Why are (the best) women so good at chess? Participation rates and gender differences in intellectual domains

4 authors, including:

Merim Bilalic
Northumbria University
42 PUBLICATIONS   **834** CITATIONS

SEE PROFILE

Fernand Gobet
University of Liverpool
322 PUBLICATIONS   **6,529** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Corpus Analysis of Early Grammatical Development in English View project

Project   Psychology of expertise View project

Why are (the best) women so good at chess? Participation rates and gender differences in intellectual domains

Merim Bilalić[1,*], Kieran Smallbone[2], Peter McLeod[1] and Fernand Gobet[3]

[1]*Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK*
[2]*Manchester Centre for Integrative Systems Biology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK*
[3]*Centre for Cognition and Neuroimaging, Brunel University, Uxbridge, Middlesex UB8 3PH, UK*

A popular explanation for the small number of women at the top level of intellectually demanding activities from chess to science appeals to biological differences in the intellectual abilities of men and women. An alternative explanation is that the extreme values in a large sample are likely to be greater than those in a small one. Although the performance of the 100 best German male chess players is better than that of the 100 best German women, we show that 96 per cent of the observed difference would be expected given the much greater number of men who play chess. There is little left for biological or cultural explanations to account for. In science, where there are many more male than female participants, this statistical sampling explanation, rather than differences in intellectual ability, may also be the main reason why women are under-represented at the top end.

Keywords: gender differences; participation (base) rates; chess; intellectual activities; intelligence; science

1. INTRODUCTION

The former president of Harvard University, Lawrence Summers, expressed a view widely held by both academic researchers (e.g. Geary 1998; Kimura 1999; Pinker 2002) and laypeople when he suggested that innate biological differences in intellectual abilities may explain why the most successful scientists and engineers are predominantly male (Summers 2005). Recent research confirms that, although there is little difference between the average scores of men and women on intelligence and aptitude tests, highly intelligent people are predominantly male (Deary *et al.* 2003; Irwing & Lynn 2005). As Irwing & Lynn (2005) said: 'different proportions of men and women with high IQs… may go some way to explain the greater numbers of men achieving distinctions of various kinds for which a high IQ is required, such as *chess Grandmasters*, Fields medallists for mathematics, Nobel prize winners and the like' (p. 519; emphasis added).

Despite the increasing proportion of women in intellectually demanding professions, they are still under-represented at the top level in science (Long 2001; Xie & Shauman 2003; Ceci & Williams 2007) and engineering (Long 2001). There are various possible explanations for this apart from innate biological differences (for critiques and negative findings see Kerkman *et al.* 2000; Spelke 2005; Lachance & Mazzocco 2006). Socialization and different interests (Benbow *et al.* 2000; Ayalon 2003), gender roles (Massa *et al.* 2005), gatekeeper effects

(Davidson & Cooper 1992; Steele 1997; Huffman & Torres 2002), cultural differences (Andreescu *et al.* 2008) and higher participation rates of men (Charness & Gerchak 1996; Chabris & Glickman 2006) have all been proposed. Here, we show that in chess, an intellectually demanding activity where men dominate at the top level, the difference in the performance of the best men and women is largely accounted for by the difference that would be expected, given the much greater number of men who participate. Despite the clear superiority of the top male players, there is, in reality, very little performance gap in favour of men for non-statistical theories to explain.

Before considering cultural or biological explanations for better performance by the top performers in one of two groups of different size, a simple statistical explanation must be considered. Even if two groups have the same average (mean) and variability (s.d.), the highest performing individuals are more likely to come from the larger group. The greater the difference in size between the two groups, the greater is the difference to be expected between the top performers in the two groups. Nothing about underlying differences between the groups can be concluded from the preponderance of members of the larger group at the far ends of the distribution until one can show that this preponderance is greater than would be expected on statistical sampling grounds.

It is difficult to quantify how participation rates influence the number of outstanding men and women in fields such as science and engineering because both achievement and participation rates are difficult to measure. But it is straightforward in chess because there is an objective measure of achievement and the number of male and female participants is known. Chess has an

* Author and address for correspondence: Section for Experimental MR, Department of Neuroradiology, Tübingen University, Hoppe-Seyler-Street 3, 72076 Tübingen, Germany (merim.bilalic@med.uni-tuebingen.de).
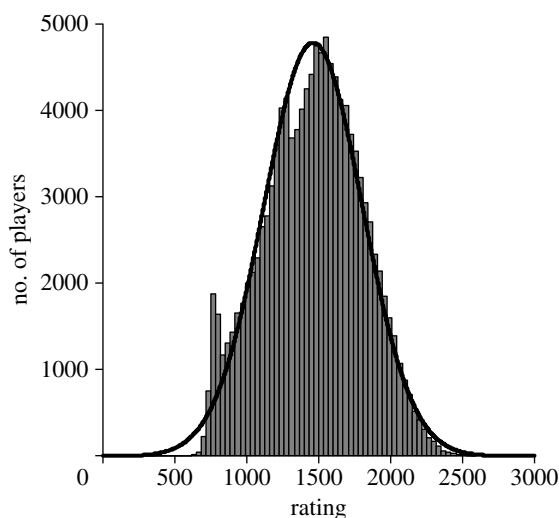
Figure 1. The distribution of the German chess rating with the best-fit normal curve superimposed. $n = 120\,399$, $\mu = 1461$, $\sigma = 342$, 16 : 1 men to women ratio.

interval scale, the Elo rating, for measuring skill level. Every serious player has an Elo rating that is obtained on the basis of their results against other players of known rating (see Elo 1986). National federations record the ratings of all players who take part in competitions. Hence, it is possible to test whether the difference between the best male and female performers is any greater than would be expected given the different numbers of male and female chess players.

2. MATERIAL AND METHODS

We developed an analytic method to estimate the expected difference between the top male and female performers based on the overall male and female participation rates using the parameters of central tendency (mean) and variability (s.d.) of the underlying population. A novelty of this method is that it can be used to estimate the values of the extreme members of very large samples (such as ours, where $n = 120\,399$). This method is described in detail in the appendix A. Our approach is based on the work of Charness & Gerchak (1996) who showed that the difference between the top male and female player in the world in 1994 was similar to that predicted by the relative numbers of male and female players in the ratings of the international chess federation (FIDE). A limitation with this study was the statistical problem that the estimate of the extreme value from a sample tends to be highly variable (Glickman & Chabris 1996; Glickman 1999). Also, the world's top female player at the time, the Hungarian Judit Polgár, was a phenomenon, by far the strongest female player the world has ever known. She is currently ranked 27th of all players in the world but she is the only female player in the top 100. The fact that the best female player is an outlier in her population, combined with the problem of high variability of the extreme value, means that conclusions drawn on the basis of the performance of the top player alone may not be applicable to top players in general. The FIDE rating list used by Charness and Gerchak only reports players of average strength and above. We used the German rating list that lists all players. Most importantly, instead of estimating just the top male and female player, we estimated the expected performance of the best 100 male and female performers.

We applied this method to the population of all German players recorded by the German chess federation (Deutscher Schachbund). With over 3000 rated tournaments in a year, the German chess federation is one of the largest and the best organized national chess federations in the world. Given that almost all German tournaments are rated, including events such as club championships, all competitive and most hobby players in Germany can be found on the rating list. The rating itself is based on the same assumptions as the Elo rating used by the international chess federation. The two correlate highly ($r = 0.93$).

We considered the players in the list published in April 2007. A small number of the best male and female chess players from all over the world participate in tournaments rated by the German chess federation and dominate the top of the rating list. We have excluded all foreign players from the analyses so that our conclusions about the expected performance of the best male and female players, based on the total number and performance of male and female German players, are only applied to German players. Figure 1 shows the distribution of ratings for the German chess population. The distribution is approximately normal with mean of 1461 and s.d. of 342. Rated men (113 386) greatly outnumber rated women (7013); that is, there are 16 male chess players for every woman.

3. RESULTS

Figure 2 shows the real difference in rating for each of the top 100 pairs of male and female players, and the difference to be expected for each pair given the much larger number of male players. The expected superiority of male players varies from approximately 270 Elo points for the best male player to approximately 440 Elo points for the 100th. Figure 2 shows that, in fact, the top three women are better than would be expected. The next 70 pairs show a small but consistent advantage for men—their superiority over the corresponding female player is a little greater than would be expected purely from the relative numbers of male and female players. From approximately the 80th pair the advantage shifts. The female players are slightly better than would be expected. Averaged over the 100 top players, the expected male superiority is 341 Elo points and the real one is 353 points. Therefore 96 per cent of the observed difference between male and female players can be attributed to a simple statistical fact—the extreme values from a large sample are likely to be bigger than those from a small one.

4. DISCUSSION

Chess has long been renowned as the intellectual activity *par excellence* (Newell *et al.* 1958) and male dominance at chess is frequently cited as an example of innate male intellectual superiority (e.g. Howard 2005; Irwing & Lynn 2005). The reason seems obvious—the best male players are indisputably better than the best female players. For example: not a single woman has been world champion; only 1 per cent of Grandmasters, the best players in the world, are female; and there is only one woman among the best 100 players in the world. When considering such a seemingly convincing example of real world male superiority, one can easily forget to consider the great disparity in the number of participants and the statistical consequences of this for the probable gender of the best players.
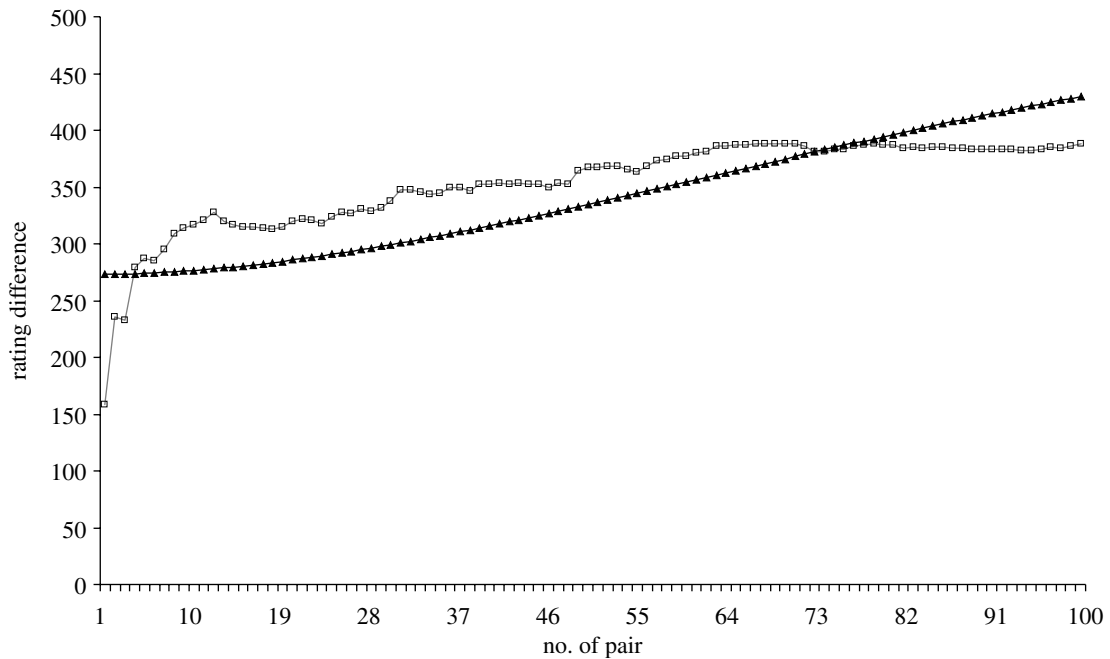
Figure 2. The differences between the real ratings of the best 100 female and male chess players and the differences expected on the basis of the common distribution of male and female ratings and the number of male and female players. The expected difference was obtained by subtracting the estimated rating for the nth female from the estimated rating for the nth male. The ratings were estimated using the participation rates of men and women and the parameters of their shared population (mean and s.d.). Black triangles, expected differences; white squares, real differences.

This was the case when the chess portal ChessBase asked some of the best female players to explain male dominance in chess (Ahmadov 2007). None of the interviewed women even mentioned the greatly differing participation rates and its consequences on the probable gender of the top performers. Similarly, at a recent gathering of more than 20 experts on gender difference to discuss the reasons for the paucity of women at the top of science, a broad range of reasons was discussed, but there was no mention of participation rates (Ceci & Williams 2007).

One way to avoid the conclusion based on participation rates would be to argue that the base participation rate for women used in this study underestimates the real participation rate. It is possible that there is a self-selection process based on the innate biological differences in intellectual abilities, and that the effects of this self-selection are already observable in the rating list we used. Women may be inferior in the intellectual abilities that are important for successful chess playing. This innate disadvantage may lead women to give up on chess in greater numbers than more successful men. The small number of women is then a consequence of their greater drop-out, which in turn is produced by their innate lack of the intellectual abilities required to succeed at chess. Differential participation rates may explain the discrepancy at the top, but the difference is itself a direct product of innate differences in intellectual abilities.

This argument sounds reasonable but it rests on a controversial assumption. It requires that there should be innate differences between men and women in the intellectual abilities required for success at chess. The topic of gender differences in cognitive abilities is a hotly debated one, which lacks conclusive evidence (for example, Geary 1998; Kimura 1999; Kerkman *et al.* 2000; Pinker 2002; Spelke 2005; Summers 2005;

Lachance & Mazzocco 2006; Ceci & Williams 2007). Even if such differences exist, it is unclear which, if any, intellectual abilities are associated with chess skill (for a recent review, see Bilalić *et al.* 2007). Whatever the final resolution of these debates, there is little empirical evidence to support the hypothesis of differential drop-out rates between male and females. A recent study of 647 young chess players, matched for initial skill, age and initial activity found that drop-out rates for boys and girls were similar (Chabris & Glickman 2006). Our study does not deal directly with the reasons why there are so few women in competitive chess. These may have to do with selective drop-out before tournament play starts in the early stages of learning to play chess. We can speculate about the reasons for low participation rates of women in competitive intellectual endeavours (as is often done, e.g. Steele 1997; Benbow *et al.* 2000; Kerkman *et al.* 2000; Massa *et al.* 2005; Spelke 2005; Summers 2005; Lachance & Mazzocco 2006; Andreescu *et al.* 2008) but empirical evidence is scarce.

This study demonstrates that the great discrepancy in the top performance of male and female chess players can be largely attributed to a simple statistical fact—more extreme values are found in larger populations. Once participation rates of men and women are controlled for, there is little left for biological, environmental, cultural or other factors to explain. This simple statistical fact is often overlooked by both laypeople and experts. In other domains such as science and engineering, where the predominance of men at the top is offered as evidence of the biological superiority of men, large differences between the number of women and men engaged in these activities are evident (Long 2001; Xie & Shauman 2003). In these areas of life, it is not possible to estimate the performance of the top women and men and their participation rates as precisely as it is in chess. But until
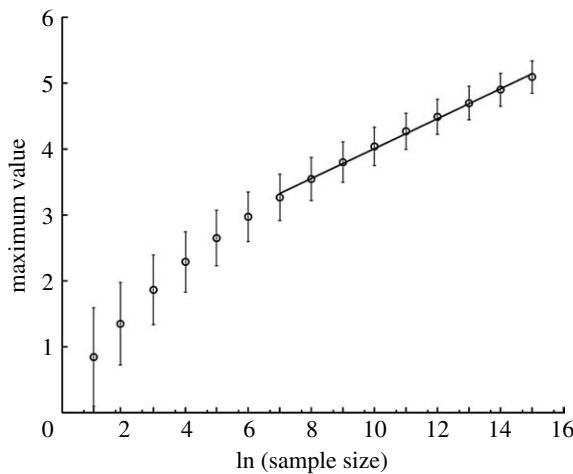
Figure 3. Plot demonstrating that the expected value of the best performer increases linearly with the logarithm of sample size, if the sample is large and drawn from a standard normal distribution. Each data point represents the mean value and s.d. of 1000 samples. The linear fit is shown using all populations that have 1000 or more members (ln $1000 \approx 7$).

the effect of participation rates has been allowed for, the greater number of men among the most successful people should not be cited as evidence of innate differences between male and female intellectual abilities.

APPENDIX A

A.1 *The expected difference between the k*th *best male and female players*

Standard methods (e.g. David & Nagaraja 2003) allow exact calculation of the expected score of the kth best player, but this proves numerically difficult for large sample sizes (such as ours, where $n = 120\ 399$). Here we use the size of the population to our advantage, deriving a simple approximation valid for large n.

Suppose we have a population of independent, identically distributed variables; in our case, the sample is all ratings from the German chess federations for both genders. We sort and relabel the sample such that $X_{n,n} \leq \cdots \leq X_{n,1}$; that is, the notation $X_{n,k}$ denotes the kth best score from the sample, which has size n. For example, $X_{n,1}$ and $X_{n,n}$ denote the best and worst ratings, respectively, in our sample. Now defining $E_{n,k} = E(X_{n,k})$, the expected value (or mean) of $X_{n,k}$, we have the recurrence relationship (Harter 1961)

$$E_{n,k+1} = \frac{1}{k}(nE_{n-1,k} - (n-k)E_{n,k}),$$

which is equivalent to

$$E_{n,k+1} = \frac{n!}{k!(n-k-1)!} \sum_{m=n-k}^{n} (-1)^{k+m-n} \binom{k}{n-m} \frac{E_{m,1}}{m}.$$

Thus we see that the expected value of the $(k+1)$th best performer of a sample size n may be written as a sum of terms of the form $E_{m,1}$—the best performer from a sample of size m—where the indexing variable takes values between $n-k$ and n.

Suppose that $X_i \sim N(0,1)$ are drawn from a standard normal distribution. If the sample size is large, then the

expected value of the best performer increases linearly with the logarithm of sample size, i.e. $E_{m,1} \approx c_1 + c_2 \ln m$, where $c_1 = 1.25$ and $c_2 = 0.287$ (figure 3).

If $n-k$ is large, then each value taken by m in the previous sum will be large. Hence, we may apply the log-linear approximation to each occurrence of $E_{m,1}$, obtaining

$$E_{n,k+1} \approx c_1 + c_2 \frac{n!}{k!(n-k-1)!} \sum_{m=n-k}^{n} (-1)^{k+m-n} \binom{k}{n-m}$$

$$\times \frac{\ln m}{m} = c_1 + c_2 \frac{n!}{k!(n-k-1)!} (-1)^k \nabla^k f(n).$$

Here, $f(x) = \ln x / x$ and $\nabla g(x) = g(x) - g(x-1)$ denotes the backward difference operator (Abramowitz & Stegun 1972). Furthermore, the backward difference operator is a close approximation of the differential operator, and hence

$$\nabla^k f(x) \approx \frac{d^k}{dx^k} f(x) = (-1)^k k! \frac{\ln x - H(k)}{x^{k+1}},$$

where $H(k) = \sum_{j=1}^{k} j^{-1}$ denotes kth harmonic number and we define $H(0) = 0$. Substituting, we find that the expected value of the kth best performer is given by

$$E_{n,k} \approx c_1 + c_2 \frac{n!}{(n-k)!n^k} (\ln n - H(k-1)).$$

This holds if $X_i \sim N(0,1)$ are drawn from a standard normal distribution. Moving to a general normal distribution, if $X_i \sim N(\mu, \sigma^2) = \mu + \sigma N(0,1)$, we find

$$E_{n,k} \approx (\mu + c_1\sigma) + c_2\sigma \frac{n!}{(n-k)!n^k} (\ln n - H(k-1)).$$

Given a distribution with known mean μ and s.d. δ, this final formula defines the expectation of the kth highest value within a sample of size n, valid provided n is large and k is relatively small. As such, it affords us a method for estimating the expected rating of a range of top players from the German chess data for each gender; indeed, we use the formula to calculate the expected ratings of the top 100 male and female players using the mean and s.d. of the population (the German chess data), in turn allowing us to determine the expected difference in rating between those players.

REFERENCES

Abramowitz, M. & Stegun, I. A. 1972 *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* New York, NY: Dover.

Ahmadov, Z. 2007 Women in chess—a matter of opinion. Retrieved 26 April 2008 from http://www.chessbase.com/newsdetail.asp?newsid=4132.

Andreescu, T., Gallian, J. A., Kane, J. M. & Mertz, J. E. 2008 Cross-cultural analysis of students with exceptional talent in mathematical problem solving. *Notices Am. Math. Soc.* **55**, 1248–1260.

Ayalon, H. 2003 Women and men go to university: mathematical background and gender differences in choice of field in higher education. *Sex Roles* **48**, 277–290. (doi:10.1023/A:1022829522556)

Benbow, C. P., Lubinski, D., Shea, D. L. & Eftekhari-Sanjani, H. 2000 Sex differences in mathematical reasoning ability at age 13: their status 20 years later. *Psychol. Sci.* **11**, 474–480. (doi:10.1111/1467-9280.00291)

Bilalić, M., McLeod, P. & Gobet, F. 2007 Does chess need intelligence?—a study with young chess players. *Intelligence* **35**, 457–470. (doi:10.1016/j.intell.2006.09.005)

Ceci, J. S. & Williams, M. W. 2007 *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: American Psychological Association.

Chabris, C. F. & Glickman, M. E. 2006 Sex differences in intellectual performance: analysis of a large cohort of competitive chess players. *Psychol. Sci.* **17**, 1040–1046. (doi:10.1111/j.1467-9280.2006.01828.x)

Charness, N. & Gerchak, Y. 1996 Participation rates and maximal performance. *Psychol. Sci.* **7**, 46–51. (doi:10.1111/j.1467-9280.1996.tb00665.x)

David, H. A. & Nagaraja, H. N. 2003 *Order statistics*. New York, NY: Wiley.

Davidson, M. J. & Cooper, C. L. 1992 *Shattering the glass ceiling—the woman manager*. London, UK: Paul Chapman Publishing.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M. & Whalley, L. J. 2003 Population sex differences in IQ at age 11: the Scottish mental survey 1932. *Intelligence* **31**, 533–542. (doi:10.1016/S0160-2896(03)00053-9)

Elo, A. E. 1986 *The rating of chessplayers, past and present*. New York, NY: Arco.

Geary, D. C. 1998 *Male, female: the evolution of human sex differences*. Washington, DC: American Psychological Association.

Glickman, M. E. 1999 Parameter estimation in large dynamic paired comparison experiments. *Appl. Stat.* **48**, 377–394. (doi:10.1111/1467-9876.00159)

Glickman, M. E. & Chabris, C. F. 1996 *Using chess ratings as data in psychological research*. Boston, MA: Department of Mathematics and Statistics, Boston University. Unpublished manuscript, available from www.wjh.harvard.edu/cfc/Glickman1996.pdf

Harter, H. L. 1961 Expected values of normal order statistics. *Biometrika* **48**, 151–165. (doi:10.1093/biomet/48.1-2.151)

Howard, R. W. 2005 Are gender differences in high achievement disappearing? A test in one intellectual domain. *J. Biosoc. Sci.* **37**, 371–380. (doi:10.1017/S0021932004006868)

Huffman, L. M. & Torres, L. 2002 It's not only 'who you know' that matters: gender, personal contacts, and job lead quality. *Gend. Soc.* **16**, 793–813. (doi:10.1177/089124302237889)

Irwing, P. & Lynn, L. 2005 Sex differences in means and variability on the progressive matrices in university students: a meta-analysis. *Br. J. Psychol.* **96**, 505–524. (doi:10.1348/000712605X53542)

Kerkman, D. D., Wise, J. C. & Harwood, E. A. 2000 Impossible 'mental rotation' problems: a mismeasure of women's spatial abilities? *Learn. Individ. Differ.* **12**, 253–269. (doi:10.1016/S1041-6080(01)00039-5)

Kimura, D. 1999 *Sex and cognition*. Cambridge, MA: MIT Press.

Lachance, J. A. & Mazzocco, M. M. 2006 A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learn. Individ. Differ.* **16**, 195–216. (doi:10.1016/j.lindif.2005.12.001)

Long, S. J. 2001 *From scarcity to visibility: gender differences in the careers of doctoral scientists and engineers*. Washington, DC: National Academy Press.

Massa, L. J., Mayer, R. E. & Bohon, L. M. 2005 Individual differences in gender role beliefs influence spatial ability test performance. *Learn. Individ. Differ.* **15**, 99–111. (doi:10.1016/j.lindif.2004.11.002)

Newell, A., Shaw, J. C. & Simon, H. A. 1958 Chess-playing programs and the problem of complexity. *IBM J. Res. Dev.* **2**, 320–335.

Pinker, S. 2002 *The blank slate: the modern denial of human nature*. New York, NY: Viking.

Spelke, E. S. 2005 Sex differences in intrinsic aptitude for mathematics and science? A critical review. *Am. Psychol.* **60**, 950–958. (doi:10.1037/0003-066X.60.9.950)

Steele, C. M. 1997 A threat in the air: how stereotypes shape intellectual identity and performance. *Am. Psychol.* **52**, 613–629. (doi:10.1037/0003-066X.52.6.613)

Summers, L. H. 2005 Remarks at NBER conference on diversifying the science and engineering workforce. Retrieved 26 April 2008 from www.president.harvard.edu/speeches/2005/nber.html.

Xie, Y. & Shauman, K. A. 2003 *Women in science: career processes and outcomes*. Cambridge, MA: Harvard University Press.