Of beauty, sex, and power: Statistical challenges in estimating small effects*

Andrew Gelman[†] David Weakliem[‡]

September 8, 2007

Abstract

A series of papers in the *Journal of Theoretical Biology* has found evidence that beautiful parents have more daughters, violent men have more sons, and other sex-ratio patterns (Kanazawa, 2005, 2006, 2007). These papers have been shown to have statistical errors, but the question remains how to interpret findings that are intriguing, potentially important, but not statistically significant. From a classical statistical perspective, these studies have insufficient power to detect the magnitudes of effects (on the order of 1 percentage point) that could be expected based on earlier studies of sex ratios. The anticipated small effects can also be handled using a Bayesian prior distribution. These concerns are relevant to other studies of small effects and also to the reporting of such studies.

Keywords: Bayesian inference, evolutionary psychology, power analysis, prior distribution, sex ratio, sociobiology, Trivers-Willard hypothesis, Type M error, Type S error

*We thank Kelly Rader and Lucy Robinson for data collection and analysis, David Dunson, Thomas Volscho, David Judkins, and Jeremy Freese for helpful comments, and the National Science Foundation, National Institutes of Health, and Columbia University Applied Statistics Center for financial support.

[†]Department of Statistics and Department of Political Science, Columbia University, New York, `gelman@stat.columbia.edu`, `www.stat.columbia.edu/~gelman`

[‡]Department of Sociology, University of Connecticut, `weakliem@uconn.edu`, `web.uconn.edu/weakliem`

Do beautiful parents have more daughters?: an interesting but unproven claim

Satoshi Kanazawa, a Reader in Management and Research Methodology at the London School of Economics, has published a series of papers in the *Journal of Theoretical Biology* with titles such as "Big and tall parents have more sons," "Violent men have more sons," "Engineers have more sons, nurses have more daughters," and "Beautiful parents have more daughters" (Kanazawa, 2005, 2006, 2007, Kanazawa and Vandermassen, 2005). More recently, he has publicized some of these claims in an article, "Ten politically incorrect truths about human nature," for *Psychology Today* and in a forthcoming book, *Why Beautiful People Have More Daughters* (Miller and Kanazawa, 2007a, b).

However, the statistical analysis underlying Kanazawa's claims has been shown to be flawed, with some analyses making the error of controlling for an intermediate outcome in estimating a causal effect, and another analysis being subject to multiple comparisons problems. In short, Kanazawa's findings are not statistically significant, and the patterns in his analysis could well have occurred by chance (Gelman, 2007a).

Had the lack of statistical significance been noticed in the review process, these articles almost certainly would not have published in the journal; however, the fact of their appearance (and their prominence in a forthcoming popular book) leads to an interesting statistical question: how should we think about research findings that are intriguing but not statistically significant? A quick answer would be to simply ignore them: after all, anyone armed with even a simple statistics package can go through public databases fishing for correlations to confirm a pre-existing hypothesis.[1]

This might be too glib, however, since even nonsignificant findings can be suggestive. To consider the example that we shall explore further in this paper, a regression analysis finds the probability of a girl birth to be 4.7% more likely for attractive than for unattractive parents—but with a standard error of 4.3%. Not statistically significant, but if we had to guess whether girls are more likely to be born to beautiful or ugly parents, the data would suggest the former.

There also are substantive reasons why Kanazawa's hypothesis should not be dismissed out of hand, even though his results are not statistically significant. First, his findings are motivated theoretically by a well-respected model of Trivers and Willard (1973); see Fawcett et al. (2007) for a recent review. In addition, beauty has been found to have effects

[1] We are not suggesting that Kanazawa did this, merely that an investigator *could* follow this strategy of searching for confirmation.

in areas including economics (Hamermesh and Biddle, 1994) and education (Hamermesh and Parker, 2005), boy and girl children have been associated with political attitudes (Oswald and Powdthavee, 2006, Washington, 2007).

The present article focuses on the question of how to interpret non-significant results. We also touch on distortions of equivocal findings by the popular media, which is partly a consequence of the difficulties that statisticians have in giving guidance in such settings. One reason this topic is important is that systematic errors such as overestimation of the magnitudes of small effects can mislead scientists and, through them, the general public.

Classical and Bayesian inference for small effects

We focus on Kanazawa's (2007) analysis of data from the National Longitudinal Survey of Adolescent Health, which included interviewers' subjective assessments of respondents' attractiveness (on a 1–5 scale) along with data including the sexes of respondents' children (if any). The analysis is restricted to the sex of the first child (to remove potential biases arising, for example, from the sex of the first child affecting the decision of whether to have more children). After rescaling attractiveness to have a mean of 0 and standard deviation of 0.5 (to be comparable to a binary predictor; see Gelman, 2007b), a regression of the proportion of girl births on attractiveness yields an estimated coefficient of 0.047 (that is, 4.7 percentage points) with a standard error of 0.043. The challenge is to interpret this finding, which is consistent with an existing theory but is not statistically significant.[2]

The key is to recognize that the effects being studied are likely to be small. There is a large literature on variation in the sex ratio of human births, and the effects that have been found have been on the order of 1 percentage point (for example, the probability of a girl birth shifting from 48.5% to 49.5%). Variation attributable to factors such as race, parental age, birth order, maternal weight, partnership status, and season of birth is estimated at from less than 0.3 percentage points to about 2 percentage points (James, 1987, Chahnazarian, 1988, Cagnacci et al., 2003, 2004, 2005, Norberg, 2004), with larger changes (as high as 3 percentage points) arising under economic conditions of poverty and famine (Ansari-Lari and Saadat, M., 2002, Catalano, 2003, Almond et al., 2007). That extreme deprivation increases the proportion of girl births is no surprise, given that male fetuses (and also male babies and adults) are more likely than females to die under adverse

[2] Kanazawa (2007) identified a statistically significant 8 percentage point difference—a 52% chance of girl births for the parents in the highest attractiveness category, compared to a 44% chance for the average of the four lower categories. However, as pointed out by Gelman (2007a), this is just one of the many possible comparisons that could be performed with these data, and the regression described here is the more standard statistical analysis and not subject to multiple comparisons issues.

conditions.[3] Based on our literature review, we would expect any effects of beauty on the sex ratio to be less than 1 percentage point, which represents the range of natural variation under normal conditions.[4]

Classical inference. Now let us return to our regression example. With an estimate of 4.7% and a standard error of 4.3%, the classical 95% interval for the difference in probability of a girl birth, comparing beautiful to ugly parents, is $[-3.9\%, 13.3\%]$. To put it another way, effects as low as -3.9% or as high as $+13.3\%$ are roughly consistent with the data. Given that we only expect to see effects in the range $\pm 1\%$, we have essentially learned nothing from this study.

Another way to frame this is to consider what would happen if repeated independent studies were performed with the same precision, and thus, approximately the same standard error of 4.3%. There is at minimum a 5% chance of obtaining a statistically significant result, which would imply an estimate of 8.4% or larger in either direction (1.96 standard deviations from zero). If multiple tests are performed, the chance of finding something statistically significant increases, as discussed by Gelman (2007a) in this example. In any case, though, the estimated effect—at least 8.4%—is much larger than anything we would realistically think the effect size could be. This is a Type M (*magnitude*) error (Gelman and Tuerlinckx, 2000): the study is constructed in such a way that any statistically-significant finding will almost certainly be a huge overestimate of the true effect. In addition there will be Type S (*sign*) errors, in which the estimate will be in the opposite direction as the true effect.

We get a sense of the probabilities of these errors by considering three scenarios of studies with standard errors of 4.3 percentage points:

1. **True difference of zero.** If there is no correlation between parental beauty and sex ratio of children, then a statistically significant estimate will occur 5% of the time, and it will always be misleading.

2. **True difference of 0.3%.** If the probability of girl births is actually 0.3% higher among beautiful than among ugly parents, then there is a 3% probability of seeing a

[3] In addition, large sex ratio differences in the other direction have been observed—an excess of boy births in parts of India, China, and other countries—but this is likely attributable to selective infanticide and abortion (Das Gupta, 2005, 2006, 2007) or possibly hepatitis (Oster, 2005, 2006) and not so relevant to the present discussion.

[4] In addition to true variation in probabilities, observed proportions show simple binomial variability, which has a standard deviation of $0.5/\sqrt{n}$, which equals 1.1% for a sample size of $n = 2000$ or 0.5% for a sample size of 10,000. This will appear as part of the noise in any statistical study.

statistically significant positive result—and a 2% chance of seeing a statistically significant negative result. In either case, the estimated effect, of at least 8.4 percentage points, will be over an order of magnitude higher than the true effect, and with a 2/5 chance of going in the wrong direction. If the result is *not* statistically significant, the chance of the estimate being in the wrong direction (a Type S error) is 47.5%, implying that the direction of the estimate provides almost zero information on the sign of the true effect.

3. **True difference of 1%.** If the probability of girl births is actually 1% higher among beautiful than among ugly parents—which, based on the literature, is on the high end of possible effect sizes—then there is a 4% chance of a statistically-significant positive result, and still over a 1% chance of a statistically-significant result in the wrong direction. Overall there is a 40% chance of a Type S error: again, the estimate gives us little information about the sign or the magnitude of the true effect.

4. **True difference of 3%.** Even if the true difference were as high as 3%, which we find implausible from the literature review, there is still only a 10% chance of obtaining statistical significance, and the overall Type S error rate is 24%.

A study of this size is thus not fruitful for estimating variation on this scale. This is one reason that successful studies of the human sex ratio use much larger samples, typically from demographic databases (e.g., Almond and Edlund, 2006, Almond et al., 2007).

Bayesian inference. We can also redo Kanazawa's analysis using a Bayesian prior distribution. To start with, given sufficiently diffuse prior distribution, the posterior distribution would be approximately normal with mean 4.7 and standard error 4.3, which would imply about an 86% probability that the true effect is positive. In general, the more concentrated the prior distribution around 0 (i.e., a presumption based on the sex-ratio literature that the true effect is likely to be small), the closer the posterior probability will be to 50%.

For example, consider a Cauchy distribution with center 0 and scale 0.3%, which implies that the true difference in percentage of girls, comparing beautiful and ugly parents, is most likely to be near zero, with a 50% chance of being in the range $[-0.3\%, 0.3\%]$, a 90% chance of being in the range $[-1\%, 1\%]$, and a 94% chance of being less than 3 percentage points in absolute value. We center the prior distribution at zero because, ahead of time, we have no particular reason to believe the coefficient will be positive or negative. (As we discuss later, convincing arguments can be given to predict a coefficient in either direction; see Freese, 2007, and Fawcett et al., 2007.) The Cauchy family has flat tails so that, if the data do convincingly show a very large effect, this will not be contradicted by the prior distribution.

We combine the prior distribution with the normal likelihood for the regression coefficient on the standardized beauty coefficient (that is, the likelihood corresponding to the normal distribution with mean 4.7% and standard deviation 4.3%). The resulting posterior distribution gives a probability that the difference is positive—that beautiful parents actually have more daughters—is only 58%—and even if the effect is positive, there is a 78% chance it is less than 1 percentage point. This analysis depends on the prior distribution but not to an extreme amount; for example, if we broaden the Cauchy prior and give it a scale of 1%, the posterior probability that the true difference is positive is still only 65%. Switching from the Cauchy to another family such as the normal distribution has little effect on the results. The key is that effects are likely to be small, and in fact the data are consistent with small results.

In any case, although Bayesian inference may be the most effective approach for any given data set, we see frequentist analysis as necessary for estimating the systemic effects of applying a statistical procedure, as argued in a different context by Rubin (1984). An ideal is for scientific understanding about a quantity (in this case, the correlation between beauty of parents and sex ratio of children) to have a recognized uncertainty that can be summarized by a probability distribution. Individual researchers can collect data or creatively analyze existing sources (as was done by Kanazawa) and publish their results, and then occasional meta-analyses can be done to review the results. This procedure smooths some of the variation which is inherent in these small-sample studies, where the probability of a positive effect can jump from 50% to 58%, then perhaps down to 38% with the next study, and so forth.

The 50 most beautiful people

One way to calibrate our thinking about Kanazawa's results is to collect more data. Every year, *People* magazine publishes a list of the fifty most beautiful people, and, because they are celebrities, it is not difficult to track down the sexes of their children, which we did for the years 1995–2000.[5]

As of 2007, the 50 most beautiful people of 1995 had 32 girls and 24 boys, or 57.1% girls, which is 8.6 percentage points higher than the population frequency of 48.5%. This

[5]Data were collected from Wikipedia, the Internet Movie Database, and celebrities' personal webpages in August, 2007. Information was missing for 2 beautiful people in 1995, 2 in 1996, 3 in 1997, 6 in 1998, 3 in 1999, and 2 in 2000. The data are available at `www.stat.columbia.edu/~gelman/research/beautiful/`.

sounds like good news for the hypothesis. But the standard error is $0.5/\sqrt{56} = 6.7\%$, so the discrepancy is not statistically significant. Let's get more data.

The 50 most beautiful people of 1996 had 45 girls and 35 boys: 56.2% girls, or 7.8% more than in the general population. Good news! Combining with 1995 yields 56.6% girls—8.1% more than expected—with a standard error of 4.3%, tantalizingly close to statistical significance. Let's continue to get some confirming evidence.

The 50 most beautiful people of 1997 had 24 girls and 35 boys—no, this goes in the wrong direction, let's keep going ... For 1998, we have 21 girls and 25 boys, for 1999 we have 23 girls and 30 boys, and the class of 2000 has had 29 girls and 25 boys.

Putting all the years together and removing the duplicates, such as Brad Pitt, *People*'s most beautiful people from 1995 to 2000 have had 157 girls out of 329 children, or 47.7% girls (with standard error 2.8%), a statistically insignificant 0.8% percentage points lower than the population frequency. So nothing much seems to be going on here. But if statistically insignificant effects with a standard error of 4.3% were considered acceptable, we could publish a paper every two years with the data from the latest "most beautiful people."

Why is this important?

Why does this matter? Why are we wasting our time on a series of papers with statistical errors that happen not to have been noticed by reviewers for a fairly obscure journal? We have two reasons: first, as discussed in the next section, the statistical difficulties arise more generally with findings that are suggestive but not statistically significant. Second, as we discuss presently, the structure of scientific publication and media attention seem to have a biasing effect on social science research.

Before reaching *Psychology Today* and book publication, Kanazawa's findings received broad attention in the news media. For example, the popular Freakanomics blog (Dubner, 2006) reported,

> "a new study by Satoshi Kanazawa, an evolutionary psychologist at the London School of Economics, suggests ... there are more beautiful women in the world than there are handsome men. Why? Kanazawa argues its because good-looking parents are 36% more likely to have a baby daughter as their first child than a baby son—which suggests, evolutionarily speaking, that beauty is a trait more valuable for women than for men. The study was conducted with data from 3,000 Americans, derived from the National Longitudinal Study of Adolescent Health, and was published in the Journal of Theoretical Biology."

Publication in a peer-reviewed journal seemed to have removed all skepticism, which is noteworthy given that the authors of *Freakanomics* are themselves well qualified to judge social science research.

In addition, the estimated effect grew during the reporting. As noted above, the 4.7% (and not statistically significant) difference in the data became 8% in Kanazawa's choice of the largest comparison, which then became 26% when reported as a logistic regression coefficient (see Gelman, 2007a), and then jumped to 36% for reasons unknown (possibly a typo in a newspaper report). The funny thing is that the reported 36% signaled to us right away that something was wrong, since it was 10 to 100 times larger than reported sex-ratio effects in the literature (James, 1987, Chahnazarian, 1988). Our reaction when seeing such large estimates was not "Wow, they've found something big!" but, rather, "Wow, this study is underpowered!"

This problem will occur again and again, and is worth thinking about now. To start with, most of the low-hanging fruit in social science have presumably been plucked, leaving researchers to study small effects. Sex ratios are of inherent interest to all of us who have or are considering having babies, as well as for their implications for the organization of society. Miller and Kanazawa's billing of their result as a "politically incorrect truth" hints at the connection to live political issues such as abortion, parental leave policies, and comparable-worth laws that turn upon judgments of the appropriate roles for men and women in society.

As discussed earlier in this article, studies with insufficient statistical power will spit out random results which will occasionally be statistically significant and, even more often, be suggestive, as in Kanazawa's beauty-and-births studies. It is tempting to interpret the directions of these essentially random findings without recognizing the fragility of the explanations. Evolutionary psychology could be used to explain a result in the opposite direction, using the following sort of argument: persons judged to be beautiful are, one could claim, more likely to be healthy, affluent, and from dominant ethnic groups, more generally having traits that are valued in the society at large. (Consider, for example, Miss Americas, who until recent decades were all white.) Such groups are more likely to exercise power, a trait that, in some sociobiological arguments, are more beneficial for men than women—thus it would be natural for more attractive parents to be more likely to have boys. We are not claiming this is true; we are just noting that the argument could go in either direction, which puts a particular burden on the data analysis.[6] This ability of

[6]Another argument that the effect could go in the other direction is based on the connection of the

the theory to explain findings in any direction is also pointed out by Freese (2007), who describes this sort of argument as "more 'vampirical' than 'empirical'—unable to be killed by mere evidence."

In statistics, you can't prove a negative. "Beautiful parents have more daughters" is a compelling headline, but "There is no compelling evidence that beautiful parents are more or less likely to have daughters" is not so appealing. As a result, public discourse can get cluttered with unproven claims, which perhaps will lead to a general skepticism that will, in boy-who-cried-wolf fashion, unfairly discredit more convincing research.

The result is a sort of asymmetrical warfare, with proponents of sex differences and other "politically incorrect" results producing a series of empirical papers that, for reasons of statistical power, give essentially random clues about true population patterns, and opponents of this line of research being reduced to statements such as "the data are insufficient" or else ruling that such research is dangerous and out of bounds, an attitude deplored by Pinker (2007). The aforementioned Freakanomics article concluded, "It is good that Kanazawa is only a researcher and not, say, the president of Harvard. If he were, that last finding about scientists may have gotten him fired." It should be possible to criticize large unproven claims in biology and social science without dismissing the entire enterprise.

Why is this not obvious?

The natural reaction of a competent quantitative researcher to the statistics in this article is probably, Duh. Everyone who has ever applied for an NIH grant knows all too much about power calculations. But, if this is so obvious, why did the mistake (by a Reader in Research Methodology, no less) result in not one, but several papers in the *Journal of Theoretical Biology*, a prominent publication with an impressive name and a respectable impact factor of 2.3 (higher than that of the *Journal of the American Statistical Association*, the *Annals of Statistics*, and the *Journal of the Royal Statistical Society*)? One problem, of course, is that referees are busy and statistical errors are not always noticed. But another problem is

generalized Trivers-Willard hypothesis to the pattern that men show more variation than women in many traits. The theory, as Kanazawa describes it, could be expressed by saying that the probability of a male child is positively related to $\mathrm{E}(N|M_q) - \mathrm{E}(N|F_q)$, where $\mathrm{E}(N|M_q)$ is the expected number of children that a male child born to parents with a particular quality would have, and $\mathrm{E}(N|F_q)$ is the expected number of children a female would have. Even if the correlation between beauty and the number of children is stronger for women (which could be plausible), the variance of the number of children is larger for men, so the slope of a regression of number of children on beauty might be larger for men. So in terms of evaluating the theory, getting empirical evidence on actual variation (if any) in $\mathrm{E}(N|M) - \mathrm{E}(N|F)$ is important. The direct analysis of parental beauty and children's sex amounts to taking a variety of possibly weak instruments as substitutes for the theoretically important variable.

the confusing connection between statistical significance and sample size. It is well known that, with a large enough sample size, one can just about always find statistical significant, if small, effects. But it is not so well realized that, when effects truly are small, there is little point in trying to find them with underpowered studies.

These problems are not new, even in the field of sex ratios. For example, in a book with the unfortunate title of *Probability, Statistics, and Truth*, von Mises (1957) studied the sex ratios of births in the 24 months of 1907–1908 in Vienna and found less variation than would be expected from the binomial distribution. He attributed this to different sex ratios in different ethnic groups. In fact, though, the variance, though less than expected by chance, was not *statistically significantly* less (based on the 23 degrees of freedom available from these data). There seems to be a human desire to find more than pure randomness in sex ratios, despite that there is no convincing evidence that boys or girls run in families or that sex ratios vary much at all except under extraordinary conditions.[7]

Realistically, a researcher on sex ratios has to make two arguments: a statistical case that observed patterns represent real population effects and cannot be explained simply by sampling variability, and a biological argument that effects on the order of 1% are substantively important. The claimed effect size of 26% should have aroused suspicion in comparison to the literature on the human sex ratio; in addition, though, the papers snuck through the review process because reviewers did not recognize that the power of the studies were such that the *only* very large estimated effects could make it through the statistical-significance filter. The result is almost a machine for producing exaggerated claims, which of course become only more exaggerated when they hit the credulous news media (with an estimate of $4.7\% \pm 4.3\%$ being ramped up to 26% and then reported as 36%).

Statisticians should take some of the blame here. Classical significance calculations do not make use of prior knowledge of effect sizes, and Bayesian analyses are often not much better. Textbook treatments of Bayesian inference (e.g., Carlin and Louis, 2001, Gelman et al., 2003) almost entirely use noninformative prior distributions and essentially ignore issues of statistical power. Conversely, power calculations are commonly used in designing studies but are rarely used to enlighten data analyses. And theoretical concepts such as Type S and Type M errors (Gelman and Tuerlinckx, 2000) have not been integrated into statistical practice.

[7]For example, recently there has been much discussion of literature on timing of intercourse to modify the sex ratio, with some studies reporting large effects, and various plausible stories to explain these effects. However, careful analyses have found that the true impact of intercourse timing is very small if anything (David Dunson, private communication).

The modern solution to difficulties of statistical communication is to have more open exchange of methods and ideas. More transparency is apparently needed, however: for example, *Psychology Today* did not seem to notice the Gelman (2007a) critique of Kanazwawa's findings in the *Journal of Theoretical Biology* or other methodological criticisms that have appeared in sociology journals (Freese and Powell, 2001, Volscho, 2005). We hope that a more systematic way of understanding estimates of small effects will provide a clearer framework for open communication.

References

Almond, D., and Edlund, L. (2006). Trivers-Willard at birth and one year: evidence from U.S. natality data 1983–2001. *Proceedings of the Royal Society B*.

Almond, D., Edlund, L., Li, H., and Zhang, J. (2007). Long-term effects of the 1959–1961 China famine: mainland China and Hong Kong. Technical report, Department of Economics, Columbia University.

Ansari-Lari, M., and Saadat, M. (2002). Changing sex ratio in Iran, 1976–2000. *Journal of Epidemiology and Community Health* **56**, 622–623.

Cagnacci, A., Renzi, A., Arangino, S., Alessandrini, C., and Volpe, A. (2003). The male disadvantage and the seasonal rhythm of sex ratio at the time of conception. *Human Reproduction* **18**, 885–887.

Cagnacci, A., Renzi, A., Arangino, S., Alessandrini, C., and Volpe, A. (2004). Influences of maternal weight on the secondary sex ratio of human offspring. *Human Reproduction* **19**, 442–444.

Cagnacci, A., Renzi, A., Arangino, S., Alessandrini, C., and Volpe, A. (2005). Interplay between maternal weight and seasons in determining the secondary sex ratio of human offspring. *Fertility and Sterility* **84**, 246–248.

Carlin, B. P., and Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*, second edition. London: CRC Press.

Catalano, R. A. (2003). Sex ratios in the two Germanies: a test of the economic stress hypothesis. *Human Reproduction* **18**, 1972–1975.

Chahnazarian, A. (1988). Determinants of the sex ratio at birth: review of recent literature. *Social Biology* **35**, 214–235.

Das Gupta, M. (2005). Explaining Asia's "missing women": a new look at the data.

Population and Development Review **31**, 529–535.

Das Gupta, M. (2006). Cultural versus biological factors in explaining Asia's "missing women": response to Oster. *Population and Development Review* **32**, 328–332.

Das Gupta, M. (2007). China's "missing girls"—son preference or hepatitis B infections? Research brief, Human Development and Public Services Research, World Bank.

Dubner, S. J. (2006). Why do beautiful women sometimes marry unattractive men? Freakanomics blog, 2 August, 9:44 am.

Fawcett, T. W., Kuijper, B., Pen, I., and Weissing, F. J. (2007). Should attractive males have more sons? *Behavioral Ecology* **18**, 71–80.

Freese, J. (2007). The problem of predictive promiscuity in deductive applications of evolutionary reasoning to intergenerational transfers: three cautionary tales. *Caring and Exchange Within and Across Generations*, ed. A. Booth et al. Washington, D.C.: Urban Institute Press.

Freese, J., and Powell, B. (2001). Making love out of nothing at all? Null findings and the Trivers-Willard hypothesis. *American Journal of Sociology* **106**, 1776–1788.

Gelman, A. (2007a). Letter to the editor regarding some papers of Dr. Satoshi Kanazawa. *Journal of Theoretical Biology* **245**, 597–599.

Gelman, A. (2007b). Scaling regression inputs by dividing by two standard deviations. Technical report, Department of Statistics, Columbia University.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*, second edition. London: CRC Press.

Gelman, A., and Jakulin, A. (2007). Bayes: liberal, radical, or conservative? *Statistica Sinica* **17**, 422–426.

Gelman, A., and Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics* **15**, 373–390.

Hamermesh, D. S., and Biddle, J. E. (1994). Beauty and the labor market. *American Economic Review* **84**, 1174–1194.

Hamermesh, D. S., and Parker, A. M. (2005). Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review* **24**, 369–376.

James, W. H. (1987). The human sex ratio. I: a review of the literature. *Human Biology* **59**, 721–752.

Kanazawa, S. (2005). Big and tall parents have more sons: further generalizations of the Trivers-Willard hypothesis. *Journal of Theoretical Biology* **233**, 583–590.

Kanazawa, S. (2006). Violent men have more sons: further evidence for the generalized the Trivers-Willard hypothesis. *Journal of Theoretical Biology* **239**, 450–459.

Kanazawa, S. (2007). Beautiful parents have more daughters: a further implication of the generalized Trivers-Willard hypothesis. *Journal of Theoretical Biology.*

Miller, A. S., and Kanazawa, S. (2007a). Ten politically incorrect truths about human nature. *Psychology Today*, July/August.

Miller, A. S., and Kanazawa, S. (2007b). *Why Beautiful People Have More Daughters.* New York: Perigee, to appear.

Kanazawa, S., and Vandermassen, G. (2005). Engineers have more sons, nurses have more daughters: an evolutionary psychological extension of Baron-Cohen's extreme male brain theory of autism. *Journal of Theoretical Biology* **233**, 589–599.

Norberg, K. (2004). Partnership status and the human sex ratio at birth. *Proceedings of the Royal Society B* **271**, 2403–2410.

Oster, E. (2005). Hepatitis B and the case of the missing women. *Journal of Political Economy* **113**, 1163–1216.

Oster, E. (2006). On explaining Asia's "missing women": comment on Das Gupta. *Population and Development Review* **32**, 323–327.

Oswald, A., and Powdthavee, N. (2006). Daughters and left-wing voting. Technical report, Department of Economics, University of Warwick, U.K.

Pinker, S. (2007). In defense of dangerous ideas. *Chicago Sun-Times*, 15 July.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* **12**, 1151–1172.

Trivers, R. L, and Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science* **179**, 90–92.

Volscho, T. W. (2005). Money and sex, the illusory universal sex difference: comment on Kanazawa. *Sociological Quarterly* **46**, 716–736.

von Mises, R. (1957). *Probability, Statistics, and Truth*, second edition. New York: Dover. Reprint.

Washington, E. L. (2007). Female socialization: how daughters affect their legislator fathers' voting on women's issues. *American Economic Review*, to appear.