

Chapter 3 Methods of Evolutionary Sciences

Jeffrey A. Simpson and Lorne Campbell

Methods in Evolutionary Psychology

Charles Darwin (1859) began formulating the theory of evolution by natural selection almost 20 years before he published *The Origin of Species*. One of the main reasons he waited to publish his iconoclastic book was that he did not have sufficient evidence to support his theory (Desmond & Moore, 1991). Testing the theory of evolution by natural selection was a truly major and complicated task, so Darwin used several different methods to marshal support for the theory. He spoke, for example, with animal breeders to learn about artificial selection. Over time, he discovered that heritable variation in domesticated traits was shaped by the preferences of breeders, a process similar to the natural selection of traits. He also surveyed the existing scientific literature on myriad species in their natural environments, carefully describing and cataloguing the large amount of variation that existed both within and between species. And he spent countless days experimenting with seeds to determine whether they germinated after being exposed to different types of conditions. Armed with a huge amount of information from all his observations, field studies, and experiments, Darwin was eventually able to marshal sufficient initial support for the basic principles of the theory of evolution by natural selection. It was partly Darwin's relentless tenacity at gathering and analyzing data from multiple sources that resulted in his theory eventually being accepted by the wider scientific community.

Both the theory and science of evolution have progressed remarkably since 1859. Indeed, Darwin's vision that his theory of evolution would provide the foundation for the study of psychology is coming to fruition in a growing number of academic disciplines. This is a very exciting time for the evolutionary sciences. However, a larger number of researchers need to emulate Darwin by adopting a more multifaceted approach when studying psychological adaptations. To do so, researchers must take advantage of all the many different investigative methods that are currently available.

To facilitate this process, this chapter revisits some of the fundamental principles and concepts that have anchored research methods in the social and behavioral sciences for several decades. Our hope is that it will also kindle (or

rekindle) greater interest in methodological issues by not only showcasing the many research methods currently available to evolutionary scientists, but also by clarifying *how* different research methods, measures, and statistical techniques can be utilized to make clearer, stronger, and more precise tests of evolutionary-based predictions.

The chapter has three overarching themes. The first is that, to provide stronger and more definitive tests of theories, multiple research methods and outcome measures must be used to test alternate models *within* ongoing programs of evolutionary research. Each major research method (e.g., laboratory experiments, surveys, computer simulations) and each type of outcome measure (e.g., self-reports, peer-ratings, behavioral ratings) have strengths and limitations. No single method or measure is optimal in every research context because different methods, measures, and techniques entail trade-offs between maximizing internal validity, external validity, and the generalizability of findings across participants. Both methodological triangulation within programs of research (i.e., adopting a multiple-method/multiple-measure approach when testing for effects) and the testing of alternative models are required to arrive at strong, clear inferences.

A second theme is that there has been a general overreliance on certain research methods (e.g., correlational approaches) and certain measures (e.g., self-reports) in some areas within the evolutionary sciences. In some cases, this mono-method/mono-measure focus has impeded the rigorous testing of certain evolutionary-based phenomena; in others, it has not allowed investigators to discern whether the results predicted by evolutionary theories fit observed data better than alternative competing theories. This problem remedied by greater knowledge and appreciation of the numerous strengths and advantages that multiple research methods and different paradigms can offer.

A third organizing theme is the need to test and provide better evidence for the “special design” properties of psychological adaptations. In some situations, a multimethod/multimeasure approach can help researchers provide better and stronger evidence for the “special design” features of certain purportedly evolved traits, behaviors, or characteristics in humans. The telltale signs of selection and adaptation should be most evident when specific stimuli (triggering events) produce specific effects (responses) across different levels of measurement (ranging from molecular to macro levels). Converging patterns of findings from well-conducted multimethod/multimeasure studies can appreciably *increase* our confidence that certain “specially designed” adaptations probably did, in fact, evolve. We now turn to the first major topic of the chapter, which centers on theory testing, special design, and “strong” research methods.

Theory Testing, Special Design, and Strong Research Methods

Many evolutionary theories confront relatively high evaluation standards given the sheer complexity—and sometimes imprecision—of the metatheories in which they are grounded. As a rule, evolutionary theories tend to be more complex than other theories, including historical origin theories that do not have an evolutionary basis such as certain social structuralist theories (e.g., Eagly, 1987). One reason for this is that inferring simple associations between distal biologically based adaptations and how current psychological processes operate is more complicated than inferring associations between cultural or social structural factors and current psychological processes. More complex theories usually generate a larger number of “internal” alternative explanations, which makes it more difficult to derive straightforward predictions about whether and how certain traits or behaviors were—or should have been—adaptive in our ancestral past (see Caporael & Brewer, 2000; Dawkins, 1989).

This problem has been magnified by the relative lack of attention often devoted to (a) clarifying how different middle-level evolutionary theories are or are not interrelated, and (b) specifying the conditions under which different theories make similar versus different predictions about specific outcomes (Simpson & Belsky, 2008). Evolutionary theories are hierarchically organized and they have several levels of explanation, ranging from broad metatheoretical assumptions, to domain-relevant middle-level principles, to specific hypotheses, to specific predictions (Buss, 1995; Ketelaar & Ellis, 2000). Most middle-level evolutionary theories (such as parental investment, attachment, parent-offspring conflict, reciprocal altruism) extend the core assumptions of their metatheories to specific psychological domains, such as the conditions under which individuals invest in their offspring, bond with them, experience conflict with them, or assist others who are not biologically related to them. In some cases, middle-level theories generate competing hypotheses and predictions. Parental investment theory, for instance, makes different predictions than reciprocal altruism theory does regarding when men should invest in young, biologically unrelated children of unattached women (see Ketelaar & Ellis, 2000). In other cases, middle-level evolutionary theories spawn hypotheses that vie with nonevolutionary theories (e.g., the debate about why homicide is so prevalent in “families”; see Daly & Wilson, 1988). Little attention is typically paid to which outcomes different competing theories or models—either evolutionary based or otherwise—logically anticipate. *Whenever possible, tests between predictions that have been logically derived from competing models should be built into evolutionary research programs.*

At times, evolutionary researchers also do not fully explain the deductive logic that connects one level of explanation (such as the basic principles of a middle-level theory) to adjacent levels (such as a specific set of hypotheses). One reason for this is that evolutionary hypotheses exist along a “continuum of confidence,” which ranges from: (a) clear and firm hypotheses that are unequivocally and directly derived from a middle-level theory, to (b) expectation-based hypotheses that can be logically deduced from a theory, but cannot be directly derived from it without making auxiliary assumptions, to (c) speculative hypotheses based on casual or intuitive hunches.

How Can More Compelling Evidence Be Generated?

How can the evolutionary sciences overcome these limitations? As a start, researchers must articulate clearer, more specific, and more detailed models of the historical events that should have produced an evolved trait or attribute (Conway & Schaller, 2002). Supportive evidence must also be gathered from a wide range of disciplines (e.g., anthropology, zoology, genetics, evolutionary biology) to justify the “starting assumptions” of a proposed historical theory or model and to explain why it is more probable than other theories or models. To accomplish this, evolutionary scientists must conduct more refined cost–benefit analyses relevant to the evolutionary history of each purported adaptation (Cronin, 1991). Specifically, greater attention must focus on the probable costs, constraints, and limitations—social, physical, behavioral, physiological, and otherwise—that might have counterweighted the conjectured benefits associated with a hypothesized adaptation (see Eastwick, 2009). After conducting these analyses, researchers must elucidate why certain adaptations should have produced better solutions to specific evolutionarily relevant problems than other possible adaptations, and good tests of alternative models should be performed.

These limitations might also be rectified if investigators structured more of their research around the predictions that specific evolutionary theories or models make regarding the onset, operation, and termination of specific psychological processes or mechanisms. When doing so, a clear conceptual distinction must be maintained between models of historical (evolutionary) events and the current psychological events or processes being examined (see Tinbergen, 1963). This can be achieved by organizing research questions around Buss’s (1995, pp. 5–6) incisive definition of evolved psychological mechanisms:

An evolved psychological mechanism is a set of processes inside an organism that: (1) Exists in the form it does because it (or other mechanisms that reliably produce it) solved a specific problem of

individual survival or reproduction recurrently over human evolutionary history; (2) Takes only certain classes of information or input, where input (a) can be either external or internal, (b) can be actively extracted from the environment or passively received from the environment, and (c) specifies to the organism the particular adaptive problem it is facing; (3) Transforms that information into output through a procedure (e.g., decision rule) in which output (a) regulates physiological activity, provides information to other psychological mechanisms, or produces manifest action and (b) solves a particular adaptive problem.

When developing and testing the deductive logic of a theory, therefore, evolutionary scientists should: (a) articulate how and why specific selection pressures should have shaped certain psychological mechanisms or processes, (b) identify the specific environmental cues that should have activated these processes in relevant ancestral environments, (c) explain how these processes should have guided thoughts, feelings, and behavior in specific social situations, and (d) specify the cues or outcomes that should have terminated these psychological processes or mechanisms. The wider adoption of this general approach could yield several benefits. First, by clarifying and more rigorously testing the deductive logic underlying an evolutionary theory or model, investigators are in a better position to articulate how and why their theory provides a *forward-thinking* account of specific psychological processes or mechanisms rather than an ad hoc, backward-thinking explanation. Second, because subtle connections between different theoretical levels are more fully explained, the theory or model being tested should have greater explanatory coherence. Third, sounder and more extensive deductive logic will help researchers to derive more novel predictions. The most powerful theories generate new and unforeseen predictions that cannot be easily derived from alternative theories. Many novel hypotheses are likely to involve statistical interactions in which certain psychological mechanisms are activated or terminated by very specific environmental inputs. And theories that predict specific types of context-dependent statistical interactions usually have fewer alternative explanations.

Adaptations, Adaptationism, and Standards of Evidence

At a conceptual level, many evolutionary psychologists adopt a general investigative orientation known as adaptationism. Using this approach, researchers try to identify the specific selection pressures that shaped the evolution of certain traits or characteristics in our ancestral past (Thornhill, 1997; Williams, 1966). This approach asks questions of the form “What is the function or purpose of this particular structure, organ, or characteristic?” Answers to such questions have

produced rapid and significant advances in many areas of science. With respect to human evolution, some adaptationist research programs have used optimization modeling (e.g., testing different formal mathematical theories of possible selection pressures in the environment of evolutionary adaptedness (EEA); Parker & Maynard Smith, 1990) to provide evidence for certain presumed adaptations in humans. Most programs, however, have simply developed plausible, intuitive arguments regarding how a given trait or characteristic might have evolved to solve specific evolutionary problems (Williams, 1966, 1992).

The general adaptationist approach has been criticized by Gould and Lewontin (1979), who claim that most adaptationist research has used weak or inappropriate standards of evidence to identify adaptations. They argue that most adaptationist research simply demonstrates that certain outcomes are consistent with theoretical predictions without fully examining competing alternative accounts. Gould (1984) has also argued that most adaptationist research has overemphasized the importance of selection pressures and underestimated the many constraints on selection forces, leading some adaptationists to presume that adaptations exist when rigorous evidence is lacking. Gould and Lewontin (1979) maintain that many constraints—genetic, physical, and developmental—may have opposed or hindered the impact that selection pressures had on most phenotypic traits and characteristics over evolutionary time. Thus, they claim that exaptations (i.e., preexisting traits that take on new beneficial effects without being modified by new selection pressures) are numerous, making it nearly impossible to recreate the selection history of a given trait or characteristic. Most adaptations are, in fact, built on earlier adaptations, exaptations, or spandrels (i.e., byproducts that happen to be associated with adapted traits). The evolutionary sciences, therefore, must use methodologies capable of documenting specific adaptations more directly (Mayr, 1983; see also Buss, Haselton, Shackelford, Bleske, & Wakefield, 1998).

What types of evidence have been gathered to test whether certain traits or psychological attributes might be adaptations? Andrews, Gangestad, and Matthews (2003) discuss six standards of evidence: (1) *Comparative standards*, which make specific phylogenetic comparisons regarding a purportedly adaptive trait across different species; (2) *Fitness maximization standards*, which identify particular traits that ought to maximize fitness returns in particular environments, including current ones; (3) *Beneficial effects standards*, which focus on the fitness benefits that a presumably adaptive trait could have produced in ancestral environments; (4) *Optimal design standards*, which test formal mathematical simulations of how different selection pressures might have produced trade-offs in evolved features and how fitness could have been increased by trading off the features of one trait against others; (5) *Tight fit standards*, which examine how closely a presumably adaptive trait's features match, and should have efficiently solved, a major evolutionary problem; and (6) *Special design standards*, which identify and test the unique functional properties of a purportedly adaptive trait.

The first five standards offer indirect evidence that a given trait might be an adaptation. The sixth standard—special design—provides much more rigorous evidence (Andrews et al., 2003). *Thus, evolutionary research programs must be developed, organized, and structured around providing more firm and direct evidence for the special design properties of possible adaptations.* As more special design features of a hypothesized adaptation are documented, each contributing to a specific function, it becomes more plausible that the hypothesized adaptation actually evolved for that function. The best and most rigorous evolutionary research programs routinely test for special design features (also see Schmitt & Pilcher, 2004).

Special Design Evidence

Organisms are living historical documents (Williams, 1992). Accordingly, adaptations should reveal remnants of the selective forces that shaped them. Before a trait can be classified as an adaptation, however, its primary evolutionary function or purpose must first be ascertained (Mayr, 1983; Thornhill, 1997). To accomplish this, the specific selection pressures that most likely generated and shaped the functional design of the trait must be inferred. Functionally designed traits tend to perform a purpose “with sufficient precision, economy, efficiency, and so forth to rule out pure chance as an adequate explanation” (Williams, 1966, p. 10). Chance factors can include processes such as phylogenetic legacy, genetic drift, byproduct effects, and mutations, any of which could be responsible for the development of a particular trait.

Several additional factors make it difficult to determine whether a particular trait is an adaptation. These include the potentially confounding effects of historically prior adaptations (e.g., those upon which more recent “secondary adaptations” may have been constructed), trade-offs between interacting adaptations (e.g., selection for camouflage from predators versus colorful ornamentation to attract mates), and counteradaptations (e.g., countervailing mating tactics that emerge between the sexes in a species). Further complicating matters, different traits may require different types of evidence to demonstrate their special design properties. For example, the special design features of many morphological traits (e.g., the human eye, body organs) have been demonstrated simply by showing that a particular trait has complex design and performs a specific function with a very high degree of precision, economy, and efficiency. Additional evidence, however, is often needed for complex behavioral and cognitive traits believed to be adaptations because domain-general learning processes (such as exapted learning mechanisms) can produce traits with considerable specificity, proficiency, and complexity (see Andrews et al., 2003). For these “complex traits,” further evidence for their special design properties is often required.

Fortunately, several sources of evidence can increase our confidence about the “special design” of certain traits (Andrews et al., 2003). First, complex trait adaptations can be documented by showing that a trait is a biased outcome of a specific developmental or learning mechanism (Cummins & Cummins, 1999). Such traits develop or are learned very easily, quickly, and reliably, and they tend to solve specific adaptive problems with much greater proficiency than other traits that could have been produced by the same underlying mechanisms. Examples include the strong and automatic propensity to fear certain objects (e.g., snakes, Öhman & Mineka, 2001), the capacity to develop grammar and language (Pinker, 1994), the environmentally specific conditioning associated with punishment (Garcia, Hankins, & Rusiniak, 1974), and the perceptual expectations and preferences of young infants (Spelke, 1990). Second, complex adaptations can be demonstrated by showing that a trait’s specially designed features would have solved major problems in ancestral environments, but tend to be dysfunctional or harmful in modern environments. One example is the strong cravings that most people—especially young children—have for foods high in fat and sugar (Drewnowski, 1997). Third, complex adaptations can be documented by revealing that alternative theories or processes do not predict or cannot explain certain outcomes (e.g., the superior spatial location memory of women, Silverman & Eals, 1992; the superior cheater detection capabilities of both sexes, Cosmides, 1989). Finally, confidence in a trait’s adaptive status increases when several traits all serve the same basic function (e.g., the factors that govern shifts in women’s mate preferences across the reproductive cycle; see Thornhill & Gangestad, 2008).

There are, of course, some drawbacks to using special design as the sole evidentiary criteria for adaptations. It might, for example, be difficult to provide unambiguous evidence for the special design features of certain adaptations. To guard against this possibility, investigators should test not only for the special design features of specific traits, but should provide some evidence for the other standards as well. Adaptations may also be difficult to identify because many complex traits may have mixed design (e.g., female orgasm, the development of the neocortex; see Andrews et al, 2003). If, for instance, a trait initially evolved as an adaptation for one effect, then was exapted for a different purpose, and then became a secondary adaptation for yet another purpose, the trait could serve multiple functions that were shaped by different—and perhaps even conflicting—selection pressures. This would obscure the trait’s specially designed features.

Validity Issues

Validity is generally defined as “the best available approximation to the truth or falsity of propositions” (Cook & Campbell, 1979, pp. 37), so it reflects the

degree of truth regarding the statements, inferences, or conclusions drawn from empirical research. Because research programs have different missions, the validity of a given study must be evaluated in the context of the broader goals, purposes, and objectives of a given research program.

A Process Model of Validity

The procedures for establishing the validity of an operationalization or measure of a construct are similar to those for developing, testing, and confirming scientific theories (Loevinger, 1957). Since the operations and measures used in any single study are imperfect and incomplete representations of the theoretical constructs they are designed to assess, theory testing is an ongoing, cyclical process in which constructs inform research operations, which generate revised constructs, which in turn suggest new and improved operations.

Two methodological traditions have influenced how validity is defined and conceptualized. One tradition, grounded in experimental and quasi-experimental research, has focused on the validity of independent variables, particularly their conceptualization, their operationalization, and how they are perceived by participants (Cook & Campbell, 1979). A second tradition, stemming from non-experimental research in personality and clinical psychology, has focused on the validity of dependent variables and psychological scales (Cronbach & Meehl, 1955; Loevinger, 1957).

Bridging these traditions, Brewer (2000) proposed a three-stage process model of how hypothetical theoretical constructs are conceptually linked with three sets of measures: (1) observable stimuli (independent variables), (2) intervening physiological or cognitive processes (those occurring *within* individuals), and (3) observable responses (dependent or outcome variables). As shown in Figure 3.1, researchers need to make three inferential connections when planning and conducting studies.

Figure 3.1 Constructs and operationalizations. The vertical lines represent hypotheses connecting observed measures with their underlying theoretical processes/constructs. Adapted from Brewer (2000) with permission. [ch03f01]

On the independent variable side, they first must make important assumptions, inferences, and decisions about how the latent causal concepts specified by their theory should be operationally defined and manifested in the independent variables. If they are interested in essentialist causation, researchers must also establish solid inferential ties between the mediation processes predicted by their theory and the measures identified as possible mediators. On the

dependent variable side, they must derive clear inferential connections between the effects anticipated by their theory and the responses (outcomes) that are measured. Numerous problems can undermine valid inferences from a study at each stage. To complicate matters, many areas of evolutionary science lack standardized measures, operations, or procedures that correspond closely with the latent theoretical constructs of interest. Because of this, evolutionary scientists must often make fairly large inferential leaps across each set of linkages.

These difficulties can create thorny methodological problems. For example, the validity of stimulus or response measures may be called into question if the variations (either manipulated or measured levels) in a given study do not mirror the typical levels of variation in the theoretical states that the stimuli or responses are designed to assess. Moreover, it may be difficult to predict the precise levels at which certain independent variables should (or should not) have causal effects on certain outcome measures. And it might be challenging to anticipate the range over which certain independent variables should have their strongest effects on specific outcomes. Given the multitude of ways in which the validity of a study can be reduced, it is often difficult to determine whether null results from a single study reflect a failure of the theory, the operationalizations at one or more of Brewer's (2000) three stages, and/or the measures employed.

Validity in Experimental and Quasi-Experimental Research

There are four basic types of validity in experimental and quasi-experimental research (Cook & Campbell, 1979): (1) internal validity, (2) statistical conclusion validity, (3) external validity, and (4) construct validity.

Internal validity reflects the degree to which a researcher can be confident that a manipulated variable (X) has a causal impact on an outcome measure (Y). The internal validity of a study is high when one can confidently conclude that variations in Y were produced by manipulated changes in the level or intensity of X (i.e., the independent variable had a causal influence on the dependent variable, *independent* of other possible causal factors). If third variables correlate with X, these confounds can generate spurious effects. Fortunately, true experiments control for the deleterious influence of third variables through random assignment of participants to experimental conditions and through careful operationalizations and manipulations of the independent variables.

Moderating and mediating variables, however, can complicate causal inferences (Preacher & Hayes, 2008). Moderating effects exist when there is a true causal connection between an independent variable (X) and a dependent variable (Y), but the relation varies at different levels of some third variable (C).

Evolutionary scientists, for instance, might posit that an experimental manipulation of high versus low physical threat should lead most highly threatened individuals to stand and defend themselves. This link, however, might be moderated by gender, with men being more likely to adopt the “stand and defend” response under high threat than women.

Mediating effects occur when a third variable (C) is needed to complete the causal process (pathway) between X and Y. That is, systematic changes in an independent variable (X) predict changes in the mediator (C), which then predicts changes in the dependent variable (Y), statistically controlling for X. Returning to our example, evolutionary scientists might also postulate that a high level of physical threat should lead most men to experience “challenge” physiological responses that prepare them to stand and defend. Such threats, however, might lead most women to experience “threat” physiological responses, leading them to engage in different tactics.

A second major type of validity, *statistical conclusion validity*, involves the degree to which a researcher can infer that two variables reliably covary, given a specified alpha level and the observed variances. Statistical conclusion validity is a special form of internal validity, one that addresses the effects of random error and the appropriate use of statistical tests rather than the effects of systematic error. This form of validity can be undermined by having insufficient statistical power (leading to Type II statistical errors), violating important assumptions of statistical tests (e.g., that errors are uncorrelated when they are actually correlated), suffering from inflated experiment-wise error rates (which occur when multiple statistical tests are performed without adjusting the *p* values for the number of tests conducted), or when measures have low reliabilities. Statistical conclusion validity can also be threatened if treatment or condition implementations are unreliably administered, if random events occur during experiments, or if respondents differ in how they interpret the meaning of treatments, independent variables, or outcome measures.

A third major form of validity, *external validity*, involves the degree to which a researcher can generalize from a study: (a) *to* particular target persons or settings, or (b) *across* different persons, settings, and times. The external validity of a study can be assessed by testing for statistical interactions (i.e., whether an effect holds across different persons, settings, or times), and it can be enhanced by conducting several heterogeneous studies. External validity is threatened when statistical interactions exist between selection and treatment (i.e., do recruitment factors make it easier for certain people to enter particular treatments or conditions?), between setting and treatment (i.e., do similar treatment or condition effects emerge across different research settings?), or between history and treatment (i.e., do effects generalize across different time periods?).

Brewer (2000) distinguishes three forms of external validity: ecological validity, relevance, and robustness. *Ecological validity* is the extent to which an effect occurs under conditions that are “typical” or “common” for a given population. *Relevance* reflects the degree to which findings are useful or applicable in solving social problems or improving the quality of life. *Robustness* (sometimes called “generalizability”) has the most important implications for evolutionary research because it reflects the degree to which a finding is replicable across different settings, people, and historical contexts.

To evaluate the robustness of an effect, theorists must clearly define the populations and settings to which it should (and should not) generalize. Within the evolutionary sciences, generalizability from one prototypical participant population at one time period (e.g., Westernized college students in current environments) to target populations from other time periods (e.g., typical hunters and gathers in our ancestral past) is one of the most common external validity concerns. Similar concerns have been raised in other fields within psychology (see Arnett, 2008). Evolutionary scientists need to articulate the principle ways in which contemporary participant populations are likely to differ from more traditional hunter/gatherer “target” populations and how these differences may qualify the interpretation of certain evolutionary findings.

The fourth type of validity—*construct validity*—is the most encompassing form of validity. Construct validity reflects the degree to which operations that are intended to represent a given causal construct *or* effect construct can be explained by alternate constructs (Cronbach & Meehl, 1955). For causal constructs, construct validity addresses the question, “Does a finding reveal a causal relation between variable X and variable Y, between variable Z and variable Y (which might also correlate with variable X), or with some other outcome variable?” For effect constructs such as outcome measures, construct validity addresses the question, “From a theoretical standpoint, does this measure/scale correlate with measures with which it should covary (convergently), and does it *not* correlate with measures with which it should not correlate (discriminantly)?”

Most independent variables are complex packages of multiple and sometimes correlated variables. For example, when an experimenter tries to induce social isolation in participants, the manipulation may produce other unanticipated states, such as heightened anxiety, depressive symptoms, or negative moods. Many of the concerns about construct validity, therefore, revolve around how independent variables are (or should be) operationalized in particular studies and how they are perceived by participants. An experimental manipulation might also elicit multiple hypothetical states in the same individual, making it nearly impossible to identify the specific causal agent that is operative in a study. Cook and Campbell (1979) claim that the most serious threat to the construct validity of causal constructs is a mono-operation bias—the recurrent use of a single method or paradigm to assess a theoretical construct. *Conceptual replications that involve*

different operationalizations of the same construct are essential to demonstrate sufficient construct validity.

Multitrait-Multimethod Approaches

Gathering evidence for the construct validity of a trait or scale requires testing its convergent and discriminant validation properties. This can be accomplished using the Multitrait-Multimethod Matrix approach (Campbell & Fiske, 1959). Measures have three sources of variance: (1) variance that a construct was intended to assess (convergent validity components), (2) variance that a construct was *not* intended to assess (systematic error variance), and (3) random error due to unreliability of the measures. All studies fall into one of four categories: (1) monotrait-monomethod (when a single trait/scale is studied using one research method), (2) monotrait-heteromethod (when a single trait/scale is studied using different methods), (3) heterotrait-monomethod (when different traits/scales are studied using one method), or (4) heterotrait-heteromethod (when multiple traits/scales are studied using multiple methods). Heterotrait-heteromethod approaches are preferable because they allow researchers to test for both the convergent and discriminant validation properties of traits/scales. Strong evidence for convergent validity exists when a trait/scale correlates with measures that tap theoretically similar constructs, even when the trait/scale is measured using different methods. Compelling evidence for discriminant validity exists when a trait/scale does *not* correlate with measures that tap theoretically independent or unrelated constructs, even when the same methods are used.

Statistical Power

Another very important issue when designing studies is statistical power, a topic that has been brought back to the forefront of discussion in recent years (e.g., Schimmack, 2012). Statistical power is the probability of rejecting the null hypothesis when it is false (Cohen, 1988), but it is more commonly discussed as the ability to detect an effect if it actually exists. Having an appropriate level of statistical power (approximately .80 or 80%), therefore, is essential to test hypotheses adequately, particularly those derived from evolutionary models that are novel or counterintuitive, such as the effects of ovulation on women's mate preferences.

The average power of most published studies in various fields is worryingly low. Cohen (1988) has lamented the low power of most published research in psychology, and he has advocated strongly for increasing the power of

studies. Despite this clarion call, little improvement in power had been achieved in the intervening years. Recently, Bakker, van Dijk, and Wicherts (2012) estimated the average power of studies in psychology to be 35%, and Button et al. (2013) estimated the average power of research in neuroscience to be even lower (21%). When the fact that over 90% of all published research reports statistically significant (non-null) findings (Sterling, Rosenbaum, & Weinkam, 1995) is combined with the low power of most studies, a non-trivial amount of published findings must be false positives (Ioannidis, 2005).

To understand how low power negatively impacts research, Button et al. (2013) identified three problems of low statistical power. First, with low power, there is a low probability of discovering true effects and, thus, a high level of false negatives. When a researcher genuinely believes that an effect exists (particularly if it has not been documented yet), it makes sense to design a well-powered study. Otherwise, a real effect may go undiscovered and the research process becomes inefficient and a waste of time for everyone involved (the researchers, ethics board evaluators, research assistants, and participants).

Second, low power combined with the prior probability of an effect being true at $p < .05$ can result in low positive predictive value (PPV), the probability that the effect found is indeed true (Ioannidis, 2005). When power is low, the probability that a set of effects are true effects decreases. Low power, in other words, makes it more difficult to find true effects (false negatives), and it can also result in discovering effects that are not true (false positives).

Third, low power can produce exaggerated estimates of the magnitude of an effect when a true effect is discovered. This phenomenon is known as the “winner’s curse” (Ioannidis, 2008). Because low powered studies can detect only large effects, when a true effect is not very large, a low powered study will overestimate the size of the effect when the results happen to pass the threshold for statistical significance (i.e., capitalizing on chance in small samples). Subsequent replication studies using the same number of participants will not be likely to detect the effect, and replication attempts using sample sizes at least 2.5 times larger than the original sample are generally required to detect the effect, if it exists (Simonsohn, 2013). Additionally, when high powered replication studies do detect the effect, it is likely to be much smaller than in the original low powered study.

Schimmack (2012) has identified another problem with low power when multiple studies are reported. Specifically, when a series of low powered studies consistently reject the null hypothesis, this undermines one’s ability to conclude that the effect (or effects) truly exists, given the low probability of obtaining this pattern of positive effects over multiple studies. Instead of obtaining a string of positive effects, the likelihood of finding null effects in some of these tests is actually *higher* when the power of the studies is *lower*. Schimmack proposes that

fewer but more adequately powered studies provide more convincing and statistically defensible support for hypotheses.

Levels of Analysis and Phylogenetic Approaches

To marshal truly compelling and complete evidence for a purportedly evolved trait or behavior, one needs to distinguish between four distinct levels of analysis—adaptive function, ontogenetic development, proximate determinants, and evolutionary history (Tinbergen, 1963). Adaptive function (ultimate) explanations are concerned with the evolved adaptive purpose of a given trait or behavior. An adaptive function explanation, for example, might focus on associations between dominance and reproductive success in males and females in a species such as the chimpanzee, noting that dominance is more critical to the reproductive success of males than females. Developmental (ontogenetic) explanations address the lifespan-specific inputs that sensitize an organism to particular cues in the environment. A developmental explanation, for instance, might address the fact that maturing male chimpanzees experience certain hormonal changes during adolescence, making them more likely to engage in dominance-related behaviors than females. Proximate explanations focus on the immediate triggers of a given trait/behavior, including its inputs, information processing procedures, and outputs. A proximate explanation, for example, might document that displays of male dominance are typically triggered by threats from other males, and that responses to other males' displays are stronger when circulating testosterone levels are higher. Finally, historical (phylogenetic) explanations consider the ancestral roots of a given trait or behavior in relation to other species. Researchers who adopt this approach, for example, might view sex differences in chimpanzee dominance relative to other primate species or other social mammals (i.e., increasingly more distant relatives), observing that males are larger and more competitive in most mammalian species. Comparative methods that address questions at the phylogenetic level are less utilized than other methods, yet they can clarify and extend our understanding of the evolutionary history of a given trait or behavior in a given species in important ways (see Eastwick, 2009).

Phylogenetic methods typically reside in section I (field studies) of the circumplex model shown in Figure 3.2. Certain traits or behaviors can be correlated within or between species for functional or nonfunctional reasons (Harvey & Pagel, 1991). Thus, documenting a correlation between two traits or behaviors does not mean they have evolved together between different species

over evolutionary time. To fully evaluate the adaptive nature of the covariation between a pair of traits or behaviors, one must model their phylogenetic relationship between specific species over time. Phylogenetic relationships are the specific patterns of descent and ancestry over very long periods of evolutionary time, which can be summarized in phylogenetic trees.

Figure 3.2 Research strategies. A = Point of maximum concern with generality across actors; B = Point of maximum concern with precision of measurement; C = Point of maximum concern with realism of the context. Source: From *Research on Human Behavior: A Systematic Guide to Method* (Figure 4-1, p. 85), by P. J. Runkel and J. E. McGrath, 1972, New York, NY: Holt, Rinehart, & Winston. [ch03f02]

AU: Can you confirm that you have acquired permission to use Figure 3.2? ED: Yes, we have permission.

Phylogenetic relationships are important to test and document in comparative studies because species that are more closely related phylogenetically (i.e., are closer to each other in phylogenetic trees) are often similar on many traits and behaviors (Blomberg, Garland, & Ives, 2003). Species may share similar morphologies and behaviors because they evolved from a common ancestor or because similar selection pressures generated the *independent* evolution of the traits/behaviors they have in common. Similarities among species, in other words, can be attributable to either *homology* (having a shared ancestry) or *analogy* (the independent evolution of the trait/behavior within each species, also known as convergent evolution).

Consider an example. Two traits may highly or even perfectly correlated in two or more living species, such that trait X (e.g., high paternal investment in offspring) always co-occurs with trait Y (e.g., adult pair-bonds between mates), and the absence of trait X always co-occurs with the absence of trait Y. If there are no species in which X is present but Y is not (or vice versa), the two traits are likely to be homologous; both traits are most likely shared (or not shared) because of some earlier ancestral species from which the current species evolved. The two traits, therefore, most likely emerged at the same time for functional reasons; they did not coevolve independently within each species.

If, however, the two traits evolved together at separate points during evolutionary history, there should be multiple occasions on which changes in one trait (e.g., paternal investment) were linked with the other trait (e.g., pair-bonding). Even though the two traits are highly or even perfectly correlated in the current species, they evolved independently in different lineages and the changes in the two traits just happen to be associated. The correlation between the two traits

across species, in other words, simply reflects their repeated and independent co-evolution (Harvey & Pagel, 1991).

Comparative phylogenetic methods can answer important and novel questions about the evolutionary history of evolved traits/behaviors. Fraley, Brumbaugh, and Marks (2005), for instance, have used these methods to investigate patterns of pair-bonding across different species. Doing so, they have found that the connection between the provision of paternal care and pair-bonding between mates is probably due to convergent evolution, whereas connection between neoteny and pair-bonding appears to be due to homology (shared ancestry). Eastwick (2009) has suggested that evolutionary scientists should build phylogenetic relationships between humans and our hominid and pongid relatives (both living and extinct) more directly into our theorizing. He proposes that if one considers the specific timing of evolutionary events along with evolutionary constraints, phylogenetic approaches can generate novel predictions about patterns of human mating and new explanations for existing findings such as adaptive workarounds, which are more evolutionary recent adaptations in a species that “manage” the maladaptive elements of pre-existing evolutionary constraints.

Research Programs Providing Good Evidence for Psychological Adaptations

Different traits or behaviors are likely to require different types of evidence to reveal their special design properties, but certain methodological strategies can facilitate the documentation of special design. The special design features of specific traits can be revealed by conducting research that: (a) uses multiple methods and multiple measures to assess and triangulate the major constructs, (b) tests for and systematically discounts alternative explanations for a trait’s uniquely designed functional features, and (c) reveals the footprints of special design at different measurement levels (ranging from neural mechanisms, to context-specific modes of information processing, to emotional reactions, to molar behavioral responses; see Wilson, 1998). Some programs of research have documented the special design properties of certain hypothesized psychological adaptations. Examples include research on the effects of father absence/involvement on daughters’ pubertal development (Ellis, McFadyen-Ketchum, Dodge, Pettit, & Bates, 1999), patterns of homicide in families with biological fathers versus stepfathers (Daly & Wilson, 1988), and mother-fetus

conflict during gestation (Haig, 1993). Two particularly laudatory programs of research are highlighted below.

Snakes and an Evolved Fear Module

Öhman, Mineka, and their colleagues have offered strong, programmatic, and compelling evidence that humans and closely related primates have an evolved “fear module” for reptiles (Öhman & Mineka, 2001). What makes this program of research exemplary is the nature, quality, and type of evidence that has been gathered for the special design features of this purported adaptation. This evidence has been strengthened by the use of multiple research methods (e.g., comparative methods, interviews, field observations, experimental laboratory studies) to test carefully derived predictions, by systematically testing and ruling out alternative theories and explanations, and by documenting the unique footprints of special design at multiple levels of measurement (ranging from neural mechanisms to general cognitive expectations and behavioral reactions).

Several interlocking findings clearly indicate that higher primates possess an evolved fear module (see Öhman & Mineka, 2001, for a review). Based on interviews with humans (Agras, Sylvester, & Oliveau, 1969), comparative field data on different primate species (King, 1997), and observations of primates living in captivity versus in the wild (Mineka, Keir, & Price, 1980), research has confirmed that humans and other higher primates have an acute fear of snakes with distant evolutionary origins. Conducting well-designed experiments, researchers have also demonstrated that lab-raised monkeys learn to fear snakes very quickly just by observing fearful expressions in other monkeys (Cook & Mineka, 1990), lab-raised monkeys show preferential conditioning to toy reptiles but not to innocuous stimuli such as toy rabbits (Cook & Mineka, 1991), and humans who receive shocks in the presence of snakes show longer, stronger, and qualitatively different conditioning responses than do humans who are shocked in the presence of non-aversive stimuli such as flowers (Öhman & Mineka, 2001). This body of findings implies that the strong association between snakes and aversive unconditioned stimuli emanates from the evolutionary history of primates rather than from culturally mediated conditioning processes.

Additional lab experiments have shown that humans automatically infer illusory associations between snakes and aversive stimuli. For example, people are more likely to perceive that fearful stimuli (snakes) co-occur with painful experiences (shocks) than is true of other non-fearful stimuli, even when there is no actual association between pairings of shock and different stimuli (Tomarken, Sutton, & Mineka, 1995). People also believe that shocks are more likely to follow exposure to dangerous stimuli such as snakes and damaged electrical equipment, but illusory correlations emerge only between snakes and shock once people have

been exposed to a random series of stimulus/shock trials (Kennedy, Rapee, & Mazurski, 1997). Experiments assessing visual detection latencies have confirmed that when people are shown large sets of stimulus pictures, snakes automatically capture their visual attention, regardless of how many distractor stimuli are present (Öhman, Flykt, & Esteves, 2001). These results suggest that humans are “prepared” to perceive associations and process visual information about snakes and aversive outcomes in systematically biased ways.

Experiments have also identified where in the brain the “fear circuit” might be located. Using backward masking techniques that present stimuli outside of conscious awareness, Öhman and Soares (1994, 1998) discovered that fear responses can be learned and activated, even when backward masking prevents images of snakes from reaching higher cortical processing. This indicates that fear responses reside in ancient neural circuits that evolved long before the full development of the neocortex.

Viewed together, this entire body of evidence strongly suggests that humans and higher primates have a fear module that evolved to reduce recurrent threats posed by dangerous and potentially lethal animals. This module is sensitive to, and is automatically activated by, a specific class of stimuli, it operates in specific areas of the brain (the amygdala) that evolved before the neocortex, and it has fairly specialized neural circuitry. This innovative program of research nicely illustrates how different research methods—lab and field experiments, field observations, comparative methods—can be used to provide compelling evidence for a specific, cross-species psychological adaptation whose footprints exist at different levels of analysis and measurement.

Mate Preferences in Women Across the Reproductive cycle

Several well-conceptualized and carefully designed studies have tested the ovulatory shift hypothesis—that women have an evolved psychological adaptation that motivates them to prefer men who have “good genes” as short-term mates, primarily when they are ovulating and could conceive a child with such men (Thornhill & Gangestad, 2008). This program of work is elegant because the predictions are carefully derived from good genes sexual selection models as well as cross-species data, the predictions are very specific (entailing specific statistical interaction patterns), the predictions and findings are difficult to derive from competing theories/models, and numerous alternative explanations have been discounted.

According to the Strategic Pluralism Model of mating (Gangestad & Simpson, 2000), women evolved to make trade-offs between two sets of attributes

when evaluating men as potential mates: men's degree of health/viability (their "good genes"), and their level of commitment/investment in the relationship and subsequent offspring. Fluctuating symmetry (FA: the extent to which individuals are bilaterally symmetrical at different locations of the body) is one good marker of health/viability (Thornhill & Gangestad, 2008). Thus, women should find more symmetrical men more attractive than less symmetrical men in short-term mating contexts, especially when they are ovulating (and could conceivably transmit the "good genes" of these men to their offspring). This model, therefore, predicts very specific statistical interaction patterns that are neither anticipated nor easily explained by alternative perspectives.

The ovulatory shift hypothesis has been tested using a variety of research methods and techniques (see Gildersleeve, Haselton, & Fales, 2014, for a review). Self-report questionnaire studies have confirmed that more symmetrical men are more likely to engage in extra-pair sex and are more prone to be chosen by women as extra-pair partners (e.g., Gangestad & Thornhill, 1997). Self-report and interview studies have revealed that women are more likely to have extra-pair affairs when they are ovulating, but they are *not* necessarily more prone to have sex with their current romantic partners during ovulation. Moreover, women report stronger sexual attraction to and fantasies about men other than their current romantic partners when they are ovulating (Gangestad, Thornhill, & Garver, 2002), a pattern that is *not* found for current partners unless they have "good genes" characteristics (e.g., Haselton & Gangestad, 2006).

To test predictions about olfactory markers of men's FA and ovulatory shifts in women, Gangestad and Thornhill (1998) had women smell unscented T-shirts worn by different men. If women were ovulating during the study, they rated the scents of more symmetrical men as more attractive than those of less symmetrical but, as predicted, this interaction effect did not emerge in nonovulating women. Providing discriminant validity evidence for this effect, Thornhill et al. (2003) also found that even though women prefer the scent of heterozygous major histocompatibility (MHC) alleles in men (which should be valued in primary partners because mating with an individual who has more diverse MHC alleles should limit infections within families), the preference for MHC did *not* increase when women were ovulating.

In a laboratory behavioral observation study, Simpson, Gangestad, Christensen, and Leck (1999) found that more symmetrical men displayed greater social presence and more direct intrasexually competitive tactics (rated by observers) than less symmetrical men when being interviewed by an attractive woman and competing against another man for a "lunch date." When a different group of women evaluated the videotaped interviews of these men and rated how attractive they found each one as a short-term and a long-term mate, women who were ovulating were significantly more attracted to men who displayed greater social presence and direct intrasexual competitiveness—the tactics displayed by

more symmetrical men—in short-term but *not* in long-term mating contexts (Gangestad, Simpson, Cousins, Garver-Apgar, & Christensen, 2004). Considered together, these findings confirm that women's mate preferences vary across the reproductive cycle in very specific and theoretically consistent ways.

Summary and Conclusions

Current research programs in the evolutionary sciences can be strengthened in several ways from a methodological standpoint. First, when feasible, researchers should use a wider range of research methods in their ongoing programs of work, especially more experimental methods and techniques. Second, a wider array of measurement and statistical techniques should be utilized. Third, sounder evidence needs to be provided regarding the validity of major manipulations, scales, and individual-item measures before they are adopted for widespread use (e.g., experimental manipulations of “social status,” self-report measures of “mate value”). Fourth, greater attention should focus on deducing, modeling, and testing the features of psychological mechanisms that are believed to be evolved adaptations. Fifth, stronger and better evidence is needed to determine how well outcomes predicted by different evolutionary theories or models fit different data sets, especially relative to competing non-evolutionary theories or models. Whenever possible, alternative constructs and explanations should be carefully derived and measured to test and adjudicate between competing constructs or models. Sixth, the special design features of purported adaptations should be directly specified and tested at different levels of analysis and measurement. Seventh, evidence for possible adaptations needs to be procured for multiple evidentiary standards. Eighth, empirical evidence for specific hypotheses should be gathered in different cultures, especially those that are more similar to the environments in which ancestral humans evolved. Finally, more effort must be devoted to developing and testing novel predictions, particularly those that cannot be easily derived or explained by competing theories.

In conclusion, evolutionary scientists need to emulate the methodological breadth and creativity of Charles Darwin. This can be accomplished in part by utilizing a broader array of research methods and statistical techniques, many of which can help investigators map out and comprehend the evolved architecture of the human mind much more precisely. To convince the wider scientific community of the value as well as the predictive, explanatory, and integrative power of evolutionary approaches, evolutionary theories and models need to be developed more carefully, derived more precisely, and tested more thoroughly than theories that do not involve historical origins. Given their tremendous explanatory and integrative power, some evolutionary theories have, at times,

proceeded ahead of good empirical evidence, especially with respect to humans. Recent advances in research and statistical methods are now closing this gap. However, evolutionary researchers must continue to refine the deductive logic of their theoretical models, revise or alter questionable or conflicting tenets of middle-level theories, discard or recast problematic hypotheses, and formulate more specific hypotheses that explicitly test the special design properties of presumed adaptations. If these goals are achieved, the evolutionary sciences will continue make rapid and significant theoretical and empirical progress in the coming years.

References

- Agras, S., Sylvester, D., & Oliveau, D. (1969). The epidemiology of common fears and phobias. *Comprehensive Psychiatry*, *10*, 151–156.
- Andrews, P. W., Gangestad, S. W., & Matthews, D. (2003). Adaptationism—how to carry out an exaptationist program. *Behavioral and Brain Sciences*, *25*, 489–504.
- Arnett, J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*, 602–614.
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543–554.
- Blomberg, S. P., Garland, T., Jr., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*, 717–745.
- Brewer, M. B. (2000). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 3–16). New York, NY: Cambridge University Press.
- Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry*, *6*, 1–30.
- Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., & Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *American Psychologist*, *53*, 533–548.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

- Caporael, L. R., & Brewer, M. B. (2000). Metatheories, evolution, and psychology: Once more with feeling. *Psychological Inquiry*, *11*, 23–26.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Conway, L. G., & Schaller, M. (2002). On the verifiability of evolutionary psychological theories: An analysis of the psychology of scientific persuasion. *Personality and Social Psychology Review*, *6*, 152–166.
- Cook, M., & Mineka, S. (1990). Selective associations in the observational conditioning of fear in rhesus monkeys. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 372–389.
- Cook, M., & Mineka, S. (1991). Selective associations in the origins of phobic fears and their implications for behavior therapy. In P. Martin (Ed.), *Handbook of behavior therapy and psychological science: An integrative approach* (pp. 413–434). Oxford, England: Pergamon Press.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition*, *31*, 187–276.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.
- Cronin, H. (1991). *The ant and the peacock: Altruism and sexual selection from Darwin to today*. New York, NY: Cambridge University Press.
- Cummins, D. D., & Cummins, R. (1999). Biological preparedness and evolutionary explanation. *Cognition*, *73*, B37–B53.
- Daly, M., & Wilson, M. (1988). *Homicide*. New York, NY: Aldine de Gruyter.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or, preservation of favoured races in the struggle for life*. London, England: Murray.
- Dawkins, R. (1989). *The selfish gene*. New York, NY: Oxford University Press.
- Desmond, A., & Moore, J. (1991). *The life of a tormented evolutionist*. New York, NY: Norton.
- Drewnowski, A. (1997). Taste preferences and food intake. *Annual Review of Nutrition*, *17*, 237–253.
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.

- Eastwick, P. W. (2009). Beyond the Pleistocene: Using phylogeny and constraint to inform the evolutionary psychology of human mating. *Psychological Bulletin*, *135*, 794–821.
- Ellis, B. J., McFadyen-Ketchum, S., Dodge, K. A., Pettit, G., & Bates, J. (1999). Quality of early family relationships and individual differences in the timing of pubertal maturation in girls: A longitudinal test of an evolutionary model. *Journal of Personality and Social Psychology*, *77*, 387–401.
- Fraley, R. C., Brumbaugh, C. C., & Marks, M.J. (2005). The evolution and function of adult attachment: A comparative and phylogenetic analysis. *Journal of Personality and Social Psychology*, *89*, 731–746.
- Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavior and Brain Sciences*, *23*, 573–587.
- Gangestad, S. W., Simpson, J. A., Cousins, A. J., Garver-Apgar, C. E., & Christensen, P. N. (2004). Women's preferences for male behavioral displays change across the menstrual cycle. *Psychological Science*, *15*, 203–207.
- Gangestad, S. W., & Thornhill, R. (1997). The evolutionary psychology of extra-pair sex: The role of fluctuating asymmetry. *Evolution and Human Behavior*, *18*, 69–88.
- Gangestad, S. W., & Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proceedings of the Royal Society of London, Series B*, *265*, 727–733.
- Gangestad, S. W., Thornhill, R., & Garver, C. E. (2002). Changes in women's sexual interest and their partners' mate retention tactics across the menstrual cycle: Evidence for shifting conflicts of interest. *Proceedings of the Royal Society of London B*, *269*, 975–982.
- Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1974). Behavior regulation of the milieu interne in man and rat. *Science*, *185*, 824–831.
- Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014). Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. *Psychological Bulletin*, *140*(5), 1205–1259.
- Gould, S. J. (1984). Only his wings remained. *Natural History*, *93*, 10–18.
- Gould, S.J., & Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, Series B* *205*, 581–598.
- Haig, D. (1993). Genetic conflicts in human pregnancy. *Quarterly Review of Biology*, *68*, 495–532.
- Harvey, P.H., & Pagel, M.D. (1991). *The comparative method in evolutionary biology*. Oxford, England: Oxford University Press.

- Haselton, M. G., & Gangestad, S. W. (2006). Conditional expression of women's sexual desires and male mate retention efforts across the ovulatory cycle. *Hormones and Behavior*, *49*, 509–518.
- Ioannidis, J. P. A. (2005) Why most published research findings are false. *PLoS Med* *2*, e124. doi:10.1371/journal.pmed.0020124
- Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*, 640–648.
- Kennedy, S. J., Rapee, R. M., & Mazurski, E. J. (1997). Covariation bias for phylogenetic versus ontogenetic fear-relevant stimuli. *Behaviour Research and Therapy*, *35*, 415–422.
- Ketelaar, T., & Ellis, B. J. (2000). Are evolutionary explanations unfalsifiable?: Evolutionary psychology and the Lakatosian philosophy of science. *Psychological Inquiry*, *11*, 1–21.
- King, G. E. (1997). *The attentional bias for primate responses to snakes*. Paper presented at the annual meeting of the American Society of Primatologists, San Diego, CA.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.
- Mayr, E. (1983). *Toward a new philosophy of biology: Observations of an evolutionist*. Cambridge, MA: Harvard University Press.
- Mineka, S., Keir, R., & Price, V. (1980). Fear of snakes in wild- and laboratory-reared rhesus monkeys (*Macaca mulatta*). *Animal Learning and Behavior*, *8*, 653–663.
- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, *131*, 466–478.
- Öhman, A., & Mineka, S. (2001). Fear, phobias and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*, 483–522.
- Öhman, A., & Soares, J. J. F. (1994). “Unconscious anxiety”: Phobic responses to masked stimuli. *Journal of Abnormal Psychology*, *103*, 231–240.
- Öhman, A., & Soares, J. J. F. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following nonrecognized fear-irrelevant stimuli. *Journal of Experimental Psychology: General*, *127*, 69–82.
- Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, *348*, 27–33.
- Pinker, S. (1994). *The language instinct*. New York, NY: William Morrow.

- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavioral Research Methods, 40*, 879–891.
- Runkel, P. J., & McGrath, J. E. (1972). *Research on human behavior: A systematic guide to method* (p. 85). New York, NY: Holt, Rinehart, & Winston.
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17*, 551–566.
- Schmitt, D.P., & Pilcher, J.J. (2004). Evaluating evidence of psychological adaptation: How do we know one when we see one? *Psychological Science, 15*, 643–649.
- Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 487–503). New York, NY: Oxford University Press.
- Simonsohn, U. (2013, December 10). *Small telescopes: Detectability and the evaluation of replication results*. Retrieved from <http://ssrn.com/abstract=2259879>
- Simpson, J. A., & Belsky, J. (2008). Attachment theory within a modern evolutionary framework. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 131–157). New York, NY: Guilford Press.
- Simpson, J. A., Gangestad, S. W., Christensen, P. N., & Leck, K. (1999). Fluctuating asymmetry, sociosexuality, and intrasexual competitive tactics. *Journal of Personality and Social Psychology, 76*, 159–172.
- Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*, 25–56.
- Sterling, T. D., Rosenbaum, W., & Weinkam, J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician, 108*–112.
- Thornhill, R. (1997). The concept of an evolved adaptation. In M. Daly (Ed.), *Characterizing human psychological adaptations* (pp. 4–22). London, England: CIBA Foundation.
- Thornhill, R., & Gangestad, S. W. (2008). *The evolutionary biology of human female sexuality*. New York, NY: Oxford University Press.
- Thornhill, R., Gangestad, S. W., Miller, R., Scheyd, G., McCullough, J. K., & Franklin, M. (2003). Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women. *Behavioral Ecology, 14*, 668–678.
- Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*, 410–433.

- Tomarken, A. J., Sutton, S. K., & Mineka, S. (1995). Fear-relevant illusory correlations: What types of associations promote judgmental bias? *Journal of Abnormal Psychology, 104*, 312–326.
- Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.
- Williams, G. C. (1992). *Natural selection: Domains, levels and challenges*. Oxford, England: Oxford University Press.
- Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York, NY: Knopf.