

Comment

Are participation rates sufficient to explain gender differences in chess performance?

In a recent paper in the *Proceedings*, Bilalić et al. (2009) argued that different participation rates of men and women represent the key factor that has to be taken into account when the comparatively small number of women at the top level of certain intellectually demanding activities needs to be explained. Their conclusion was based on the results of an analysis of ratings of German chess players. According to Bilalić et al. (2009), 96 per cent of the observed differences in performance between the top 100 pairs of male and female players could be attributed to differential participation rates. The first purpose of this comment is to argue that their conclusion was premature and caused by an inappropriate statistical approach. The second purpose is to propose a more adequate method of analysis and to show that participation rates only explain two-thirds of the observed differences.

Bilalić et al. (2009) assumed that the ratings of German chess players are realizations of normally distributed random variables. Then, they calculated approximately the expected rating of the kth best male and the kth best female player. Fig. 2 of their paper contained the differences of these two values for k = 1, ..., 100. What these authors did not mention, however, is that their model predicts a rating of 3031 for the best male German player and a rating above 2700 for the 16 best male German players. Currently, there are only 33 players in the world with a rating above 2700 and there is no German belonging to this elite group. The highest rating ever achieved by a human player is 2851, which is significantly lower than the expected rating of 3031 predicted for the best German player according to the model of Bilalić et al. (2009). Therefore, this model seems inadequate to describe the upper tail of the distribution of ratings of German chess players.

I will now describe an analytical approach that does not rely on the questionable assumption of a normal distribution for the rating. Assume there are $n_{\rm f}$ female players and $n_{\rm m}$ male players, and let R_k denote the rank of the kth best female player in the ordered combined list of male and female players. Under the assumption that gender has no effect on rating performance, it follows that the distribution of R_k is a negative hypergeometric distribution (Johnson & Kotz

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2009.2257 or via http://rspb.royalsocietypublishing.org.

1969), i.e.:

$$P(R_{k} = s) = \frac{\binom{n_{\rm f}}{k-1} \cdot \binom{n_{\rm m}}{s-k}}{\binom{n_{\rm f} + n_{\rm m}}{s-1}} \cdot \frac{n_{\rm f} - k + 1}{n_{\rm f} + n_{\rm m} - s + 1}$$

$$= \frac{\binom{s-1}{k-1} \cdot \binom{n_{\rm m} + n_{\rm f} - s}{n_{\rm f} - k}}{\binom{n_{\rm f} + n_{\rm m}}{n_{\rm f}}},$$

$$(1)$$

for $k \le s \le n_m + k$. The expected value is $ER_k = k$. $(n_{\rm m}+n_{\rm f}+1)/(n_{\rm f}+1)$ and it is straightforward to calculate the 0.05 per cent quantile $r_{k,0.0005}$ and the 99.95 per cent quantile $r_{k,0,9995}$ for this distribution. If there is no gender effect on rating performance, then with probability of at least 99.9 per cent, the rank R_k of the kth best female player would be between $r_{k,0.0005}$ and $r_{k,0.9995}$. Figure 1 compares the observed rank r_k of the kth best female player with its expected value and with the quantiles $r_{k,0.0005}$ and $r_{k,0.9995}$. The discrepancy is evident. With the exception of the best female German player (even for her, the observed rank $r_1 = 87$ is considerably above her expected rank of $ER_1 = 21$ but at least is within the interval $[r_{1,0.0005}, r_{1,0.9995}] = [1,157]$), the observed ranks of the best 100th female players are above their 99.9% confidence intervals. For example, for the 100th best female player, the observed rank is 5505, whereas her expected rank is only 2116 and, assuming no gender effect, her rank is between $r_{100,0.0005} = 1510$ and $r_{100,0.9995} = 2849$ with a probability of at least 99.9 per cent.

Perfect agreement between the observed rank and the rank expected under the assumption that gender has no effect on rating performance for the kth best female player would occur if she possessed rating f_k of the round(ER_k)th best player in the combined list (where round(ER_k) denotes rounding ER_k to the nearest integer). Analogously, let ER_k^* denote the expected rank of the kth best male player and m_k the rating of the round(ER_k^*k)th best player in the combined list. Then, $d_k = m_k - f_k$ may be considered to represent that part of the rating difference between the kth best male and the kth best female player, which can be attributed to differential participation rates of men and women. Figure 2 compares these differences d_k with the differences between the actual ratings of the best 100 female and male players. Only between

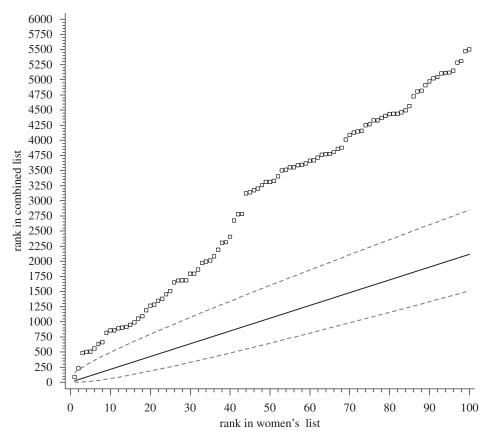


Figure 1. Observed and expected rank of the best 100 female chess players. Squares represent the observed rank of the kth best female player in the combined list of male and female players. The unbroken line gives the expected rank ER_k and the two dotted lines denote the quantiles $r_{k,0.0005}$ and $r_{k,0.9995}$.

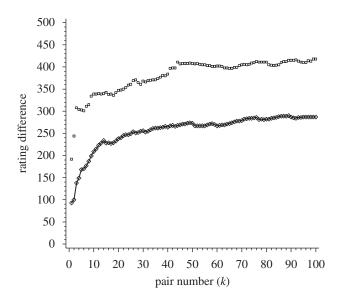


Figure 2. The differences between the actual ratings of the best 100 female and male German chess players and the differences d_k attributable to different participation rates. Squares, real differences; diamonds, attributable differences.

41 and 71.1 per cent (mean value: 66.9%) of the actual rating differences are explained by different participation rates of men and women, which is substantially lower than the 96 per cent obtained by Bilalić *et al.*

(2009). The unexplained gap between the two curves varies between 99 and 170 rating points (mean value over 100 pairs: 124.5). If two players with a rating difference of 124.5 points compete in a match over 100 games, the expected result is 67: 33 in favour of the higher rated player. Therefore, the conclusion of Bilalić *et al.* (2009) that 'there is little left for biological or cultural explanations to account for', appears to be premature.

I am grateful to Karen Hirschmann for pointing my attention to the work of Bilalić *et al.* (2009).

Michael Knapp*
Institute for Medical Biometry, Informatics and Epidemiology,
University of Bonn, Sigmund-Frud-Str. 25, 53105 Bonn,
Germany
*knapp@uni-bonn.de

REFERENCES

Bilalić, M., Smallbone, K., McLeod, P. & Gobet, F. 2009 Why are (the best) women so good at chess? Participation rates and gender differences in intellectual domains. *Proc. R. Soc. B* **276**, 1161–1165. (doi:10.1098/rspb. 2008.1576)

Johnson, N. L. & Kotz, S. 1969 Discrete distributions. New York, NY: Wiley.