



This is a repository copy of *Restricting range restricts conclusions*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/159116/>

Version: Published Version

Article:

Vaci, N. orcid.org/0000-0002-8094-0902, Gula, B. and Bilalić, M. (2014) Restricting range restricts conclusions. *Frontiers in Psychology*, 5. 569.

<https://doi.org/10.3389/fpsyg.2014.00569>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Restricting range restricts conclusions

Nemanja Vaci, Bartosz Gula and Merim Bilalić*

Department of General Psychology and Cognitive Science, Institute of Psychology, Alpen-Adria University Klagenfurt, Klagenfurt, Austria

*Correspondence: merim.bilalic@aau.at

Edited by:

David Zachary Hambrick, Michigan State University, USA

Reviewed by:

Fred Oswald, Rice University, USA

Fernand Gobet, University of Liverpool, UK

Keywords: expertise, skill acquisition, chess, Elo rating, gender differences, gerontology, talent

THE EXPERTISE APPROACH AND SKILL ACQUISITION

Research on expertise is by definition focused on a restricted sample of individuals. Experts are people who consistently produce outstanding performance in their domains (Ericsson, 2006) and as such are without exception located on the positive side of the skill distribution. The usual approach in the study of expertise is to compare the extreme group of the skill distribution, experts, with the extreme group at the other end, that of novices. This contrasting approach, which we have called the “expertise approach” (Bilalić et al., 2010, 2012), has a long tradition (Chase and Simon, 1973; Simon and Chase, 1973; De Groot, 1978; Preacher et al., 2005). Its main advantage over the common approach in cognition, where all participants are at the same skill level, is the presence of a control group of novices that enables falsification of results obtained on experts (Wason, 1960; Kuhn, 1970; Campitelli and Spelman, 2013). In that way, the expertise approach is not unlike the neuropsychological approach that contrasts results obtained on patients with the results of “normal” participants (Shallice, 1988).

The main goal of the expertise approach is to provide evidence relating to the cognitive and neural mechanisms behind processes such as object and pattern recognition, which would be difficult to obtain from subjects who possess approximately the same level of expertise. The skill acquisition process, which is one of the main topics of expertise (William and Harter, 1899), is of secondary importance in the expertise approach. This is understandable as the contrast between

experts and novices captures only the beginning and the end product of the skill acquisition process. It is unrealistic to follow people for the length of time required in order to achieve expertise in a given domain. However, expertise researchers have recently started employing an archival approach that provides a more complete picture of the skill acquisition process (Charness and Gerchak, 1996; Chabris and Glickman, 2006; Howard, 2008, 2009; Bilalić et al., 2009). In the game of chess, a domain commonly studied in expertise research, there are precise records of all practitioners from an early age (Howard, 2006a; Bilalić et al., 2009). These records include not only personal information such as gender and age, but also skill levels at different stages, numbers of games played, and corresponding skill change. The records provide a wealth of data for investigating the influence of factors such as age, gender, and even talent, on the skill acquisition process. Here we want to draw attention to the fact that some of the databases used in previous research only provide records of the very best practitioners. In the expertise approach, such restriction is an integral part of the methodology, but restricting the range of population in the archival approach could have grave consequences for the conclusions about the nature of skill acquisition.

DIFFERENT DATABASES, DIFFERENT CONCLUSIONS

One of the advantages of chess as a domain is that there is an objective and reliable measure of skill. Skill is measured on an interval scale that reflects the performance of a player against other players. The Elo

rating, named after Arpad Elo who introduced the scale as a measure of chess skill (Elo, 1978), is measured in the same way all over the world. A beginner is supposed to have 600–800 Elo points, average players about 1500 Elo points, master players above 2200 Elo points, while the very best players, called grandmasters, have ratings above 2500 Elo points. Expert players are considered to have above 2000 rating points.

The most frequently used database in skill acquisition studies is the database of the International Chess Federation, FIDE (for more information, see Howard, 2006a). This database, like other chess databases we will mention here, offers multiple advantages for skill acquisition research. Firstly, it gathers records from the 1970s to the present, and so it is possible to obtain trajectories of ratings over the course of players’ lives. Secondly, it represents the whole population of the very best players in the world. Thirdly, this database contains multiple measurement points from players, so it can be used to observe individual skill trajectories. Besides rating points, numerous other variables are recorded (e.g., number of games played, gender of participants, nationality, title, rating change, rating rank) which could be used for research purposes (Howard, 2006a). In other words, the FIDE database offers a fruitful basis for exploration and description of the multiple factors and processes behind chess skill acquisition.

For all its advantages, the FIDE database provides only the records of the very best players. Due to technical and logistical reasons, the FIDE database at the beginning logged only master level players

(above 2200 Elo). Only in the 1990s was the level lowered to expert level players (2000 Elo) and then in the last decade to the level of average players (1500 Elo and below). In other words, the worst players in the FIDE database are still average practitioners.

The ideal situation would be to have records of all players from the very beginning of their careers, not only when they reach a particular level of expertise. In this scenario, the database would also encompass people who for whatever reasons do not become experts. Fortunately, there are such databases. National databases, such as the databases of the German Chess Federation and the United States Chess Federation (USCF), keep records of all their members and thus represent the whole population of competitive (national) players. They provide all the information the FIDE database offers without restricting the range of skill.

If one is trying to examine factors that influence skill acquisition, a database that contains only average and above-average players should not be the starting point

of investigations. As is well known in the field of statistics, the conclusions obtained on data with restricted range could be misleading especially if they are not obtained by appropriate analyses (Long, 1997; Sackett and Yang, 2000). The best possible way to examine differences in effects that researchers could obtain while analyzing the data is through the comparison of effects made on different databases. Here we illustrate the possible pitfalls of skill range restriction by comparing distributions of ratings from a database with restricted range (FIDE) and a database with unrestricted range (German).

The FIDE and German databases contain similar number of practitioners, around 120,000 (see **Figure 1**). However, the ratings of the two distributions overlap only at the highest values of the German distribution and the lowest values of FIDE distribution. Not only are the mean and variance of both distribution vastly different, but also other parts of the distribution, such as quartiles, are also extremely different.

Restricting the database to the best players also has a consequence for the skill trajectories of players. One needs time to become an expert and the players in the FIDE database are older (37 years) than the players in the German database (32 years). The differences between the two databases are also evident when we compare typical skill trajectories. **Figure 2A** shows the FIDE players entering the database at around age 10, having already become competent players (rating of 1900 Elo), with a subsequent shallow increase to the peak at age 39. In contrast, the German players have a steeper increase, since they are entered the database as novices and learn faster at beginning skill level, as implied by the power law of practice (Newell and Rosenbloom, 1981), until the same peak at age 39. The decline in later years is also different in the two databases with FIDE players declining faster than their German counterparts.

One could say that the FIDE and German players have vastly different ratings that make the comparison between them difficult. One way around this

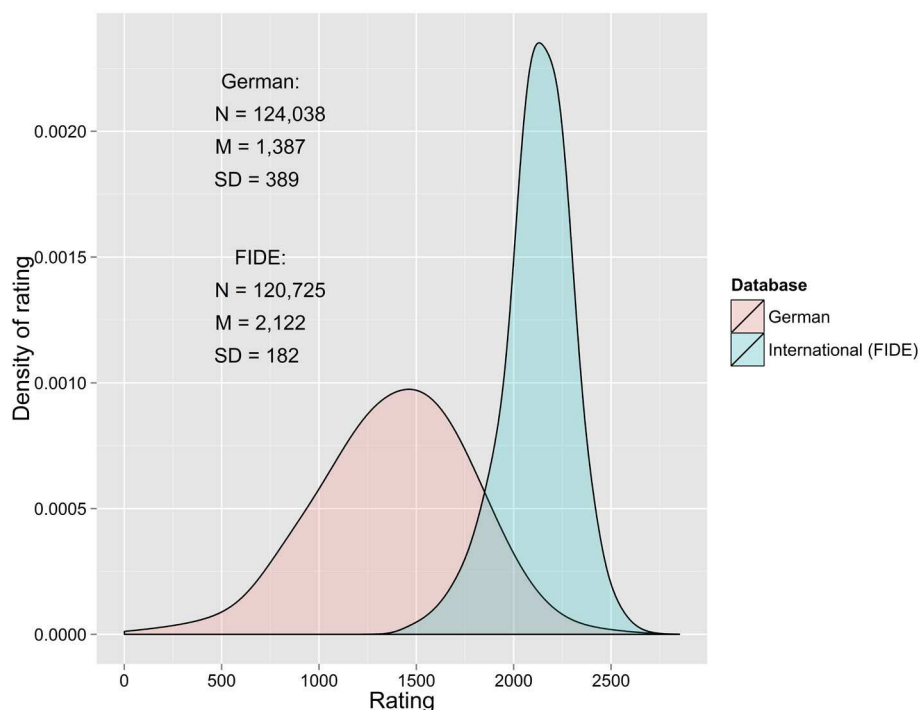
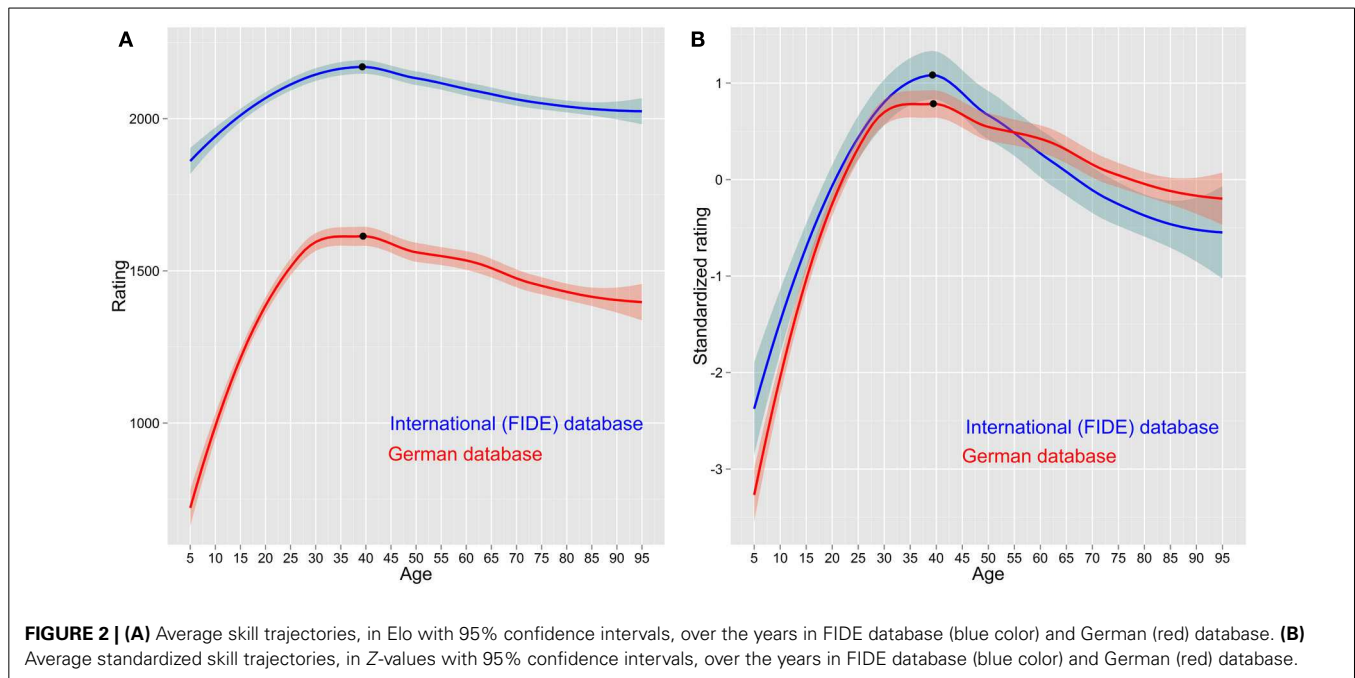


FIGURE 1 | Distribution of chess skill as measured by Elo rating in FIDE (blue color) and German (red) databases (as of 2008). The databases contain similar number of players, but differ vastly in the

distribution shape and coverage—the only overlap is at the highest values of the German database and lowest values of the FIDE database.



problem is to standardize ratings in each database separately and check if the skill trajectories are similar in both datasets. **Figure 2B** shows the standardized rating as a function of age. Again, on average FIDE players start at a higher skill level but improve more slowly. They also have a higher peak but their decline is so rapid that in the latter stages of their careers their standardized performance is lower than the standardized performance of their German colleagues.

EXPLAINING CONTRADICTIONARY RESULTS

We have demonstrated that there are vast differences between the databases commonly used in the archival approach to skill acquisition. The two datasets are completely different in the range of values as well as in the number of participants that are obtaining a particular rating. The restricted databases, such as FIDE, do not represent the whole skill range and may provide inadequate answers to the questions under investigation. The restriction of range and its consequences may also explain some of the inconsistencies and contradictory findings in the field.

For example, Roring and Charness (2007) used the FIDE databases to investigate age effects on skill acquisition. They demonstrated the peak age in chess skill

to be around 43 years, much later than previously proposed peak around 35 years (e.g., Howard, 2005). Another surprising result was the fact that the decline is steeper for initially lower rated participants than for higher rated participants. In other words, initially more able participants were declining significantly more slowly than their initially weaker colleagues. Our illustrations (**Figures 2A,B**) indicate that both peak age and declining rate are influenced by the range restriction in the FIDE database. The conclusion would be significantly different if a whole range database, such as the German database, were used.

To further illustrate possible consequences of the range restriction, we can consider the inconsistent findings in the research on gender differences in skill acquisition. It is notable that the studies using the restricted FIDE database regularly find gender differences in skill acquisition (Howard, 2005, 2006b, 2014 but see, Bilalić and McLeod, 2006, 2007). Furthermore, the studies using the national German and USCF databases (Chabris and Glickman, 2006; Bilalić et al., 2009) also noted the differences in the mean and highest ratings of women and men. However, using the unrestricted range and the full lifespan data, they observed that factors such as participation

rates and dropout rates could explain the differences. This kind of analysis is impossible with the FIDE database where dropouts are not recorded because the people concerned stopped playing chess before they achieved expert level.

Researchers using the FIDE database to investigate talent (Howard, 2008, 2009) face a similar problem. The time required to reach a certain level and the amount of practice (as measured by games played) may well provide clues about the different natural endowments of certain players. This in turn may allow us to speculate about different levels of talent. It is, however, impossible to make any certain conclusions if we lack the very first part of their skill acquisition process, as we do in the FIDE database. As with the gender factor in skill acquisition process, the differences in the early stages may as well overshadow the differences at the highest level. Similarly, the causes behind dropouts may remain unresolved because the data of the people who for whatever reasons stopped playing chess is not available.

Both the expertise and archival approaches are important vehicles for the investigation of expertise and cognition in general. The restricted range of focus in the expertise approach is a fundamental part of the methodology and an advantage over usual research on cognition. The

archival approach offers the possibility of capturing the full cycle of the long-term skill acquisition process. In this paper we have demonstrated that the results obtained on the restricted range do not necessarily generalize to the whole range of values. The effects obtained with restricted range cannot and should not be used to make inferences about the mechanisms and factors that influence skill acquisition. When we restrict our data, we restrict our conclusions too.

ACKNOWLEDGMENT

We would like to thank Frank Hoppe for providing the German database and Robert Gaschler for the FIDE database.

REFERENCES

- Bilalić, M., Langner, R., Erb, M., and Grodd, W. (2010). Mechanisms and neural basis of object and pattern recognition: a study with chess experts. *J. Exp. Psychol. Gen.* 139, 728–742. doi: 10.1037/a0020756
- Bilalić, M., and McLeod, P. (2006). How intellectual is chess?—a reply to Howard. *J. Biosoc. Sci.* 38, 419–421. doi: 10.1017/S0021932005001185
- Bilalić, M., and McLeod, P. (2007). Participation rates and the difference in performance of women and men in chess. *J. Biosoc. Sci.* 39, 789–793. doi: 10.1017/S0021932007001861
- Bilalić, M., Smallbone, K., McLeod, P., and Gobet, F. (2009). Why are (the best) women so good at chess? participation rates and gender differences in intellectual domains. *Proc. Biol. Sci.* 276, 1161–1165. doi: 10.1098/rspb.2008.1576
- Bilalić, M., Turella, L., Campitelli, G., Erb, M., and Grodd, W. (2012). Expertise modulates the neural basis of context dependent recognition of objects and their relations. *Hum. Brain Mapp.* 33, 2728–2740. doi: 10.1002/hbm.21396
- Campitelli, G., and Spelman, C. (2013). Expertise paradigms for investigating the neural substrates of stable memories. *Front. Hum. Neurosci.* 7:740. doi: 10.3389/fnhum.2013.00740
- Chabris, C. F., and 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., and Gerchak, Y. (1996). Participation rates and maximal performance: a log-linear explanation for group differences, such as Russian and male dominance in chess. *Psychol. Sci.* 7, 46–51. doi: 10.1111/j.1467-9280.1996.tb00665.x
- Chase, W. G., and Simon, H. A. (1973). Perception in chess. *Cogn. Psychol.* 4, 55–81.
- De Groot, A. (1978). *Thought and Choice in Chess, 2nd Edn.* The Hague: Mouton De Gruyter.
- Elo, A. E. (1978). *The Rating of Chessplayers, Past and Present*, Vol. 3. London: B.T. Batsford Ltd. Available online at: <http://www.getcited.org/pub/101876597>
- Ericsson, K. A. (2006). *The Cambridge Handbook of Expertise and Expert Performance.* New York, NY: Cambridge University Press.
- Howard, R. W. (2005). Objective evidence of rising population ability: a detailed examination of longitudinal chess data. *Pers. Individ. Dif.* 38, 347–363. doi: 10.1016/j.paid.2004.04.013
- Howard, R. W. (2006a). A complete database of international chess players and chess performance ratings for varied longitudinal studies. *Behav. Res. Methods* 38, 698–703. doi: 10.3758/BF03193903
- Howard, R. W. (2006b). IQ, visuospatial ability and the gender divide: a reply to Bilalić and McLeod. *J. Biosoc. Sci.* 38, 423–426. doi: 10.1017/S0021932005001197
- Howard, R. W. (2008). Linking extreme precocity and adult eminence: a study of eight prodigies at international chess. *High Abil. Stud.* 19, 117–130. doi: 10.1080/13598130802503991
- Howard, R. W. (2009). Individual differences in expertise development over decades in a complex intellectual domain. *Mem. Cognit.* 37, 194–209. doi: 10.3758/MC.37.2.194
- Howard, R. W. (2014). Gender differences in intellectual performance persist at the limits of individual capabilities. *J. Biosoc. Sci.* 46, 386–404. doi: 10.1017/S0021932013000205
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions.* Chicago, IL: University of Chicago Press.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*, Vol. 7. Thousand Oaks, CA: Sage.
- Newell, A., and Rosenbloom, P. S. (1981). “Mechanisms of skill acquisition and the law of practice,” in *Cognitive Skills and Their Acquisition*, ed J. R. Anderson (Hillsdale, NJ: Erlbaum), 1–55.
- Preacher, K. J., Rucker, D. D., MacCallum, R. C., and Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychol. Methods* 10, 178–192. doi: 10.1037/1082-989X.10.2.178
- Roring, R. W., and Charness, N. (2007). A multi-level model analysis of expertise in chess across the life span. *Psychol. Aging* 22, 291–299. doi: 10.1037/0882-7974.22.2.291
- Sackett, P. R., and Yang, H. (2000). Correction for range restriction: an expanded typology. *J. Appl. Psychol.* 85, 112–118. doi: 10.1037/0021-9010.85.1.112
- Shallice, T. (1988). *From Neuropsychology to Mental Structure.* New York, NY: Cambridge University Press.
- Simon, H. A., and Chase, W. G. (1973). Skill in chess: experiments with chess-playing tasks and computer simulation of skilled performance throw light on some human perceptual and memory processes. *Am. Sci.* 61, 394–403.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Q. J. Exp. Psychol.* 12, 129–140. doi: 10.1080/17470216008416717
- William, L. B., and Harter, N. (1899). Studies on the telegraphic language: the acquisition of a hierarchy of habits. *Psychol. Rev.* 6, 345–375. doi: 10.1037/h0073117

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 05 April 2014; accepted: 22 May 2014; published online: 12 June 2014.

Citation: Vaci N, Gula B and Bilalić M (2014) Restricting range restricts conclusions. *Front. Psychol.* 5:569. doi: 10.3389/fpsyg.2014.00569

This article was submitted to *Cognition*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Vaci, Gula and Bilalić. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.