DARWIN (1859) BEGAN THINKING and writing about the theory of evolution by natural selection years before he published *The Origin of Species*. Although he had stumbled upon an idea that would forever change biology and the life sciences, Darwin kept his thoughts to himself for several years because he did not have sufficient evidence to support his iconoclastic views (Desmond & Moore, 1991). He needed some provisional empirical evidence for his grand theory.

Testing the theory of evolution by natural selection was no small task, so Darwin used several different methods for this purpose. For example, Darwin spoke with animal breeders to learn about artificial selection. He eventually discerned that heritable variation in domesticated traits was shaped by the preferences of breeders and likened this process to the natural selection of traits. He also surveyed the existing scientific literature on species in their natural environments, describing and cataloguing the vast amount of variation that existed within and between species. Additionally, he spent countless hours experimenting with seeds to determine whether they germinated after being exposed to various conditions. Armed with information from his observations, field studies, and experiments, Darwin was able to provide initial support for the basic premises underlying the theory of evolution by natural selection. Indeed, it was Darwin’s relentless perseverance gathering data from multiple sources that permitted his theory ultimately to be embraced by the wider scientific community.

Both the theory and science of evolution have progressed tremendously since the publication of *The Origin of Species*. Darwin’s belief that evolution would provide a foundation for the study of psychology is now coming to fruition as many researchers in a growing number of academic disciplines are beginning to discover the evolutionary structure and architecture of the human mind. This is a very exciting time for the evolutionary sciences. However, it is also a time for
researchers to emulate the research practices of Darwin by adopting a multifaceted approach to the study of psychological adaptations. To do so, researchers must be aware of the different investigative methods that are available. To facilitate this process, we revisit some of the fundamental principles and concepts that have anchored research methods in the social and behavioral sciences for several decades. We hope to kindle (or rekindle) greater interest in methodological issues by not only showcasing the myriad research methods available to evolutionary scientists, but also by clarifying how different research methods and measures can be used to make clearer and stronger tests of various evolutionary-based predictions.

The chapter contains several overarching themes. One theme is that, in order 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. As we shall see, 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 all research contexts because the use of different methods and measures requires making trade-offs between maximizing internal validity, external validity, and the generalizability of findings across people. Both methodological triangulation within programs of research (i.e., adopting a multiple-method/multiple-measure approach when testing predicted effects) and the testing of alternative models are needed to make strong, clear inferences.

A second theme is that there has been an over-reliance on certain research methods (e.g., correlational methods) and certain measures (e.g., self-reports) in some quarters of the evolutionary sciences. In some cases, this mono-method/mono-measure predilection has impeded the rigorous testing of certain evolutionary-based phenomena; in others, it has not allowed investigators to determine whether the results predicted by evolutionary theories fit observed data better than those predicted by alternate, competing theories. This over-reliance can be remedied through greater knowledge and appreciation of the many strengths and advantages that multiple research methods and paradigms can offer.

A final organizing theme is the need to test and provide better evidence for the special design properties of purported psychological adaptations. In some cases, a multimethod/multimeasure approach should help researchers offer clearer and stronger evidence for the special design features of certain evolved traits or characteristics in humans. The telltale signs of selection and adaptation should be most evident when specific stimuli (triggering events) produce specific effects or outcomes (responses) at different levels of analysis (ranging from molecular to macro levels). Converging patterns of results from well-conducted multimethod/multimeasure studies will increase our confidence in the evolution of certain "specially designed" adaptations.

This chapter has four major sections: (1) we discuss how and why multimethod/multimeasure evidence for the special design of certain traits or behaviors offers more compelling evidence for their status as adaptations. We also discuss the unique inferential and methodological challenges associated with testing evolutionary theories. (2) We review classic concepts and issues surrounding validity. In doing so, we review the major types of validity and explain why evidence for each type is required to establish construct validity. (3) We re-
view different types of investigative (research) methods that are organized around an adapted model proposed by Runkel and McGrath (1972). This model highlights the trade-offs entailed by the use of each research method. (4) Finally, we showcase two programs of evolutionary research that have applied multimethod/multimeasure strategies to document the special design features of certain hypothesized psychological adaptations. We conclude by proposing several ways in which current programs of evolutionary research might be strengthened and improved from a methodological standpoint.

THEORY TESTING, SPECIAL DESIGN, AND STRONG RESEARCH METHODS

Before reviewing the methodological strategies and techniques available to evolutionary researchers, it is important to understand why it is so challenging to provide compelling empirical support for evolutionary theories. Accordingly, we open this section by explaining why it is more difficult to marshal persuasive evidence for evolutionary-based theories than for theories that do not have explicit "historical origin" components. We then discuss the concept of evolved adaptations, review criticisms of the adaptationism approach, and propose how stronger and more persuasive evidence can be mounted by testing the special design properties of purported adaptations.

WHY IS IT DIFFICULT TO PROVIDE COMPELLING EVIDENCE FOR EVOLUTIONARY THEORIES?

Evolutionary scholars frequently lament that they must provide better or more empirical evidence for their articles to survive the peer-review process. This perception probably contains some truth. Evolutionary theories have a different logical structure than most other theories, especially those that make no explicit assumptions about the distal origins of human traits, attributes, or behaviors. This difference profoundly affects the way in which evolutionary theories are viewed and evaluated by scientists. Most scientists believe that a theory is more likely to be true if (1) empirical results repeatedly confirm what the theory predicts and (2) the results cannot be explained by competing theories. Methodologically strong research programs, therefore, are structured not only around testing and confirming the specific outcomes predicted by a given theory, but also around demonstrating that outcomes have sound discriminant validity properties.

When judging the truth (veracity) of a theory, scientists usually make inferential judgments at three levels (Conway & Schaller, 2002):

1. At the most basic level, they evaluate the perceived truth of a given hypothesis, which is a product of: the perceived amount of empirical support for the hypothesis, and the inverse of the perceived plausibility of alternate explanations.

2. At the next level, scientists make inferences about the perceived veracity of the theory from which the hypothesis is derived. This assessment hinges on: how well and clearly the hypothesis and related hypotheses are logically derived from the theory and how easily they can be derived from
alternate theories or models. Inferential evaluation usually stops here because most theories focus on contemporary psychological events and processes and make few, if any, explicit assumptions about their historical origins.

3. At the most abstract level, scientists judge the plausibility of the historical origins specified by a theory. This is a function of how well and clearly models of contemporary psychological events and processes have been logically derived from their historical origin theories or models, and the degree to which these theories or models cannot be deduced from alternate ones that posit different historical origins. Hence, the more temporally "removed" empirical findings are from the theoretical models from which they are derived, the more difficult it is to convince people that a theory is true.

This extra layer of inference explains why most theories that contain large historical origin components face higher hurdles in the scientific evaluation process. Because more levels of inference must be traversed, positive results from a single study are viewed as providing less support in relation to the full theory, even if they are entirely consistent with what the theory predicts. To complicate matters, the starting assumptions about the historical antecedents of a purportedly adaptive trait or psychological mechanism (e.g., assumptions about the specific environmental problems our human ancestors faced, their social and living arrangements, the most prevalent and countervailing selection pressures) are less likely to be assumed true given that the historical antecedents cannot be directly observed, manipulated, or measured. Thus, the perceived likelihood that some other nonevolutionary theory might explain a hypothesized adaptation increases as scientists move from hypothetical, to psychological, to historical levels of inference.

Many evolutionary theories also face stiffer evaluation hurdles because of the sheer complexity—and sometimes the imprecision—of the metatheories on which they are based. Evolutionary theories tend to be more complex than other theories, including many historical origin theories that do not have evolutionary bases (e.g., certain social structuralist theories; see Eagly, 1987). One reason for this is that inferring simple associations between distal, biologically based adaptations and the operation of current psychological processes is more complicated than inferring associations between cultural or social structural factors and current psychological processes. More complex theories typically generate a larger number of "internal" alternative explanations, making 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 (Caporael & Brewer, 2000; Dawkins, 1989).

This problem has been further compounded by the relative lack of attention devoted to (1) clarifying how different middle-level evolutionary theories should

---

1 Some origin theorists occasionally can use fossils or other artifacts to make inferences about the historical origins of a given trait or behavior. This is less likely in the evolutionary sciences, where purported adaptations (e.g., domain-specific psychological mechanisms) leave sparse artifact trails (but see Andrews et al., 2003a; Williams, 1966, 1992).

2 This does not mean that these perceptions are well founded. Moreover, some theories that could or should have historical origin components fail to make such assumptions explicit, rendering these models more diffuse and difficult to test.
and should not interrelate, and (2) specifying the conditions under which different theories make similar versus divergent predictions about specific outcomes (Simpson, 1999). Evolutionary theories are hierarchically organized and contain several levels of explanation, ranging from broad metatheoretical assumptions, to domain-relevant middle-level principles, to specific hypotheses, to highly specific predictions (Buss, 1995; Ketelaar & Ellis, 2000). Most middle-level evolutionary theories (e.g., 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 biologically unrelated others. In some cases, middle-level theories generate competing hypotheses and predictions. Parental investment theory, for instance, makes different predictions than reciprocal altruism theory about 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 (Holcomb, 1998).

At times, evolutionary researchers also do not fully articulate 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 concrete hypotheses). One reason for this is that evolutionary hypotheses exist along a “continuum of confidence,” which ranges from: (1) clear and firm hypotheses that are unequivocally and directly derived from a middle-level theory; to (2) expectation-based hypotheses that can be logically deduced from a theory, but cannot be directly derived from it without making important auxiliary assumptions; to (3) speculative hypotheses based on casual and intuitive guesses (Ellis & Symons, 1990). If researchers are testing speculative hypotheses, rigorous deductively based predictions will be more difficult to derive from a theory.

**Generating More Compelling Evidence**

How can the evolutionary sciences overcome these limitations? To begin with, researchers must develop clearer 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 procured from a wider 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 others. 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 conceivable costs, constraints, and limitations—social, physical, behavioral, physiological, and otherwise—that might have counterweighted the conjectured benefits associated with a hypothesized adaptation. After conducting these analyses, researchers must elucidate why certain adaptations would have produced better solutions to specific evolutionarily relevant problems than other possible adaptations, and direct tests of alternative models must be performed.
The forgoing 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. This might be achieved by organizing more 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:
1. Articulate how and why specific selection pressures should have shaped certain psychological mechanisms or processes;
2. Identify the specific environmental cues that should have activated these processes in relevant ancestral environments;
3. Explain how these processes should have guided thoughts, feelings, and behavior in specific social situations; and
4. 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 should be 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 would be more fully explained, the theory or model being tested ought to have greater explanatory coherence. Third, sounder and more extensive deductive logic should allow researchers to derive more novel predictions. 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. Theories that predict specific types of context-dependent statistical interactions usually have fewer alternative explanations (Conway, Schaller, Tweed, & Hallett, 2001).

Adaptations, Adaptationism, and Standards of Evidence

At a conceptual level, most evolutionary psychologists subscribe to a general investigative orientation known as adaptationism. Using this approach, researchers attempt 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 EEA; Parker & Maynard Smith, 1990) to marshal evidence for certain purported 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 merely 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. 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, probably built on earlier adaptations, exaptations, or spandrels (i.e., by-products that happen to be associated with adapted traits). The evolutionary sciences, therefore, must use methodologies that are capable of documenting specific adaptations more directly (Mayr, 1983).

What types of evidence have been gathered to test whether certain traits or psychological attributes could be adaptations? Andrews, Gangestad, and Matthews (2003a) 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 should 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.3

3There are additional criteria and techniques that can aid in documenting adaptations (Andrews et al., 2003b). These include identifying powerful developmental biases or the maladaptive outcomes associated with certain injuries or disorders, recreating the phylogenetic history of a purported adaptation (to confirm that function emerged before structure), marshaling molecular genetic evidence for an adaptation or, in limited cases, showing evidence of possible homology with other higher primates.
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., 2003a). 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 and 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.

**Special Design Evidence**

Organisms are living historical documents (Cronin, 1991; Williams, 1992). Accordingly, adaptations should reveal remnants of the selective forces that shaped them. Before a trait can be classified as an adaptation, 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, etc. to rule out pure chance as an adequate explanation" (Williams, 1966, pp. 10). Chance factors can include processes such as phylogenetic legacy, genetic drift, by-product effects, and mutations, any of which could be responsible for the development of a particular trait.

Several additional factors also 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" might have been constructed), trade-offs between interacting adaptations (e.g., selection for camouflage from predators versus colorful ornamentation to attract mates), and counter-adaptations (e.g., countervailing mating tactics that emerge between the sexes in a species). 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 that are believed to be adaptations because domain-general learning processes (such as exapted learning mechanisms) can produce traits that have considerable specificity, proficiency, and complexity (see Andrews, Gangestad, & Matthews, 2003a, 2003b; Kruschke,

---

4The term adaptive trait refers to the underlying decision rules and information processing algorithms that should have been selected to guide context-specific thoughts, feelings, and behaviors. The deployment of these rules should typically have increased individuals' reproductive fitness in ancestral environments (Andrews et al., 2003a).

5We are not suggesting that a trait must be shaped only by natural selection forces in order to be considered an adaptation. Clearly, natural selection must operate on existing variation, most of which is generated by chance factors.
For these “complex traits,” further evidence for their special design properties is typically required.

Fortunately, several sources of evidence can increase our confidence about the special design of certain traits (Andrews et al., 2003a). First, “complex” trait adaptations can be documented by demonstrating that a trait is a biased outcome of a specific developmental or learning mechanism (Cummins & Cummins, 1999). These traits develop or are learned very easily, quickly, and reliably, and they tend to solve specific adaptive problems with 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; Ohman & 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 deleterious in modern-day environments. One example is the dire 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 adaptive 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; Gangestad, Thornhill, & Garver-Apgar, Chapter 11, this volume).

There are some drawbacks to using special design as the sole evidentiary criteria for adaptations. For one, it might be difficult to provide unambiguous evidence for the special design features of certain adaptations. To guard against this possibility, investigators should not only test for the special design features of specific traits, but should provide some evidence for the other standards as well. Adaptations might also be difficult to identify because many “complex” traits may possess mixed design (e.g., female orgasm, the development of the neocortex; see Andrews et al., 2003a). If, for example, 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 unless a fine-grained analysis of its design features was performed.

**ISSUES OF VALIDITY**

Validity is generally defined as “the best available approximation to the truth or falsity of propositions” (T. D. Cook & Campbell, 1979, pp. 37). Therefore, validity reflects the degree of truth regarding the statements, inferences, or conclusions drawn from empirical research. Since research programs have different missions, the validity of a given study must be evaluated in the context of the larger goals, purposes, and objectives of a research program.
GOALS AND OBJECTIVES OF RESEARCH

Most studies or programs of research are designed to achieve one of three general objectives: (1) to demonstrate the existence of a hypothesized effect or association, (2) to provide evidence about what causes or produces an effect, and (3) to explain the intervening psychological processes or mechanisms that mediate or moderate a causal link (Brewer, 2000). Research designed to demonstrate a phenomenon simply attempts to document a predicted association or effect. Demonstration research often is descriptive, such as revealing the frequency with which an event or behavior occurs or demonstrating the strength of association between two variables. Although many demonstration studies are staged in field settings, they can also be conducted in the laboratory.

A second major objective of research is to establish cause-effect relations between specific variables (e.g., if independent variable X is manipulated in a certain way, dependent variable Y should then change in predictable ways). This focus on causation reflects the “utilitarian” view of causal processes (T. D. Cook & Campbell, 1979). According to this approach, reliable causal associations between variables are confirmed, but little if any attention is devoted to explaining how they are generated or why they occur. A third fundamental objective of research—one that guides the thinking of most evolutionary scientists—is explanation (i.e., to understand intervening processes that mediate or moderate and, thus, explain how and why X causes Y). This “essentialist” view of causation (T. D. Cook & Campbell, 1979) is integral to theory testing, which usually is concerned with identifying the conditions under which causal relations do and do not exist.

A PROCESS MODEL OF VALIDITY

The procedures for establishing the validity of an operationalization or measure of a construct are similar to the procedures and rules for developing, testing, and confirming scientific theories (see Loevinger, 1957). Different types of validity offer unique sources of evidence about the general construct validity of a given scale, measure, or psychological process. Because 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 produce revised constructs, which then suggest new and improved operations, and so on.

Two methodological traditions have influenced the way in which validity is defined and conceptualized. One tradition, grounded in experimental and quasi-experimental research, has focused primarily on the validity of independent variables, particularly their conceptualization, their operationalization, and how they are perceived by research participants (T. D. Cook & Campbell, 1979). A second tradition, emanating from nonexperimental 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) has proposed a three-stage process model that describes how hypothetical theoretical constructs are conceptually linked with three sets of measures: (1) observable stimuli (independent variables), (2) intervening physiological or cognitive processes (those that occur within individuals), and (3) observable responses (dependent or outcome variables). As shown
in Figure 4.1, researchers must make three *inferential connections* when planning and conducting studies. 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. Especially 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 chosen as potential mediating variables. On the dependent variable side, they must derive clear inferential connections between the effects anticipated by their theory and the responses (outcomes) measured in their study. Assorted problems can undermine valid inferences from a study at any of these stages. To complicate matters, many areas of evolutionary science lack standardized measures, operations, or procedures that correspond closely with the latent theoretical constructs of interest. As a result, evolutionary scientists must frequently make large inferential leaps across each set of linkages.

These difficulties can create thorny methodological problems. For example, the validity of stimulus or response measures might be questioned 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 tap. It also may be difficult to predict the precise levels at which certain independent variables should (or should not) exert causal effects on specific outcome measures. And it might be challenging to anticipate the range over which certain independent variables should have their strongest effects on specific outcome measures (Rakover, 1981). Given the multitude of ways in which the validity of a study can be reduced, it is often difficult to know whether null results from a
single study reflect a failure of the theory, a failure of the operationalizations at one or more of Brewer’s (2000) three stages, or a failure of the measures used.

**Validity in Experimental and Quasi-Experimental Research**

There are four major types of validity of primary concern in experimental and quasi-experimental research (T. D. 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 some 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 \(X\) (that is, that the independent variable had a causal impact on the dependent variable, independent of other possible causal factors). If third variables correlate with \(X\), these confounds could 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 independent variables. Researchers must also be cognizant of the many other factors that can threaten the internal validity of a study (see T. D. Cook & Campbell, 1979).

Moderating and mediating variables can complicate causal inferences (Baron & Kenny, 1986). Moderating effects exist when there is a true causal association between an independent variable \((X)\) and a dependent variable \((Y)\), but this relation varies at different levels of a third variable \((C)\). Evolutionary scientists, for instance, might posit that an experimental manipulation of high versus low impending physical threat should lead most highly threatened individuals to stand and defend themselves. However, this association could be moderated by gender, with men being more likely to adopt the "stand and defend" response under high threat than women.

Mediating effects, by comparison, 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 in turn predicts changes in the dependent variable \((Y)\), statistically controlling for \(X\). Returning to the earlier example, evolutionary scientists might also postulate that a high level of impending physical threat should lead most men to experience “challenge” physiological responses that prepare them to stand and defend. Conversely, such threats might lead most women to experience “threat” physiological responses, motivating 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 several factors, such as 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 in fact 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 (increasing the variance or meaning of treatments/conditions), or if respondents are heterogeneous on one or more dimensions that could affect how they interpret the meaning of treatments, independent variables, or outcome measures.

A third major form of validity, known as external validity, entails the degree to which a researcher can generalize from a study: (1) to particular target persons or settings or (2) across different persons, settings, and times. Researchers typically are interested in the latter form of external validity. The external validity of a study can be assessed by testing for statistical interactions, that is, whether an effect holds across different persons, settings, or times. Conducting many small, heterogeneous studies rather than a small number of large-scale ones enhances external validity. It 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 involves the extent to which an effect occurs under conditions that are “typical” or “common” for a given population. This form of external validity is most relevant to research that has descriptive goals. Relevance reflects the degree to which findings are useful or applicable in solving social problems or improving the quality of life. Robustness has the greatest implications for evolutionary research. It reflects the degree to which a finding is replicable across different settings, people, and historical contexts.

To evaluate robustness, theorists must clearly define the populations and settings to which an effect should or should not generalize. In the evolutionary sciences, generalizability from one prototypical subject 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 gatherers in our ancestral past) is one of the most important external validity concerns. Although evolutionary scientists are now conducting more cross-cultural research, some areas still have fairly narrow databases from which to draw general inferences about human evolution. Similar issues plague other fields. The results of many social psychological studies, for example, must be qualified by the attributes of typical research participants (i.e., fairly intelligent, well-educated, young individuals who have fluid attitudes and self-concepts and emerging group identities; Sears, 1986). Evolutionary scientists should continue to articulate the cardinal ways in which contemporary participant populations may differ from more traditional hunter/gatherer “target” populations and how these differences might qualify how certain evolutionary studies are interpreted.

The fourth type of validity—construct validity—is the most general and 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 (see 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)?"

Many independent variables are complex packages of multiple, sometimes correlated variables. When an experimenter tries to induce social isolation in participants, for example, the manipulation may generate other unanticipated states in individuals, such as heightened anxiety, depressive symptoms, or negative moods. Many of the concerns about construct validity, therefore, center on how independent variables are (or should be) operationalized in particular studies and how they are perceived by participants. An experimental manipulation might also evoke multiple hypothetical states in the same individual. If this occurs, it can be nearly impossible to identify the specific causal agent that is operative in a study. For this reason, T. D. Cook and Campbell (1979) contend 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 in order 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 with the multitrait-multimethod matrix approach (Campbell & Fiske, 1959). Measures contain 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. Studies can be partitioned into one of four categories: (1) monotrait-monomethod (when a single trait/scale is studied using a single 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 a single method), or (4) heterotrait-heteromethod (when multiple traits/scales are studied using multiple methods). Heterotrait-heteromethod approaches are preferable because they allow investigators to test for both the convergent and the discriminant validation properties of traits/scales. Strong evidence for convergent validity emerges when a trait/scale correlates with measures that tap theoretically similar constructs, even when the trait/scale is measured with different methods (e.g., life-event data, observational data, behavioral test situation data, self-report data; Cattell, 1972). 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.

Multitrait-multimethod techniques are not used nearly as much as they should be. One particularly pervasive problem is the over-reliance on monomethod data, especially purely self-report data. As Campbell and Fiske (1959) lament, constructs that are measured with the same method correlate more highly than when they are measured with different methods due to shared method variance. Some reported "effects" in monomethod studies, therefore, may attain conventional levels of statistical significance due to the added boost from shared method
variance. Needless to say, this reduces the validity of the conclusions reached from the study.

INVESTIGATIVE METHODS

Different investigative methods (research strategies) can be classified based on the procedures and techniques that are used to describe behavior and test hypotheses. One particularly elegant classification system has been developed by Runkel and McGrath (1972; see also McGrath, 1982). They propose that there are eight higher level research strategies, each of which resides within a circumplex model (see Figure 4.2). The eight strategies fall into four quadrants, with each quadrant containing two similar strategies. The strategies are structured around two orthogonal axes: (1) the degree to which each strategy uses obtrusive versus unobtrusive procedures, and (2) the degree to which each one contains particular versus universal behavioral systems. Each strategy tends to maximize one of three mutually conflicting research goals, which are labeled A, B, and C. Ideally, researchers yearn to maximize the generalizability of results across populations.

Figure 4.2 Research Strategies. A = Point of maximum concern with generality across actors; B = Point of maximum concern with precision of measurement; and C = Point of maximum concern with realism of the context. Source: From Research on Human Behavior: A Systematic Guide to Method (1972) by P. J. Runkel and J. E. McGrath (Figure 4-1, p. 85), New York: Holt, Rinehart, and Winston.
(A), the precision with which variables are measured (B), and the realism of the context in which variables are assessed (C). As shown in Figure 4.2, however, all three goals cannot be simultaneously maximized by any single research strategy. In trying to maximize one goal, the other goals are compromised. Because different research strategies come with particular advantages and liabilities, investigators must develop research programs that take advantage of the special strengths and qualities inherent in each strategy. In the following discussion, we give examples of how human mating behavior has been studied using different investigative methods.

**Types of Strategies**

The eight research strategies fall into four quadrants. The two strategies in Quadrant I involve situations that are real for research participants. Data is collected in natural settings from people who typically are not aware that they are being observed. Researchers may observe people in social contexts that are manipulated by the experimenter (in field experiments) or that naturally occur (in field studies). Field experiments, therefore, involve some form of experimental manipulation in which the researcher systematically alters the social context and then observes participants' behavior. Field studies, in contrast, do not have experimental manipulations and tend to be less obtrusive. In both cases, participants have usually chosen to be in the research setting. A major strength of these Quadrant I strategies is that they maximize mundane realism (the degree to which the research setting and operations resemble daily life events).

Clark and Hatfield (1989) provide an excellent example of a field experiment. They tested for sex differences in receptivity to sexual offers made by unknown opposite-sex individuals. More specifically, they hypothesized that men should be more willing than women to accept a sexual invitation from a total stranger. Male and female experimental accomplices of average attractiveness approached participants in a public setting and asked them one of three questions: "Would you like to go out tonight?", "Will you come over to my apartment?", or "Would you go to bed with me?." Not a single female participant accepted the offer of immediate sex; over 70% of men did. The advantages of using this methodology were that participants did not initially know they were in a study and they thought their decisions had real and immediate consequences.

Most field studies, on the other hand, exert minimal influence on participants while observing their behavior. Givens (1983) and Perper (1985), for example, have studied courtship behaviors in more naturalistic settings (e.g., bars). They have discovered five distinct phases of courting behavior: attention getting, recognition, talking, touching, and keeping time. Many of these behaviors are quite similar to courtship behaviors observed using similar unobtrusive observational methods in hunter/gatherer tribes (Eibl-Eibesfeldt, 1989). By observing people in natural settings, important patterns of behavior can be identified that might be difficult or impossible to witness in the laboratory.

There are drawbacks to field-based methods. These methods typically have less precision of measurement and control (B), and it is more difficult to generalize results across disparate populations (A). For example, because they did not randomly select participants from all possible populations, Clark and Hatfield could make inferences about only those people who entered their public setting.
Individuals who frequent public places may differ in systematic ways from those who do not, which might explain some of the variation in the huge gender differences they found.

Quadrant II strategies are defined by a high degree of experimenter control. They differ from Quadrant I strategies in that the situations to which participants are exposed are more removed from routine, everyday experiences. The main focus of Quadrant II strategies is on how different social or situational contexts systematically influence participants’ thoughts, feelings, and behavior. Participants are usually randomly assigned to different experimental conditions. All procedures are identical across the experimental conditions except for the critical manipulation(s), which are introduced and tested to determine whether they have a causal impact on predicted outcomes. In a repeated measures experimental design, each participant is exposed to every experimental condition, and comparisons are then made within participants across conditions to ascertain the effect of the independent variables on the dependent measures. Experiments can also be designed to place participants in more than one context, exposing them to more than one level of a second independent variable (mixed factorial designs). Any differences that emerge either between participants placed in different experimental conditions or within participants exposed to more than one condition can be attributed to the manipulations. The high precision of measurement coupled with the ability to draw clear cause-effect conclusions is the main advantage of experimental methods. The principle difference between the two strategies in this quadrant is that experimental simulations retain some degree of mundane realism.

Evolutionary scientists are conducting more laboratory experiments to test assorted predictions. Roney (2003), for example, has recently examined on how men view and describe themselves after being exposed to highly attractive women. Men were randomly assigned to review advertisements featuring either young, attractive women or older, less attractive women. They then completed questionnaires that assessed how important having a large income and being financially successful was to them, how ambitious they felt at that moment, and which traits best described their personality. This study contained two levels (young models versus old models) of a single independent variable (model attractiveness). The men in each experimental condition received the same instructions; the only difference was the advertisements they saw. Thus, any differences in the outcome measures between the two groups must be attributed to the different advertisements. As predicted, after being exposed to young, attractive women, men rated financial concerns as more important to them, felt more ambitious, and claimed that traits related to extraversion were more descriptive of their personality—all characteristics associated with increased status, a feature that most women find appealing in prospective mates (Buss, 1989).

Although carefully conducted experiments maximize precision of measurement (B), they do not maximize generalizability across populations (A). Most experiments involve small samples of participants who are recruited from academic settings or reside in a relatively narrow geo-cultural context. As a result, it is difficult to make broad generalizations about universal psychological mechanisms from just a few experiments, even if they are well controlled. Furthermore, most experiments suffer from low mundane realism (C) because the social context is contrived. Most people rarely find themselves in many of the situations created
by experimenters, and the extraneous variables that experiments try to control are often present outside the laboratory.

Quadrant III strategies are designed to minimize the influence of the social context by controlling extraneous variables (in judgment tasks) or by asking questions that are unrelated to the context in which they are assessed (in sample surveys). In judgment studies, participants' responses describe or evaluate a stimulus (instead of responding to it). In sample survey studies, participants respond to self-descriptive questions.

Singh's (1993) research on waist-to-hip ratio (WHR) is a good exemplar of an evolutionary-relevant judgment task. Singh contends that body fat distribution should be a marker of female attractiveness to the extent that body shape conveys critical information about a woman's fertility and youth. Hence, participants (judges) rated the attractiveness and health of women with different WHRs. Images near the .70 range were, in fact, rated more attractive and healthy than those that deviated from this evolutionary "ideal" value.

Buss (1989) has conducted research that exemplifies the survey methodology. He hypothesized that men and women should covet slightly different attributes in long-term mates, with men placing more emphasis on youth and beauty (cues of fertility) and women valuing status and resources (cues of provisioning ability). This hypothesis was tested in 37 cultures. Participants were asked a series of self-report questions about their personal preferences for a long-term mate. In every culture, men rated youth and beauty in long-term mates more highly than women did, whereas women rated long-term mates' status and resources more highly than men did.

One relatively new survey approach involves collecting self-report responses from participants across time (e.g., event sampling or diary studies). This approach is rapidly gaining favor among evolutionary researchers (e.g., Keller, Nesse, & Hofferth, 2001; Shackelford & Larsen, 1999). Another survey-based methodological strategy, known as meta-analysis (Cooper & Hedges, 1994), involves reviewing all empirical studies that test a similar set of hypotheses and pooling the results to calculate representative effect sizes. Valuable meta-analyses have been conducted by evolutionary scientists (e.g., Feingold, 1992; West & Sheldon, 2002). Moreover, in their classic analysis of homicide rates, Daly and Wilson (1988) have shown the usefulness of surveying public records to test evolutionary hypotheses. The study of birth order effects (Sulloway, 1996) and social prominence (Simonton, 1994) have also benefited greatly from surveying archival records. One strength of these Quadrant III strategies is that research findings are more generalizable, especially when researchers assess large samples from diverse geographical and cultural regions. Two limitations of Quadrant III strategies are their lack realism of context coupled with the fact that measurement control/precision is attenuated.

Quadrant IV strategies are theoretical rather than empirical strategies. Formal theory includes efforts to develop general theories of behavior. Computer simulations, in contrast, involve attempts to model specific behavior systems, given a particular set of theoretical constraints or parameters. Formal theory usually attempts to describe universal behavior systems and, accordingly, seeks to reach broad generalizations across different populations of people. Computer simulations are also designed to generalize across populations while manipulating different contexts, which often are prescribed by the theoretical parameters built into the models being tested.
The evolutionary sciences are rich in formal theory. Hamilton (1964), for example, introduced the concept of inclusive fitness and demonstrated how helping behaviors are governed by the degree of genetic relatedness between members of the same species. Trivers (1971) introduced the theory of reciprocal altruism to explain helping behaviors between genetically unrelated members of the same species and, in some cases, between species. Trivers (1972) also developed parental investment theory to explain certain gender differences in mating behavior. More recently, Tooby and Cosmides (1992) and Pinker (2002) have outlined the limitations of viewing the mind as a general learning device, and have proposed that the mind is equipped with domain-specific psychological mechanisms that become activated and guide thoughts, feelings, and action in circumscribed social contexts. Each of these theories posits that certain psychological processes are universal (generalizable across populations).

Kenrick and his colleagues (Kenrick, Li, & Butner, 2003; Kenrick et al., 2002) provide excellent examples of how computer simulations can be used to model and test evolutionary hypotheses. Blending dynamical systems theory with evolutionary principles, they have shown how attitudes toward restricted versus unrestricted mating orientations in women and men can be influenced by the attitudes of nearby others over time. These simulations are important because they demonstrate that different outcomes are possible when very minor changes occur in local environments.

**Research Programs Providing Evidence for Psychological Adaptations**

Different traits are likely to require different types of evidence to reveal their special design properties. Nevertheless, certain methodological strategies can facilitate the documentation of special design. The special design features of specific traits can be tested more rigorously by conducting research that: (1) uses multiple methods and multiple measures to assess and triangulate the major constructs, (2) tests for and systematically discounts alternative explanations for a trait's uniquely designed functional features, and (3) reveals the footprints of special design at different levels of analysis (ranging from neural mechanisms, to context-specific modes of information processing, to emotional reactions, to molar behavioral responses; see Wilson, 1998). Some research programs have begun to document the special design properties of certain hypothesized psychological adaptations. Select 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 next.

**Snakes and an Evolved Fear Module**

Ohman, Mineka, and their colleagues have offered strong, programmatic, and compelling evidence that humans and closely related primates possess an evolved “fear module” for reptiles (Ohman & Mineka, 2001, 2003). 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 different levels of analysis (ranging from neural mechanisms to general cognitive expectations and behavioral reactions).

Several interlocking lines of evidence clearly point to an evolved fear module in higher primates (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 that probably has distant evolutionary origins. Conducting well-designed experiments, researchers have also demonstrated that lab-raised monkeys learn to fear snakes very quickly merely by observing fearful expressions in other monkeys (M. Cook & Mineka, 1990), lab-raised monkeys show preferential conditioning to toy reptiles but not to innocuous stimuli (e.g., toy rabbits; M. 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 nonaversive stimuli (e.g., flowers; Öhman & Mineka, 2001). This body of evidence implies that the strong connection between snakes and aversive unconditioned stimuli most likely emanates from the evolutionary history of primates rather than from culturally mediated conditioning processes.

Further lab experiments have shown that humans readily infer illusory associations between snakes and aversive stimuli. For example, individuals are more likely to perceive that fearful stimuli (snakes) co-occur with painful experiences (shocks) than is true of other nonfearful stimuli, even when there is no covariation between pairings of shock and different stimuli (Tomarken, Sutton, & Mineka, 1995). Individuals also believe that shocks are more likely to follow exposure to dangerous stimuli (snakes and damaged electrical equipment), yet illusory correlations emerge only between snakes and shock after individuals have been exposed to a random series of stimulus/shock trials (Kennedy, Rapee, & Mazurski, 1997). Recent experiments assessing visual detection latencies have found 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.

Recent 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) have discovered that fear responses can be learned and activated, even when backward masking prevents images of snakes from reaching higher cortical processing. This evidence indicates that these fear responses may reside in ancient neural circuits that evolved well before 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 very specific class of stimuli, it operates in spe-
cific areas of the brain (the amygdala) that evolved before the neocortex, and it appears to have 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.

MATE PREFERENCES IN WOMEN ACROSS THE REPRODUCTIVE CYCLE

Gangestad, Thornhill, and their colleagues have conducted a series of well-conceptualized and carefully designed studies to test whether women have a psychological adaptation that leads them to prefer certain types of men as short-term mates during certain phases of their reproductive cycles. This line of work is elegant because the predictions are carefully derived from theoretical models (good genes sexual selection) and cross-species data, the predictions are quite specific (involving specific patterns of statistical interactions), the predictions and results are difficult to derive from competing theories or models, and logical alternative explanations have been systematically ruled out. Because Gangestad, Thornhill, and Garver-Apgar (Chapter 11, this volume) discuss this research in detail, we highlight only a few of these findings.

The Strategic Pluralism Model of mating (Gangestad & Simpson, 2000) proposes that women should have evolved to make trade-offs between two sets of attributes when evaluating men as potential mates: men’s degree of general health/viability, and their degree of commitment/investment to the relationship and possible offspring. Fluctuating symmetry (FA: the extent to which individuals are bilaterally symmetrical at different locations of the body) is believed to be one possible marker of viability (see Gangestad & Simpson, 2000). If so, women should find more symmetrical men more attractive than less symmetrical men in short-term mating contexts, especially when they are ovulating (and, therefore, could conceivably transmit the “good genes” of these more viable men to their offspring). Thus, this model predicts very specific patterns of statistical interactions, predictions that cannot be easily derived a priori from alternative perspectives.

This hypothesis has been tested using a variety of research methods and techniques. Self-report questionnaire studies have confirmed that more symmetrical men are more attractive than less symmetrical men in short-term mating contexts, especially when they are ovulating (and, therefore, could conceivably transmit the “good genes” of these more viable men to their offspring). Thus, this model predicts very specific patterns of statistical interactions, predictions that cannot be easily derived a priori from alternative perspectives.

To test predictions about olfactory markers of men’s FA and women’s reproductive cycle, Gangestad and Thornhill (1998) had women smell unscented T-shirts worn by men who differed in FA. If women were ovulating during the study, they rated the scents of more symmetrical men as more attractive than the
scents of less symmetrical but, as predicted, this interaction effect was not found in nonovulating women. Providing strong discriminant validity evidence for this effect, Thornhill et al. (2001) have confirmed 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 does not increase when women are 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 man as both a short-term and a long-term mate, women who were ovulating were 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, in press). Considered together, these findings confirm that women’s mate preferences vary across the reproductive cycle in very specific and theoretically consistent ways.

CONCLUSIONS

The methodology of research in the evolutionary sciences can be strengthened in several ways:

- When feasible, researchers should use a wider range of research methods in their ongoing programs of work. In particular, more research programs need to be structured around experimental methods and techniques.
- A wider array of measurement and statistical techniques should be utilized.
- More solid evidence needs to be provided for the validity of major manipulations, scales, and individual-item measures before they are adopted by other researchers (e.g., experimental manipulations of “social status,” self-report measures of “mate value”).
- Greater attention should focus on deducing, modeling, and testing the properties of psychological mechanisms that are believed to be adaptations.
- Stronger and better evidence is needed to ascertain how well outcomes predicted by different evolutionary theories or models fit different data sets, especially in relation to competing nonevolutionary theories or models. Whenever possible, alternative constructs and explanations should be carefully derived and measured to test—and hopefully discount—competing constructs or models.
- The special design features of purported adaptations must be directly specified and tested at different levels of analysis, ranging from possible neural structures in the brain, to information processing biases or modal tendencies, to physiological responses, to covert thoughts and emotional reactions, to overt behavioral responses.
• Evidence for possible adaptations needs to be gathered for multiple evidentiary standards.
• Empirical evidence for specific hypotheses should be provided across different cultures, especially those that are more similar to the environments in which humans probably evolved.
• More effort must be devoted to deriving and testing novel predictions, particularly predictions that cannot be easily derived or explained by competing theories.

In conclusion, evolutionary scientists must emulate the methodological breadth and creativity of Darwin. This can be accomplished by using a wider assortment of research methods and statistical techniques, many of which will help investigators more clearly map out and understand the architecture of the human mind. To convince the larger scientific community of the value and the predictive, explanatory, and integrative power of evolutionary approaches, evolutionary theories and models must be developed more carefully, derived more precisely, and tested more thoroughly than most other theories. Given their tremendous explanatory and integrative power, evolutionary theories have at times proceeded ahead of solid empirical evidence, especially in the case of humans. Recent advances in research methods can close this gap. Evolutionary researchers, however, must sharpen the deductive logic underlying their theoretical models, revise and refine questionable or conflicting tenets of middle-level theories, discard or revamp problematic hypotheses, and formulate more specific and refined hypotheses that more directly test the special design properties of purported adaptations. If these goals are realized, the evolutionary sciences will experience rapid and significant theoretical and empirical advances in the coming years.

REFERENCES


