wagner, D. M. (2003). Free mutual mate preferences in house mice affect reproductive success and offspring performance. *Animal Behaviour*, 65, 105–114. <http://dx.doi.org/10.1006/anbe.2002.2027>
- Feynman, R. R. (1974). *Cargo cult science*. Retrieved from <http://calteches.library.caltech.edu/51/2/CargoCult.pdf>
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford, England: Oxford University Press.
- Gowaty, P. A. (2003). Sexual natures: How feminism changed evolutionary biology. *Signs: Journal of Women in Culture and Society*, 28, 901–921. <http://dx.doi.org/10.1086/345324>
- Gowaty, P. A. (2011). What is sexual selection and the short herstory of female trait variation. *Behavioral Ecology*, 22, 1146–1147. <http://dx.doi.org/10.1093/beheco/arr113>
- Gowaty, P. A. (2014). Standing on Darwin's shoulders: The nature of selection hypotheses. In T. Hoquet (Ed.), *Current perspectives on sexual selection: What's left after Darwin* (pp. 103–118). New York, NY: Springer.
- Gowaty, P. A., Anderson, W. W., Bluhm, C. K., Drickamer, L. C., Kim, Y.-K., & Moore, A. J. (2007). The hypothesis of reproductive compensation and its assumptions about mate preferences and offspring viability. *Proceedings of the National Academy of Sciences, USA*, 104, 15023–15027. <http://dx.doi.org/10.1073/pnas.0706622104>
- Gowaty, P. A., Drickamer, L. C., & Schmid-Holmes, S. (2003). Male house mice produce fewer offspring with lower viability and poorer performance when mated with females they do not prefer. *Animal Behaviour*, 65, 95–103. <http://dx.doi.org/10.1006/anbe.2002.2026>
- Gowaty, P. A., & Hubbell, S. P. (2005). Chance, time allocation, and the evolution of adaptively flexible sex role behavior. *Integrative and Comparative Biology*, 45, 931–944. <http://dx.doi.org/10.1093/icb/45.5.931>
- Gowaty, P. A., & Hubbell, S. P. (2009). Reproductive decisions under ecological constraints: It's about time. *Proceedings of the National Academy of Sciences, USA*, 106(Suppl. 1), 10017–10024. <http://dx.doi.org/10.1073/pnas.0901130106>
- Gowaty, P. A., & Hubbell, S. P. (2010). Killing time: A mechanism of sexual conflict and sexual selection. In J. L. Leonard & A. Córdoba-Aguilar (Eds.), *The evolution of primary sexual characters in animals* (pp. 79–96). New York, NY: Oxford University Press.
- Gowaty, P. A., & Hubbell, S. P. (2013a). Bayesian animals sense ecological constraints to predict fitness and organize individually flexible reproductive decisions. *Behavioral and Brain Sciences*, 36, 215–216. <http://dx.doi.org/10.1017/S0140525X12002385>
- Gowaty, P. A., & Hubbell, S. P. (2013b). The evolutionary origins of mating failures and multiple mating. *Entomologia Experimentalis et Applicata*, 146, 11–25. <http://dx.doi.org/10.1111/eea.12023>
- Gowaty, P. A., Kim, Y.-K., & Anderson, W. W. (2012). No evidence of sexual selection in a repetition of Bateman's classic study of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA*, 109, 11740–11745. <http://dx.doi.org/10.1073/pnas.1207851109>
- Gowaty, P. A., Kim, Y.-K., & Anderson, W. W. (2013). Mendel's law reveals fatal flaws in Bateman's 1948 study of mating and fitness. *Fly*, 7, 28–38. <http://dx.doi.org/10.4161/fly.23505>
- Gowaty, P. A., Kim, Y.-K., Rawlings, J., & Anderson, W. W. (2010). Polyandry increases offspring viability and mother productivity but does not decrease mother survival in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences, USA*, 107, 13771–13776. <http://dx.doi.org/10.1073/pnas.1006174107>
- Gowaty, P. A., Steinichen, R., & Anderson, W. W. (2002). Mutual interest between the sexes and reproductive success in *Drosophila pseudoobscura*. *Evolution: International Journal of Organic Evolution*, 56, 2537–2540. <http://dx.doi.org/10.1111/j.0014-3820.2002.tb00178.x>
- Gowaty, P. A., Steinichen, R., & Anderson, W. W. (2003). Indiscriminate females and choosy males: Within- and between-species variation in *Drosophila*. *Evolution: International Journal of Organic Evolution*, 57, 2037–2045. <http://dx.doi.org/10.1111/j.0014-3820.2003.tb00383.x>
- Hessellmann, G., & Moors, P. (2015). Definitely maybe: Can unconscious processes perform the same functions as conscious processes? *Frontiers in Psychology*, 6, 584. <http://dx.doi.org/10.3389/fpsyg.2015.00584>
- Hilborn, R., & Mangel, M. (1997). *The ecological detective: Confronting models with data* (Vol. 28). Princeton, NJ: Princeton University Press.
- Hoquet, T., & Gowaty, P. A. (2016). *Bateman's data are inconsistent with Bateman's principles*. Manuscript in preparation.
- Hrdy, S. B. (1981). *The woman that never evolved*. Cambridge, MA: Harvard University Press.
- Hrdy, S. B. (1990). Sex bias in nature and in history: A late 1980s reexamination of the "biological origins" argument. *American Journal of Physical Anthropology*, 33(Suppl. S11), 25–37. <http://dx.doi.org/10.1002/ajpa.1330330504>
- Hrdy, S. B. (1999). *The woman that never evolved* (rev. ed.). Cambridge, MA: Harvard University Press.
- Hrdy, S. B. (2006). Empathy, polyandry, and the myth of the coy female. In E. Sober (Ed.), *Conceptual*

- issues in evolutionary biology (3rd ed., pp. 131–160). Cambridge, MA: MIT Press.
- Hubbell, S. P., & Johnson, L. K. (1987). Environmental variance in lifetime mating success, mate choice, and sexual selection. *American Naturalist*, 130, 91–112. <http://dx.doi.org/10.1086/284700>
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581–592. <http://dx.doi.org/10.1037/0003-066X.60.6.581>
- Janicke, T., Häderer, I. K., Lajeunesse, M. J., & Anthes, N. (2016). Darwinian sex roles confirmed across the animal kingdom. *Science Advances*, 2, e1500983. <http://dx.doi.org/10.1126/sciadv.1500983>
- Johnson, L. K. (1982). Sexual selection in a brentid weevil. *Evolution: International Journal of Organic Evolution*, 36, 251–262. <http://dx.doi.org/10.2307/2408043>
- Johnson, L. K., & Hubbell, S. P. (1984). Male choice. *Behavioral Ecology and Sociobiology*, 15, 183–188. <http://dx.doi.org/10.1007/BF00292973>
- Jones, A. (2003). Validation of Bateman's principles: Comparative evidence from taxa with conventional and reversed sex roles. *Integrative and Comparative Biology*, 43, 834.
- Jones, A. G., Arguello, J. R., & Arnold, S. J. (2002). Validation of Bateman's principles: A genetic study of sexual selection and mating patterns in the rough-skinned newt. *Proceedings of the Royal Society B: Biological Sciences*, 269, 2533–2539. <http://dx.doi.org/10.1098/rspb.2002.2177>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, S., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45, 142–152. <http://dx.doi.org/10.1027/1864-9335/a000178>
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Loehle, C. (1987). Hypothesis testing in ecology: Psychological aspects and the importance of theory maturation. *Quarterly Review of Biology*, 62, 397–409. <http://dx.doi.org/10.1086/415619>
- López-López, W. L. (2014). Scientific publishing and the commitments of researcher communities. *Universitas Psychologica*, 13, 323–324.
- Makel, M. C. (2014). The empirical march: Making science better at self-correction. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 2–7. <http://dx.doi.org/10.1037/a0035803>
- Marquet, P. A., Allen, A. P., Brown, J. H., Dunne, J., Enquist, B. J., Gillooly, J. F., . . . West, G. B. (2014). On theory in ecology. *Bioscience*, 64, 701–710. <http://dx.doi.org/10.1093/biosci/biu098>
- Martin, E. (1991). The egg and the sperm: How science has constructed a romance based on stereotypical male-female roles. *Signs: Journal of Women in Culture and Society*, 16, 485–501. <http://dx.doi.org/10.1086/494680>
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., . . . Van der Laan, M. (2014). Social science: Promoting transparency in social science research. *Science*, 343, 30–31. <http://dx.doi.org/10.1126/science.1245317>
- Moore, A. J., Gowaty, P. A., Wallin, W. G., & Moore, P. J. (2001). Sexual conflict and the evolution of female mate choice and male social dominance. *Proceedings of the Royal Society B: Biological Sciences*, 268, 517–523. <http://dx.doi.org/10.1098/rspb.2000.1399>
- Murayama, K., Pekrun, R., & Fiedler, K. (2014). Research practices that can prevent an inflation of false-positive rates. *Personality and Social Psychology Review*, 18, 107–118. <http://dx.doi.org/10.1177/1088868313496330>
- Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, 13, 88. <http://dx.doi.org/10.1186/s12915-015-0196-3>
- Parish, A. R. (1994). Sex and food control in the “uncommon chimpanzee”: How bonobo females overcome a phylogenetic legacy of male dominance. *Ethology and Sociobiology*, 15, 157–179. [http://dx.doi.org/10.1016/0162-3095\(94\)90038-8](http://dx.doi.org/10.1016/0162-3095(94)90038-8)
- Parish, A. R., de Waal, F. B., & Haig, D. (2000). The other “closest living relative”: How bonobos (*Pan paniscus*) challenge traditional assumptions about females, dominance, intra- and intersexual interactions, and hominid evolution. *Annals of the New York Academy of Sciences*, 907, 97–113. <http://dx.doi.org/10.1111/j.1749-6632.2000.tb06618.x>
- Parker, G. (1979). Sexual selection and sexual conflict. In M. S. Blum & N. A. Blum (Eds.), *Sexual selection and reproductive competition in insects* (pp. 123–166). <http://dx.doi.org/10.1016/B978-0-12-108750-0.50010-0>
- Parker, G. A., Baker, R. R., & Smith, V. G. (1972). The origin and evolution of gamete dimorphism and the male-female phenomenon. *Journal of Theoretical Biology*, 36, 529–553. [http://dx.doi.org/10.1016/0022-5193\(72\)90007-0](http://dx.doi.org/10.1016/0022-5193(72)90007-0)
- Platt, J. R. (1964). Strong inference. *Science*, 146, 347–353. <http://dx.doi.org/10.1126/science.146.3642.347>
- Popper, K. R. (1945). *The open society and its enemies*. Abingdon, England: Routledge & Kegan Paul.
- Popper, K. R. (1959). *The logic of scientific discovery*. Abingdon, England: Routledge.
- Ryan, K. K., & Altmann, J. (2001). Selection for male choice based primarily on mate compatibility in the oldfield mouse, *Peromyscus polionotus rhoadsi*.

- Behavioral Ecology and Sociobiology*, 50, 436–440. <http://dx.doi.org/10.1007/s002650100385>
- Schärer, L., Rowe, L., & Arnqvist, G. (2012). Anisogamy, chance and the evolution of sex roles. *Trends in Ecology and Evolution*, 27, 260–264. <http://dx.doi.org/10.1016/j.tree.2011.12.006>
- Schiebinger, L. (1991). *The mind has no sex? Women in the origins of modern science*. Cambridge, MA: Harvard University Press.
- Schiebinger, L. (2001). *Has feminism changed science?* Cambridge, MA: Harvard University Press.
- Schiebinger, L. L. (1993). *Nature's body: Gender in the making of modern science*. New Brunswick, NJ: Rutgers University Press.
- Snyder, B. F., & Gowaty, P. A. (2007). A reappraisal of Bateman's classic study of intrasexual selection. *Evolution: International Journal of Organic Evolution*, 61, 2457–2468. <http://dx.doi.org/10.1111/j.1558-5646.2007.00212.x>
- Sutherland, W. J. (1985). Chance can produce a sex difference in variance in mating success and explain Bateman's data. *Animal Behaviour*, 33, 1349–1352. [http://dx.doi.org/10.1016/S0003-3472\(85\)80197-4](http://dx.doi.org/10.1016/S0003-3472(85)80197-4)
- Trivers, R. (1972). *Parental investment and sexual selection* (Vol. 136). Cambridge, MA: Biological Laboratories, Harvard University.
- Turner, M. E., & Anderson, W. W. (1983). Multiple mating and female fitness in *Drosophila pseudoobscura*. *Evolution: International Journal of Organic Evolution*, 37, 714–723. <http://dx.doi.org/10.2307/2407913>
- Wade, M. J., & Shuster, S. M. (2005). Don't throw Bateman out with the bathwater! *Integrative and Comparative Biology*, 45, 945–951. <http://dx.doi.org/10.1093/icb/45.5.945>
- Wicherts, J. M., & Bakker, M. (2014). Broken windows, mediocre methods, and substandard statistics. *Group Processes and Intergroup Relations*, 17, 388–403. <http://dx.doi.org/10.1177/1368430213502557>
- Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.