Journal of Social, Evolutionary, and Cultural Psychology
2012, 6(1), 1-12.

Original Article

HOW OBVIOUS ARE HYPOTHESES IN EVOLUTIONARY PSYCHOLOGY?

David Trafimow
Department of Psychology, New Mexico State University

Daniel Gambacorta
Department of Psychology, New Mexico State University

Abstract
Evolutionary psychology critics often have accused evolutionary psychology of being unfalsifiable, whereas evolutionary psychology aficionados have responded that it is no more unfalsifiable than are other areas of psychology. The arguments on both sides largely have been at the philosophical level. However, a careful analysis of the notion of falsification implies the possibility of empirical tests of falsification claims centered on the issue of whether the hypotheses are or are not obvious. We present two empirical tests, each carried out with (presumably) less informed (undergraduate students) or more informed (graduate students) samples. The findings strongly support that at least some evolutionary psychology hypotheses are not obvious, thereby rendering them as potentially destructive tests of evolutionary psychology claims. We also tested undergraduate students on their reactions to highly cited evolutionary hypotheses in Studies 3 and 4. These highly cited hypotheses were neither extremely obvious nor extremely surprising. We conclude that although it is wrong to say that all evolutionary psychology hypotheses are obvious, there is room for improvement.

Keywords: Evolutionary psychology, falsification, auxiliary assumptions, obvious hypotheses, philosophy of science

Introduction

There has been considerable debate about whether evolutionary psychology is falsifiable, with critics providing reasons why they believe it is not (Gould, 1997; Kitcher 1985; Schlinger, 1996; Gannon, 2002) and apologists providing reasons why evolutionary psychology is just as falsifiable as anything else in psychology (Ketelaar & Ellis, 2000; Confer et al., 2010; Conway & Schaller, 2002; Schmitt, 2002; Sesardic, 2003). We eventually will argue that what seems like a philosophical issue—the falsifiability of an area of psychology—largely can be settled empirically. However, to

AUTHOR NOTE: Please direct correspondence to David Trafimow, Department of Psychology, MSC 3452, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003-8001. E-mail: dtrafimo@nmsu.edu

see why this is so, it is necessary to understand the falsification issue with some precision.

Popper (1959; 1962; 1972; 1983; also see Ackermann, 1976; Agassi, 2008) considered the issue by pointing out a basic asymmetry between verification and falsification. To see the asymmetry, suppose that a researcher derives a hypothesis from a theory, supports it, and concludes that the theory is true. This reasoning is an example of the fallacy of affirming the consequent; the fact that the finding supported the hypothesis does not logically necessitate the truth of the theory because the finding could have occurred for some reason other than the truth of the theory. In contrast, had the finding failed to support the hypothesis, the theory would be disconfirmed by the logic of *modus tollens*. In principle (not considering auxiliary assumptions), although a single wrong prediction can disconfirm a theory, even a million confirmations cannot prove the theory to be true. Because Popper thought it was better to use valid than invalid logic, he suggested that scientists should attempt to falsify rather than verify their theories. However, falsification only makes sense if the theory under investigation actually is capable of being falsified and so a basic prerequisite of a good theory is that it must be capable of being falsified – it must be *falsifiable*.

Given that a good theory must be falsifiable, what makes it so? According to Popper, the theory must generate strong hypotheses – that is, hypotheses that would be unlikely to be true if the theory is not true. Popper's idea was to subject the theory to the most potentially destructive tests possible, and if the theory survives those destructive tests, then it is "corroborated," though not proven. If a theory does not generate strong hypotheses, then it cannot be subjected to destructive tests, which is an undesirable state of affairs. So given that a theory is not falsifiable unless it generates strong hypotheses that can lead to potentially destructive tests, there is the issue of what makes a hypothesis "strong." Popper did not provide a mathematical definition but did provide a verbal one; hypotheses are strong if they are unlikely to be true if the theory from which they were derived is not true (Popper, 1959; 1972; 1983 also see Meehl, 1990; Platt, 1964; Roberts & Pashler, 2000). Strong hypotheses provide "severe" tests of theories whereas obvious hypotheses do not (see especially, Popper, 1962, pp. 242-245). In support of his argument, Popper provided positive examples from the history of science where confirmation of hypotheses that were not obvious (and might even be considered to be shocking) provided strong corroboration for theories such as Einstein's theory of relativity and he also provided negative examples where theories that only made obvious predictions failed to generate impressive research.

Philosophers and psychologists have long known that falsification is not dichotomous, but rather is continuous; depending on a variety of factors, empirical findings can provide a greater or lesser degree of corroboration or falsification. An implication is that theories should not be considered as falsifiable or not falsifiable, but as more falsifiable or less falsifiable (e.g., Duhem, 1954; Lakatos, 1970; 1978; Meehl, 1990; Quine, 1980; Trafimow, 2003). In turn, this implies the existence of factors underlying falsifiability, at least one of which is not dichotomous. Popper was well aware of the issue and devoted effort to arguing that an important underlying factor is the obviousness of the hypotheses that one can derive from the theory. If a theory only suggests obvious hypotheses, then (a) it is unlikely that the hypotheses will be found to be wrong and (b) supporting the hypotheses fails to provide much support that the theory is able to pass potentially destructive tests. As a dramatic example, if a theory predicts that the 6:00 news will come on at approximately 6:00, then (a) the hypothesis is unlikely to be wrong

and (b) finding that the hypothesis is correct fails to provide impressive support for the theory. In contrast, a hypothesis that is unlikely to be true if the theory is not true (a) is more likely to be wrong and (b) if found to be correct, provides relatively strong corroboration (not proof) for the theory. For example, when Einstein's hypothesis involving the discrepancy between the apparent and real position of a star measured during an eclipse was confirmed, it provided relatively strong corroborative support for the theory of relativity because physicists judged that the hypothesis would be unlikely to be true if the theory were not true.

Interestingly, the importance of the conditional probability of the hypothesis given that the theory is not true can be supported via Bayes's theorem. Trafimow (2003) provided a mathematical demonstration that as this conditional probability increases, confirmation of the hypothesis has less of a positive impact on the posterior probability of the theory (and can even decrease the posterior probability of the theory under special conditions); conversely, as this conditional probability decreases, confirmation of the hypothesis more positively impacts the posterior probability of the theory. In simpler words, experimentally confirming obvious hypotheses fails to provide impressive support for the theory whereas confirming non-obvious or surprising hypotheses does so. Therefore, for the best possible tests of theories, researchers should prefer non-obvious hypotheses to obvious ones. Researchers who test hypotheses that are not obvious are behaving in accordance with the spirit of Popper's philosophy of falsification whereas researchers who test obvious hypotheses are in violation of this spirit. We add, parenthetically, that even from alternative philosophical perspectives (e.g., verification, Bayesian, and others), empirical confirmation of obvious hypotheses is less compelling than empirical confirmation of non-obvious ones (Trafimow, 2003).

Because whether hypotheses are or are not obvious is so important from a variety of philosophical perspectives, how can one determine the obviousness of hypotheses? We see at least two ways though others doubtless exist. First, people could be presented with hypotheses and asked to respond whether they are obvious. A potential problem with this method is that it might be subject to hindsight bias. Secondly, people could be presented with both the hypotheses and opposing ones and asked to choose which they believe was supported by data. To the extent that people make the right choice, the hypothesis clearly is obvious whereas to the extent that people make the wrong choice, it is less obvious.

To relate this discussion back to the issue of whether evolutionary psychology only predicts obvious hypotheses, we can propose two possibilities. If evolutionary psychology researchers only propose obvious hypotheses, participants will (a) state that the hypotheses are obvious when presented with them and (b) guess correctly when choosing between evolutionary psychology hypotheses versus opposing ones. Alternatively, if at least some evolutionary psychologists are proposing hypotheses that are not obvious, participants will (a) state that they are not obvious and (b) fail to guess correctly on a substantial number of cases when choosing between evolutionary psychology hypotheses versus opposing ones. Four studies were conducted to test these possibilities. Studies 1 and 2 address whether there are any non-obvious evolutionary psychology hypotheses. Studies 3 and 4 address whether evolutionary psychology hypotheses are typically non-obvious.

Study 1 Method

There were two versions of Study 1. Undergraduate and graduate students rated the obviousness of evolutionary psychology hypotheses. In addition, we included some hypotheses that we were confident were obvious but we also included some that we thought might not be obvious. Including both kinds of hypotheses allowed us to make comparisons between them. In addition, by defining the two groups of hypotheses on an *a priori* basis, we avoided capitalizing on chance which would have been a risk had we formed the groups *a posteriori*.

Participants

Twenty-one undergraduate psychology students and 10 graduate psychology students participated in Study 1. The undergraduate students received course credit for their participation.

Procedure

Participants were presented with ten evolutionary psychology hypotheses and instructed to rate each one on an "obviousness" scale, though without any reference to "evolutionary psychology." Participants rated the findings on a seven-point scale with -3 representing "extremely obvious," 0 representing the neutral point, and 3 representing "extremely surprising."

The researchers judged five of the hypotheses as obvious and five of them as less obvious. An example of an obvious item is "Men typically desire a greater number of sex partners than women do" (Buss & Schmitt, 1993). An example of a less obvious item is "Men are better than women at navigating through strange territories" (Silverman et al., 2000). A final question asked participants whether they were previously familiar with any of the hypotheses, and if so, to list them. Finally, all participants were debriefed.

Study 1 Results

After completing this study we were made aware that the researchers had retracted one of the findings that made up one of the "surprising" items (Pollet & Nettle, 2010). Data for this item was removed from our analyses.

Consistent with expectations, undergraduate students rated the obvious hypotheses as more obvious than the other less obvious hypotheses (M = -1.74 and M = .43, $t(20) = -9.53$, $p < .001$, $d = -2.08$). The graduate students also responded with similar ratings (M = -1.74 and M = .45, $t(9) = -7.46$, $p < .001$, $d = -2.36$). Thus, what we had considered to be less obvious hypotheses really were rated as more surprising than the obvious hypotheses were rated.

The more important issue, however, concerns whether the less obvious hypotheses really were not obvious. To test this, we compared undergraduate and graduate students' ratings against the neutral point of zero and found that both sets of ratings were significantly positive (both p values < .05), thereby indicating that the participants found these hypotheses to not be obvious in an absolute sense as well as relative to the obvious hypotheses.

We repeated all of the foregoing analyses but excluding responses to hypotheses with which participants had claimed prior familiarity. These analyses resulted in similar findings.

Study 2 Method

Researchers sometimes complain that participants' ratings cannot be trusted. Consequently, Study 2 involved presenting participants with hypotheses from evolutionary psychology and opposing hypotheses and asking them to choose, out of each pair, the correct hypothesis. As in Study 1, we used hypotheses that we were confident were obvious or not obvious.

Participants

New and independent samples of 15 undergraduate psychology students and 9 graduate psychology students participated in Study 2. The undergraduate students received course credit for their participation.

Procedure

Participants were presented with ten pairs of hypotheses. One hypothesis in each pair was from the evolutionary psychology literature and was used in Study 1 whereas the other hypothesis made an opposing prediction of a null finding. For each pair of hypotheses, participants were instructed to guess which of the two hypotheses were true. As in Study 1, the researchers prejudged five of the evolutionary hypotheses as being more obvious or less obvious. An example of a pair containing an obvious hypothesis is "Children are typically *more* fearful of male strangers than of female strangers" and "Children are typically *no more* fearful of male strangers than of female strangers" (the first is correct; Heerwagen & Orians, 2002). An example of a pair containing a less obvious hypothesis is "More attractive people tend to be *equally* intelligent as unattractive people" and "More attractive people tend to be *more* intelligent than unattractive people" (the second is correct; Kanazawa, 2010). A final question asked participants whether they were previously familiar with any of the hypotheses, and if so, to list them. Finally, all participants were debriefed.

Study 2 Results

As in Study 1, we removed data for one of the "surprising" pairs of hypotheses we presented because the finding that made up this item had been retracted by the researchers (Pollet & Nettle, 2010).

Consistent with findings from Study 1, undergraduate students were correct a greater proportion of times for the pairs involving obvious rather than the pairs involving less obvious hypotheses ($M = 4.33$ and $M = 1.00$, $t(14) = 7.48$, $p < .001$, $d = 1.93$). We had entertained the possibility that graduate students would be very good at the task and therefore would be correct almost all of the time for both groups of hypotheses. In fact, this was not so; the graduate student data were not very different from the data obtained from the undergraduate students ($M = 4.33$ and $M = 1.22$, $t(8) = 7.79$, $p < .001$, $d = 2.60$).

If the participants simply guessed on each pair of hypotheses, they should have been correct approximately 50% of the time. This fact implies three competing possibilities. The first of these is that perhaps the less obvious hypotheses were nevertheless reasonably obvious, which implies that participants should have guessed more than 50% of them correctly. Secondly, participants may simply have had no clue whether the hypotheses or their opposites were correct, in which case their scores should have averaged around 50%. Finally, it is possible that the less obvious hypotheses not only were not obvious, but actually ran counter to participants' intuitions; according to this third possibility participants should have guessed correctly on fewer than 50% of the hypotheses. In fact, both undergraduate and graduate students were correct on fewer than 50% of the items (percentage correct was 25% for undergraduates, $\chi^2(1) = 15.00$, $p < .05$, and 24% for graduates, $\chi^2(1) = 9.00$, $p < .05$), thereby supporting that the less obvious evolutionary psychology predictions actually ran counter to the intuitions of both undergraduate and graduate students.

As in Study 1, we repeated the foregoing analyses but excluding responses to hypotheses with which participants had claimed prior familiarity. The findings were similar.

Study 3 Method

In Studies 1 and 2, we tested hypotheses that we had prejudged to be non-obvious against ones that we had prejudged to be obvious. The data demonstrate that it is possible to find non-obvious hypotheses, which support that the set of falsifiable evolutionary psychology hypotheses exceeds zero. It is less clear, however, that this represents evolutionary psychology in its typical form. To explore this latter issue, it is necessary to compare highly cited hypotheses against obvious ones, as opposed to comparing non-obvious hypotheses – that we had searched for carefully – against obvious ones. Consequently, Studies 3 and 4 were similar to Studies 1 and 2, respectively, except that we were interested in highly cited evolutionary psychology hypotheses rather than non-obvious ones.

Participants

Forty-five undergraduate psychology students participated in Study 3. These students received course credit for their participation. (We did not test graduate students because we had used them up in the course of performing Studies 1 and 2. In addition, we obtained similar patterns of findings from the graduate and undergraduate students in Studies 1 and 2.)

Procedure

Similar to Study 1, participants were presented with ten evolutionary psychology hypotheses and instructed to rate each one on an "obviousness" scale. No explicit reference was made to evolutionary psychology per se in any of the findings presented. Participants rated the findings on a seven-point scale with -3 representing "extremely obvious," 0 representing the neutral point, and 3 representing "extremely surprising."

As in Study 1, the researchers prejudged five of the hypotheses as being obvious. The same five "obvious" items that were used in Study 1 were also used in Study 3. The

other five items were drawn from some of the most highly cited papers in evolutionary psychology's flagship journal *Evolution and Human Behavior* and its predecessor *Ethology and Sociobiology*. The "highly cited" items were chosen on the basis of three criteria. The first criterion was that the finding came from the most highly cited papers in the aforementioned journals. The second criterion was that the paper from which the findings were drawn made empirical rather than methodological or philosophical claims. The third criterion was that the finding had to be translatable into a one-sentence description that a psychology undergraduate could understand. The "highly cited" findings used in this study came from the most highly cited papers from the aforementioned journals, which also fit the second two criteria. An example of a highly cited item is "People whose bodies are more symmetrical have more attractive faces than people whose bodies are less symmetrical" (Gangestad, Thornhill & Yeo, 1994). No more than one finding was drawn from any one paper.

A final question asked participants whether they were previously familiar with any of the hypotheses, and if so, to list them. Finally, all participants were debriefed.

Study 3 Results

Consistent with expectations, participants rated the highly cited hypotheses as less obvious than the ones we had prejudged to be obvious (M = -.37 and M = -1.51, $t(44)$ = -8.02, $p < .001$, d = 1.19). We also tested the highly cited hypotheses against the neutral point of zero. In contrast to our findings in Study 1, we found that the highly cited hypotheses were rated as being slightly obvious ($p < .05$).

We repeated all of the foregoing analyses but excluding responses to hypotheses with which participants had claimed prior familiarity. We again found that the highly cited items were rated as being less obvious than the obvious items. However, we did not find that the ratings of the highly cited items differed significantly from zero. This analysis did not take into account responses to items with which participants were previously familiar, and it showed that the highly cited hypotheses were neither obvious nor surprising.

Study 4 Method

The Study 3 findings indicate that although the highly cited evolutionary psychology hypotheses fared well against ones we had prejudged to be obvious, they were not considered particularly surprising. In Study 4, we tested the obviousness of highly cited evolutionary psychology findings by having participants guess the correct hypothesis, as in Study 2.

Participants

A new and independent sample of 47 undergraduate psychology students participated in Study 4. These students received course credit for their participation.

Procedure

The methodology of Study 4 replicated that of Study 2. Participants were presented with ten pairs of hypotheses. One hypothesis in each pair was from the

evolutionary psychology literature and was used in Study 3 whereas the other hypothesis made an opposing prediction of a null finding. For each pair of hypotheses, participants were instructed to guess which of the two hypotheses were true. As in Study 3, the researchers prejudged five of the evolutionary hypotheses as being obvious and five of the findings were drawn from the some of most highly cited papers in *Evolution and Human Behavior* and *Ethology and Sociobiology*. An example of a pair containing a highly cited hypothesis is "Women are *better* than men at remembering where objects are located in an array" and "Women are *no better* than men at remembering where objects are located in an array" (the first is correct; Eals & Silverman, 1994). A final question asked participants whether they were previously familiar with any of the hypotheses, and if so, to list them. Finally, all participants were debriefed.

Study 4 Results

Participants were correct a greater proportion of times for the pairs involving obvious rather than the pairs involving highly cited hypotheses ($M = 4.34$ and $M = 3.09$, $t(46) = 6.74$, $p < .001$, $d = .98$). As in Study 2, we compared performance on highly cited hypotheses to guessing (50%). We found that participants answered these items correctly at a rate higher than would be predicted by chance (percentage correct was 62%, $\chi^2(1) = 13.83$, $p < .05$). Therefore, these items were at least somewhat obvious.

As in the other studies, we repeated the foregoing analyses but excluding responses to hypotheses with which participants had claimed prior familiarity. The findings were similar.

Discussion

The data from Studies 1 and 2 demonstrate that there exist some non-obvious evolutionary psychology hypotheses that have been confirmed empirically. In Popper's terms, then, evolutionary psychology has passed some potentially destructive tests and hence it would be incorrect to characterize the whole area as not being falsifiable.

It is important not to exaggerate the findings as meaning more than they really mean. For example, the fact that some evolutionary psychology hypotheses are not obvious does not mean that all of them are surprising. In fact, some of the most famous and widely cited ones clearly were not particularly surprising and fooled significantly fewer than 50% of the participants in Study 4. Consequently, if a critic wished to argue that much (though not all) of evolutionary psychology research is not in the spirit of Popper's falsification philosophy, the present findings should not be interpreted as providing a strong defense against this particular argument.

With the foregoing caveat in mind, let us return to the present findings and consider potential criticisms. The most noticeable criticism pertains to the obviousness ratings in Study 1. It is possible that people were influenced by hindsight bias; everything is obvious in hindsight. Or, merely being exposed to the hypotheses might have made them seem more obvious. Note, however, that these alternative explanations push for more negative obviousness ratings. Consequently, although they potentially explain the highly negative ratings for the more obvious hypotheses, they do not explain the positive ratings for the non-obvious hypotheses.

A more general criticism of the obvious ratings might be that ratings never can be trusted and so they cannot be trusted here either. Because of the general nature of this

criticism, and its failure to provide a specific explanation of what we actually found, it naturally is more difficult to refute. Nevertheless, we remind the reader of the following. In the first place, the obvious hypotheses really were rated negatively and the less obvious ones really were rated positively, and so these findings are incompatible with a general and extreme bias in one direction or the other. Secondly, the findings replicated with a different task in Study 2 where participants actually had to choose between hypotheses. Even if there were a problem with the ratings in Study 1, the critic also would have to explain the findings from Study 2.

One last clarification remains. From the specific point of view of falsification, or even the more general theory testing approach, obvious hypotheses clearly are bad. Good theories must be shown to make at least some predictions that have a reasonable chance of being wrong (Roberts & Pashler, 2000). But researchers might have goals other than theory testing and using obvious hypotheses might fit them (Ichheiser, 1943; Richard, Bond, & Stokes-Zoota, 2001). We do not actually advocate a falsification perspective but merely point out that the falsification criticism of evolutionary psychology undeniably depends on a prior assumption that falsification is important. Therefore, we went ahead and assumed this perspective for the sake of completing the present research. Of course, if one does not believe that the falsification issue matters very much, then the criticism that evolutionary psychology is not falsifiable loses its force anyhow.

In conclusion, the present findings demonstrate that the set of non-obvious hypotheses from evolutionary psychology exceeds zero, though the ones that are highly cited were shown to be not particularly surprising. Therefore, the whole area cannot validly be tarred with the accusation of not being falsifiable unless one insists on absolute falsification that never can be achieved anyhow (Trafimow, 1999). Nor can evolutionary psychology correctly be characterized only as telling us what we already know; clearly there are some surprising hypotheses that have received empirical confirmation. In addition, the fact that a few non-obvious predictions were uncovered here strongly implies that additional ones can be discovered too. Thus, we see no reason for evolutionary psychologists to worry about whether the area, as a whole, is vulnerable to criticisms pertaining to obvious hypotheses. But this conclusion does not absolve individual evolutionary psychologists (or any other psychologists for that matter) from the responsibility of carefully considering the quality of the specific hypotheses that each one of them proposes or makes the topic of their own research. Our finding that highly cited hypotheses are not surprising and were guessed correctly significantly more than 50% of the time, suggests that perhaps evolutionary psychologists might be more discriminating in the future.

Received August 4, 2010; Revision received May 25, 2011; Accepted October 1, 2011

References

Ackermann, R. J. (1976). The philosopy of Karl Popper. Amherst, MA: University of Massachusetts Press.

Agassi, J. (2008). *A philosopher's apprentice: In Karl Popper's workshop* (revised, extended and annotated edition). Amsterdam: Rodopi.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses testing in 37 cultures. *Behavioral and Brain Sciences, 12*, 1-49.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review*, *100*, 204-232.

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*, 152–66.

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, Part II: Case study: A computational theory of social exchange. *Ethology and Sociobiology, 10*, 51-97.

Duhem, P. (1954). *The aim and structure of physical theory* (P.P. Wiener, Trans.). Princeton, NJ: Princeton University Press. (Original work published 1906)

Eals, M., Silverman, I. (1994) The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology, 15*(2), 95–105.

Euler, H. A., & Weitzel, B. (1996). Discriminative grandparental solicitude as reproductive strategy. *Human Nature, 7*, 39-59.

Gangestad, S. W., Thornhill, R., & Yeo, R. (1994). Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethology and Sociobiology, 15,* 73-85.

Gannon, L. (2002). A critique of evolutionary psychology. *Psychology, Evolution & Gender*, *4*(2), 173–218.

Geary, D. C. (2000). Evolution and proximate expression of human paternal investment. *Psychological Bulletin, 126*(1), 55-77.

Grammer, K., & Thornhill, R. (1994). Human facial attractiveness and sexual selection: The roles of averageness and symmetry. J*ournal of Comparative Psychology, 108*, 233-242.

Heerwagen, J. H., & Orians, G. H. (2002). The ecological world of children. In P. H. Kahn, Jr., & S. R. Kellert (Eds.), *Children and nature: Psychological, sociocultural, and evolutionary investigations* (pp. 29-64). Cambridge, MA: MIT Press.

Ichheiser, G. (1943). Why psychologists tend to overlook certain "obvious" facts. *Philosophy of Science*, *10*, 204-207.

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and Human Behavior, 22*, 251-267.

Kanazawa, S. (2011). Intelligence and physical attractiveness, *Intelligence*, *39*, 7-14.

Ketelaar, T., & Ellis, B. J. (2000). Are evolutionary explanations unfalsifiable? Evolutionary psychology and the Lakatosian philosophy of science. *Psychological Inquiry, 11,* 1–21.

Kitcher, P. (1985). *Vaulting ambition: Sociobiology and the quest for human nature.* Cambridge, MA: MIT Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-195). London: Cambridge University Press.

Lakatos, I., 1978. *The methodology of scientific research programmes*. Cambridge, UK: Cambridge University Press.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant using it. *Psychological Inquiry*, *1*, 108–141.

Miller, A. S., & Kanazawa, S. (2008). *Why beautiful people have more daughters: From dating, shopping, and praying to going to war and becoming a billionaire—two evolutionary psychologists explain why we do what we do*. New York, NY: Penguin.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.

Pollet, T. V., & Nettle, D. (2009). Partner wealth predicts self-reported orgasm frequency in a sample of Chinese women. *Evolution and Human Behavior, 30,* 146-151.

Pollet, T. V., & Nettle, D. (2010). Correction to Pollet and Nettle (2009): "Partner wealth predicts self-reported orgasm frequency in a sample of Chinese women." *Evolution and Human Behavior*, *31*, 149.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Popper, K. R. (1962). *Conjectures and refutations*. New York: Basic Books.

Popper, K. R. (1972) *Objective knowledge*. Oxford, UK: Oxford University Press.

Popper, K. R. (1983). *Realism and the aim of science*. London: Routledge.

Quine, W. V. O. (1980). Two dogmas of empiricism. In W.V.O. Quine (Ed.), *From a logical point of view* (2nd ed., pp. 20–46). Cambridge,MA: Harvard University Press. (Original work published 1953)

Richard, F. D., Bond, C. F. Jr., & Stokes-Zoota, J. J. (2001). "That's completely obvious… and important": Lay judgments of social psychological findings. *Personality and Social Psychology Bulletin*, *27*, 497-505.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367.

Schlinger, H. D. (1996). How the human got its spots: A critical analysis of the just so stories of evolutionary psychology. *Skeptic*, *4*, 68-76.

Schmitt, D. P. (2002). How shall I compare thee? Evolutionary psychology viewed as a psychological science. *Psychology, Evolution & Gender*, *4*(2), 219-230.

Sesardic, N. (2003). Evolution of human jealousy: A just-so story or a just-so criticism? *Philosophy of the Social Sciences, 33*(4), 427–443.

Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., & Olshansky, E. (2000). Evolved mechanisms underlying wayfinding: Further studies on the hunter-gatherer theory of spatial sex differences. *Evolution and Human Behavior, 21*, 201-213.

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review*, *110*, 526–535.

Trafimow, D. (2009). The theory of reasoned action: A case study of falsification in psychology. *Theory & Psychology*, *19*, 501-518.

Wilson, M., & Daly, M. (1985). Competitiveness, risk taking, and violence: The young male syndrome. *Ethology and Sociobiology, 6*, 59-73.

Appendix

Obvious, Non-Obvious, and Highly Cited Hypotheses

Obvious hypotheses

1. Women typically invest more time and energy into raising children than men do (Geary, 2000).
2. People find symmetrical faces to be more attractive than asymmetrical faces (Grammer & Thornhill, 1994).
3. Men typically desire a greater number of sex partners than women do (Buss & Schmitt, 1993).
4. Children are typically more fearful of male strangers than of female strangers (Heerwagen & Orians, 2002).
5. Men typically show a preference for youthful women when seeking a sexual partner (Buss, 1989).

Non-obvious Hypotheses

1. More attractive people tend to be more intelligent than unattractive people (Kanazawa, 2010).
2. Maternal grandmothers (a child's mother's mother) typically invest more time and resources into their grandchildren than do paternal grandmothers (a child's father's mother) (Euler & Weitzel, 1996).
3. Men are better than women at navigating through strange territories (Silverman et al., 2000).
4. When an attractive man and an attractive woman have a baby they are more likely to have a daughter as compared to couples that are not as attractive (Miller & Kanazawa, 2008).
5. Women will orgasm more frequently while having sex with wealthier men compared to having sex with less wealthy men (Pollet & Nettle, 2009). This finding was retracted and not used in our analyses (see Pollet & Nettle, 2010).

Highly Cited Hypotheses

1. People whose bodies are more symmetrical have more attractive faces than people whose bodies are less symmetrical (Gangestad, Thornhill & Yeo, 1994).
2. Women are better than men at remembering where objects are located in an array (Eals & Silverman, 1994).
3. Women who are in the fertile phase of their menstrual cycle (capable of becoming pregnant) prefer the faces of more masculine looking men than women who are not in the fertile phase of their menstrual cycle (Johnston, Hagel, Franklin, Fink & Grammer, 2001).
4. Young men commit more murders compared to both older men and women (Wilson & Daly, 1985).
5. People are better at reasoning about problems related to social situations compared to abstract problems not related to social situations, even when the problems are technically exactly the same (Cosmides & Tooby, 1989).