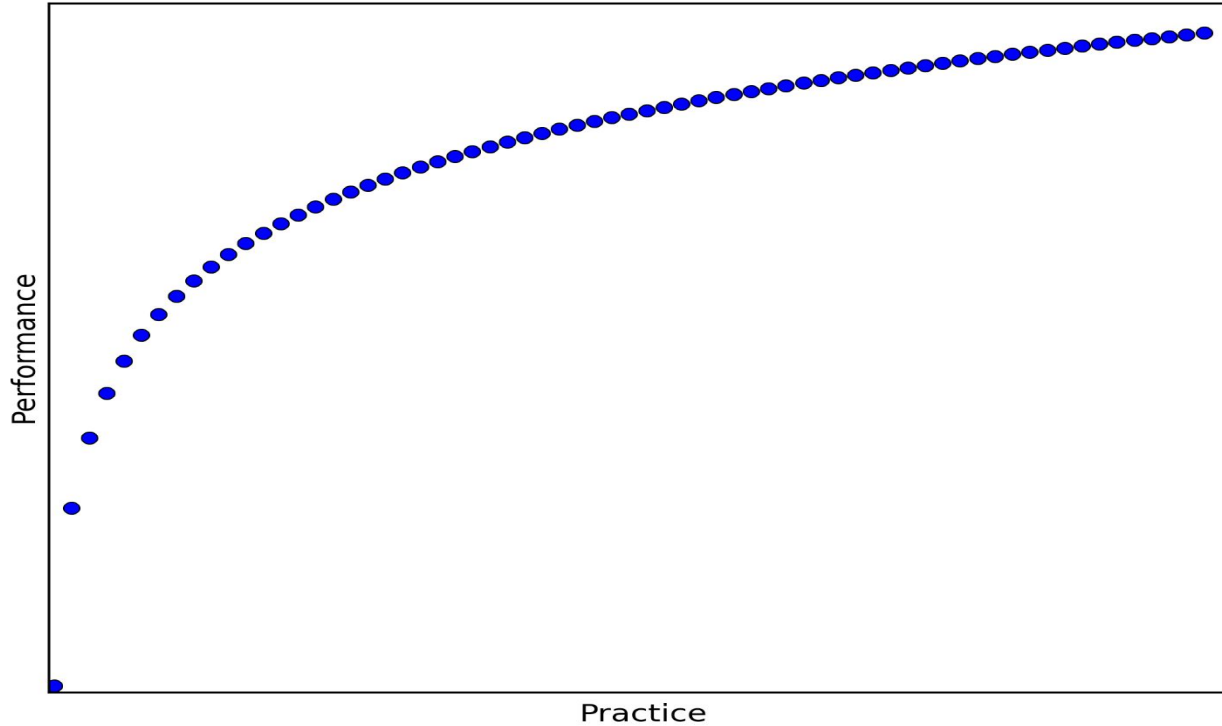
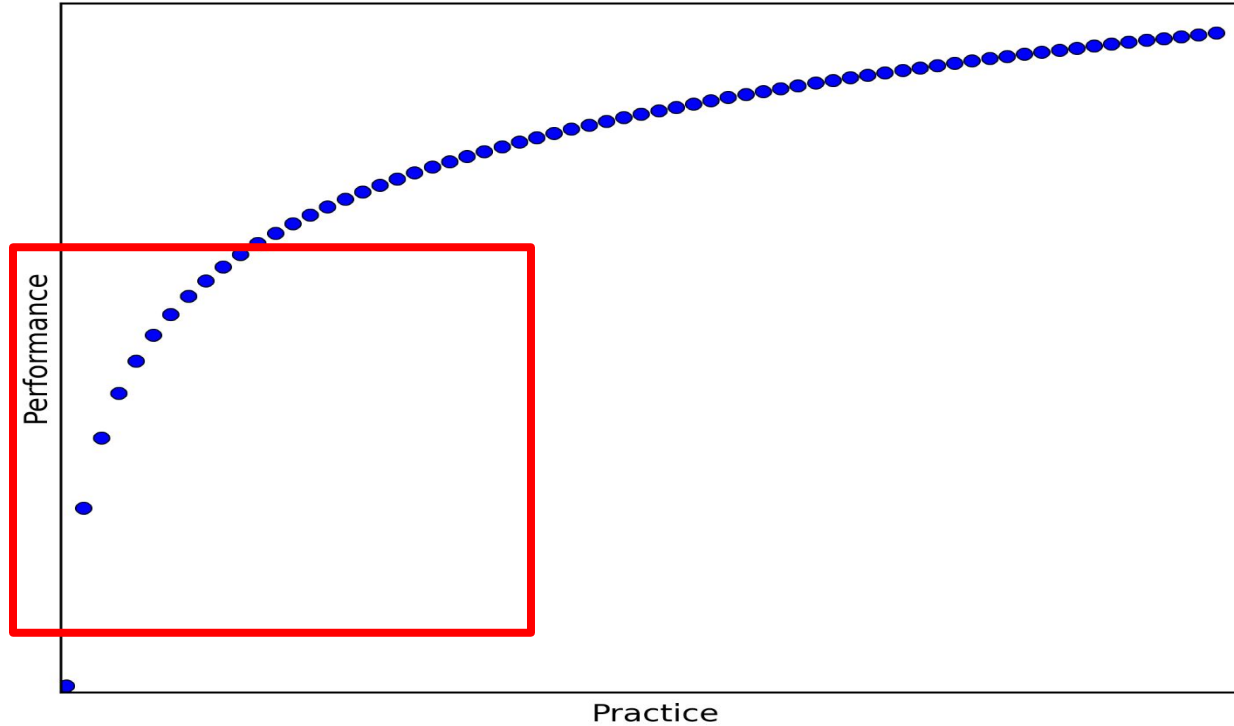

Learning Curves From Game Data

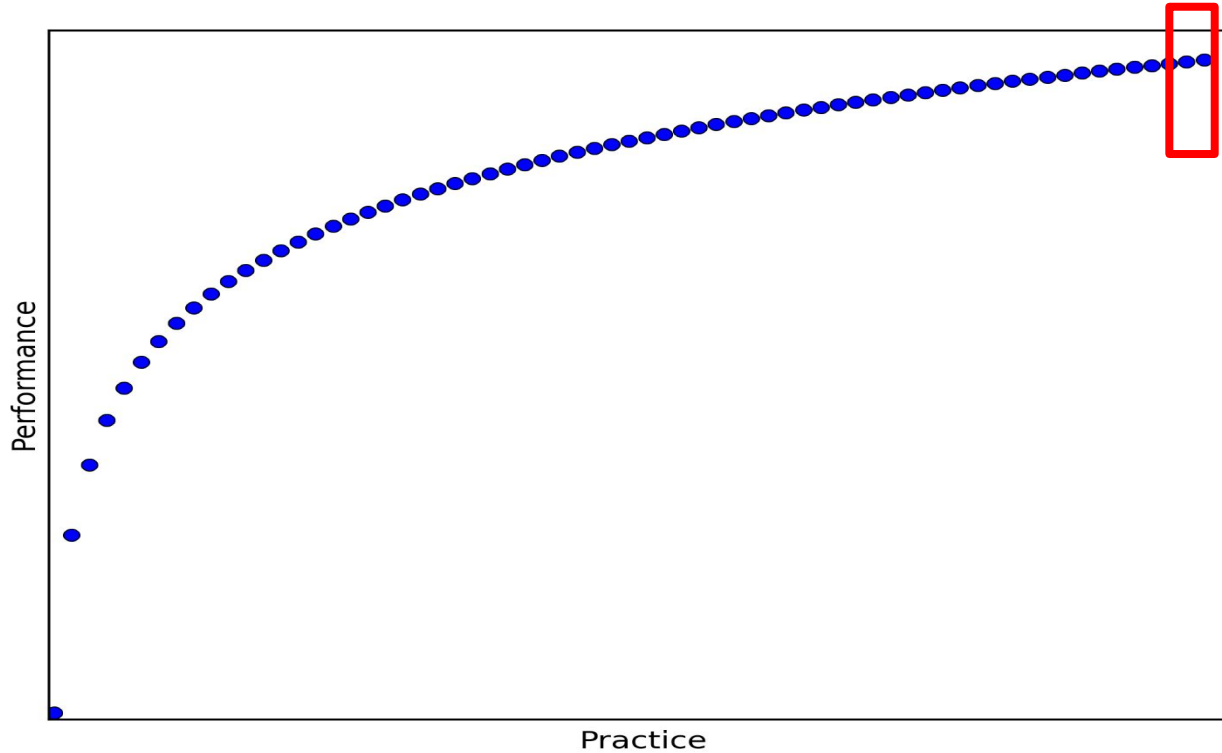
Tom Stafford, University of Sheffield
Beyond the Lab, 11th July 2017



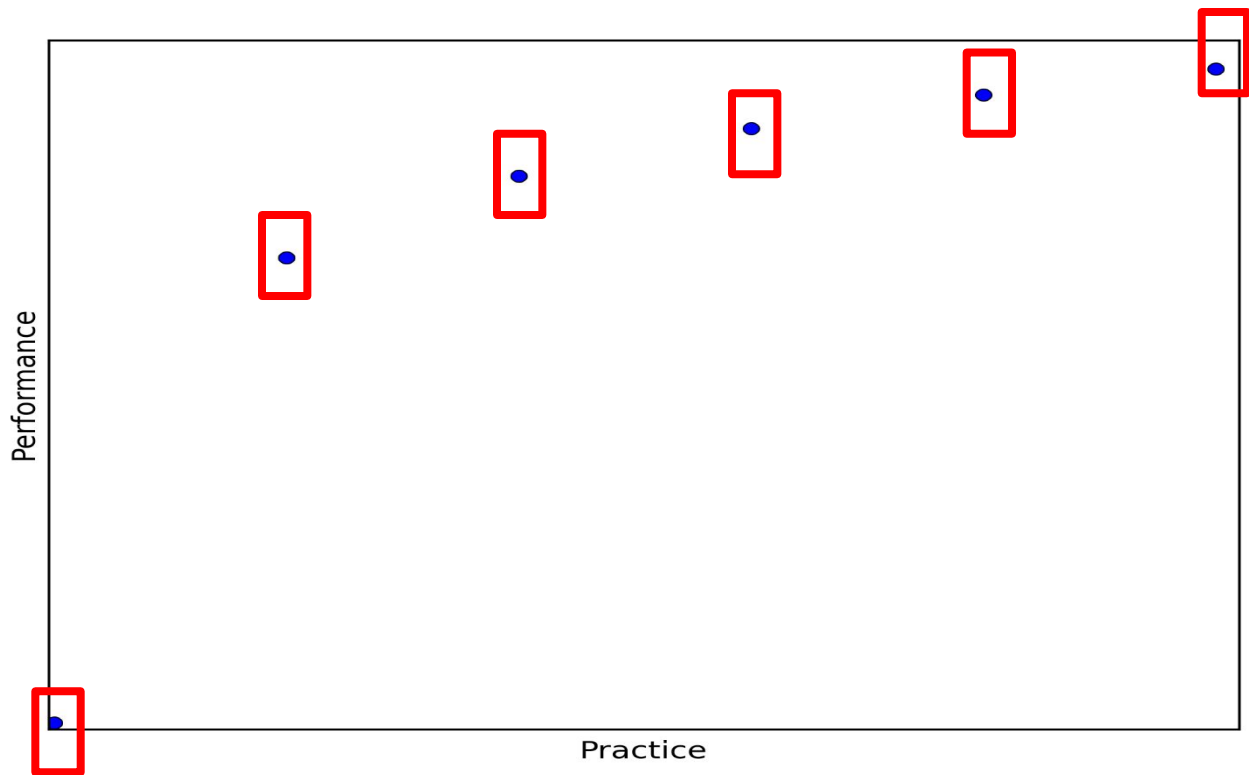
Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Lawrence Erlbaum



Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Lawrence Erlbaum



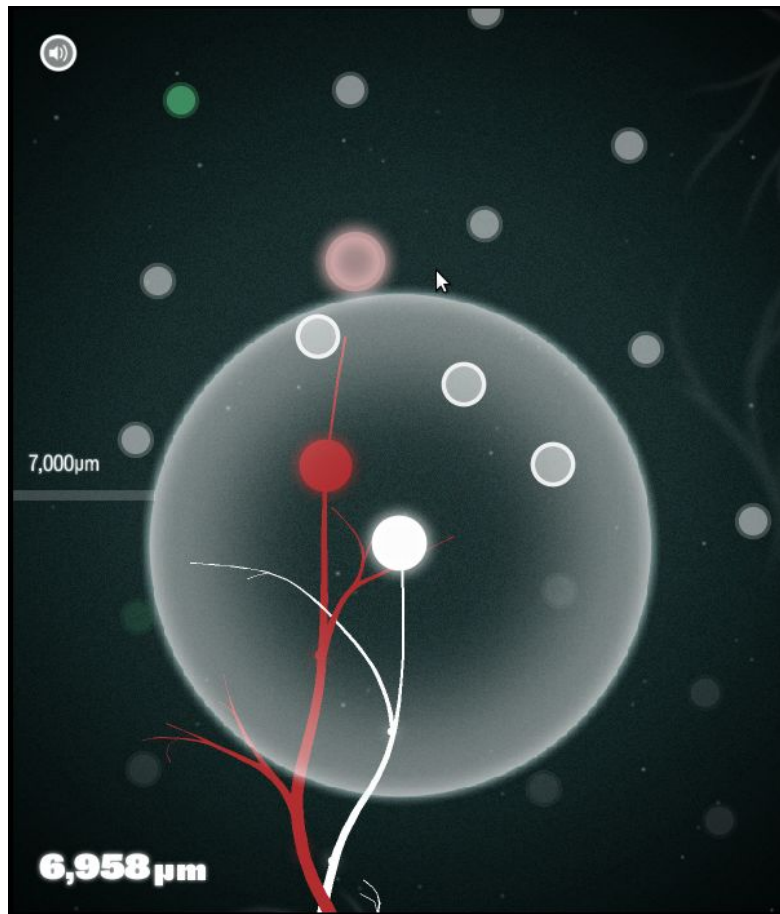
Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (p. 683-703). Cambridge University Press



Crossman, E. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, 2(2), 153-166

The image shows a screenshot of the ARIKA arcade game interface. On the left, a menu titled "KATANA" lists several options: "KATANA", "KATANA", "KATANA", "KATANA", "KATANA", "KATANA", "KATANA", "KATANA", "KATANA", "KATANA". In the center, there is a "TGM" (The Great Mission) section with a "TGM" button and a "TGM" button. Below this, there is a "LEVEL" section with a "LEVEL" button and a "LEVEL" button. On the right, a character selection screen shows a character with green hair and a yellow headband. The character is holding a sword. The background is dark with some text and graphics. At the bottom, there is a URL and the ARIKA logo.

<http://www.arika.co.jp> **ARIKA**



Stafford, T. & Dewar, M. (2014).
[Tracing the trajectory of skill learning with a very large sample of online game players.](#)
Psychological Science, 25(2)
511 - 518.



Mike Dewar

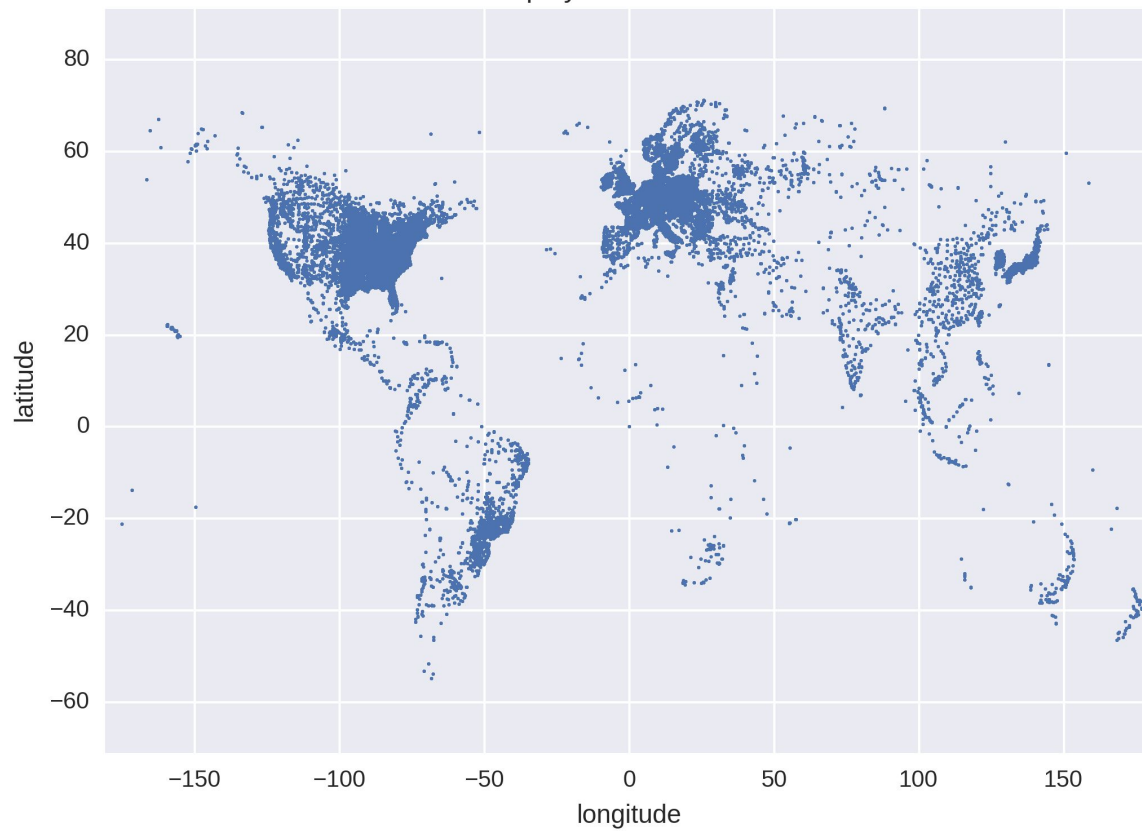


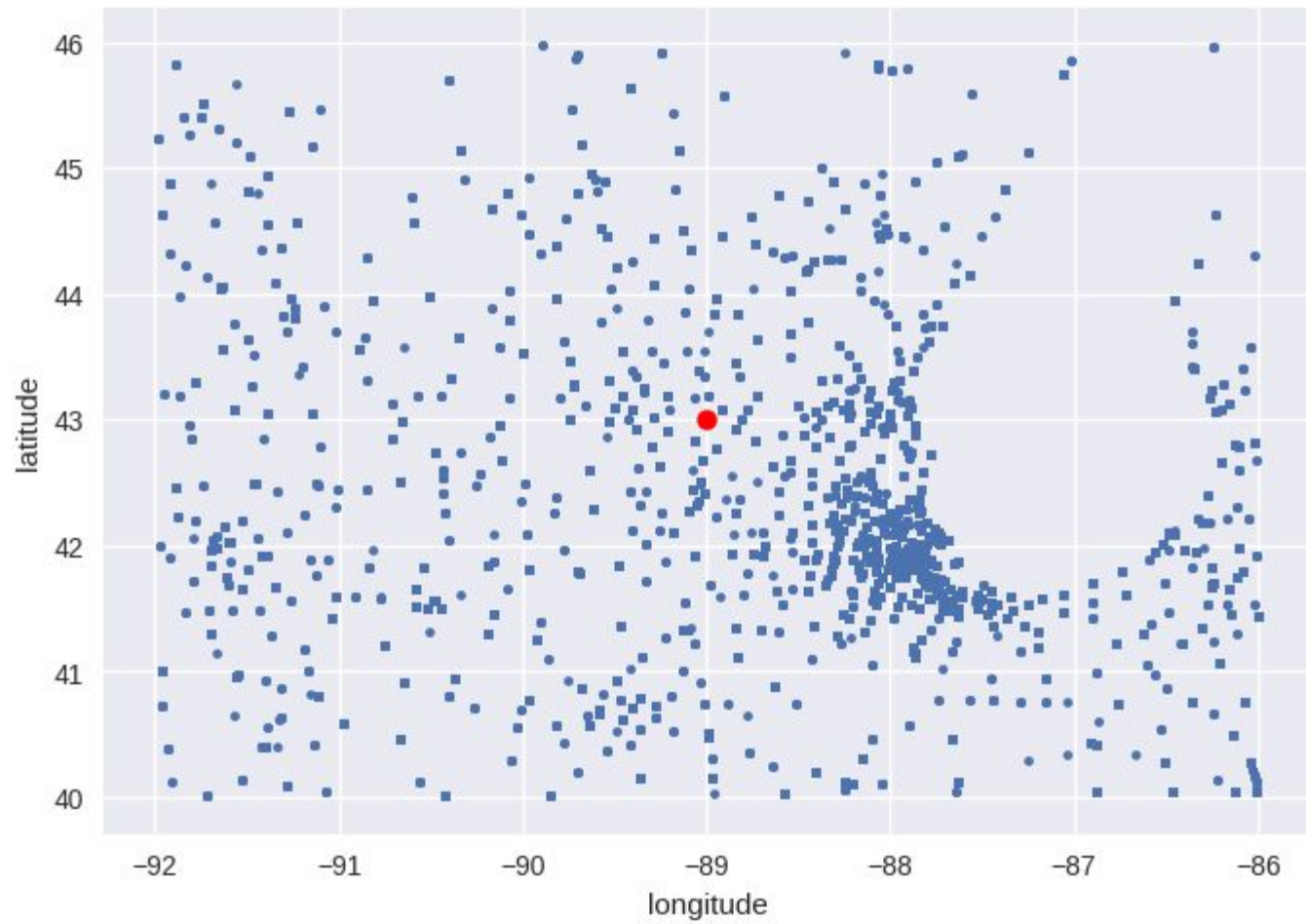
preloaded.com

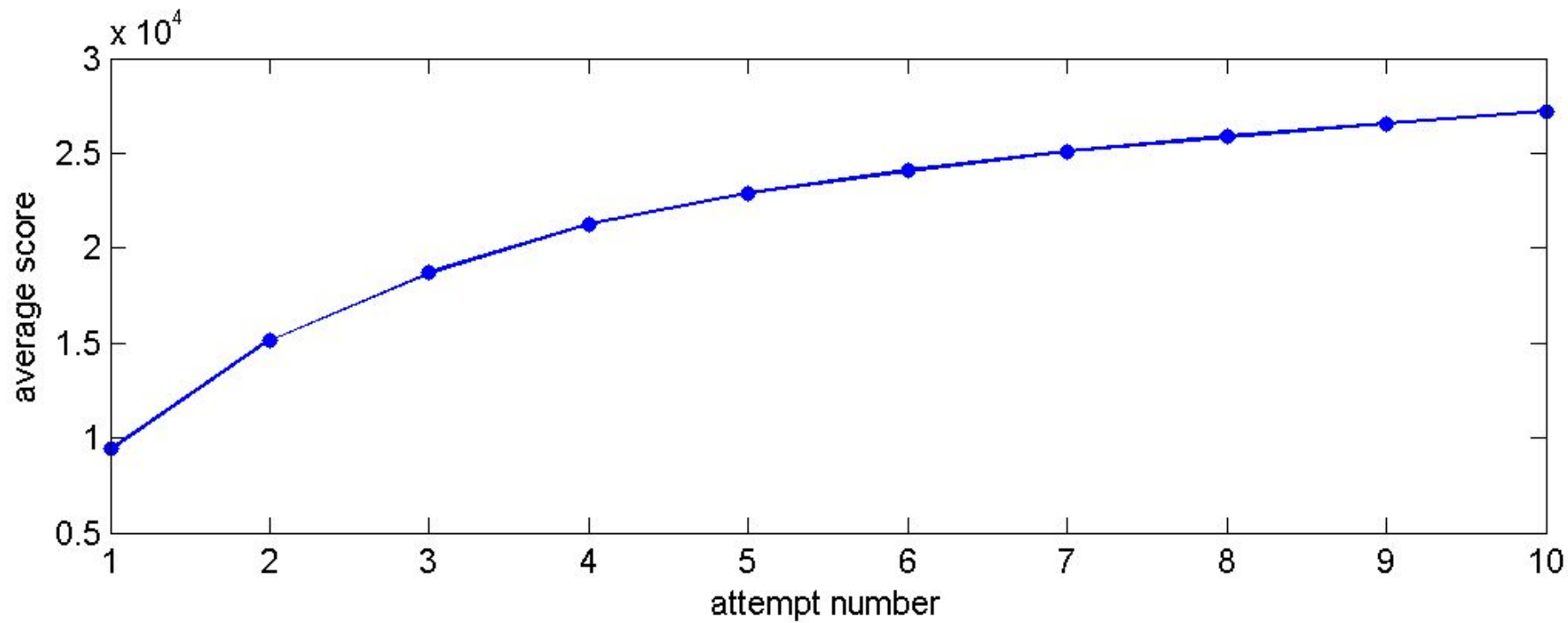
wellcome^{trust}

<http://axon.wellcomeapps.com/>

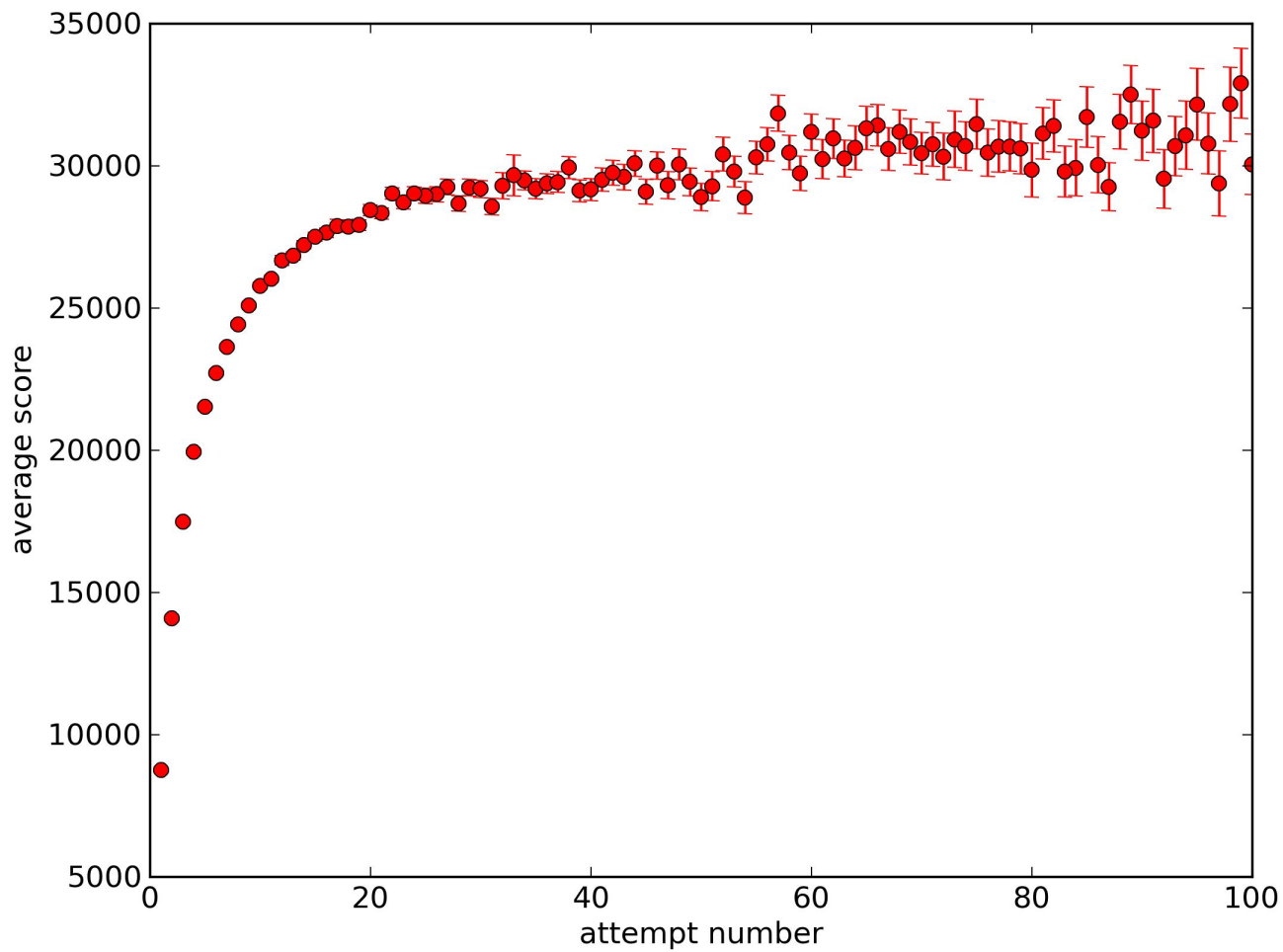
plays = 4038802







$n=854064$



(First, the
good news)

1. Test theories

(in a natural context)

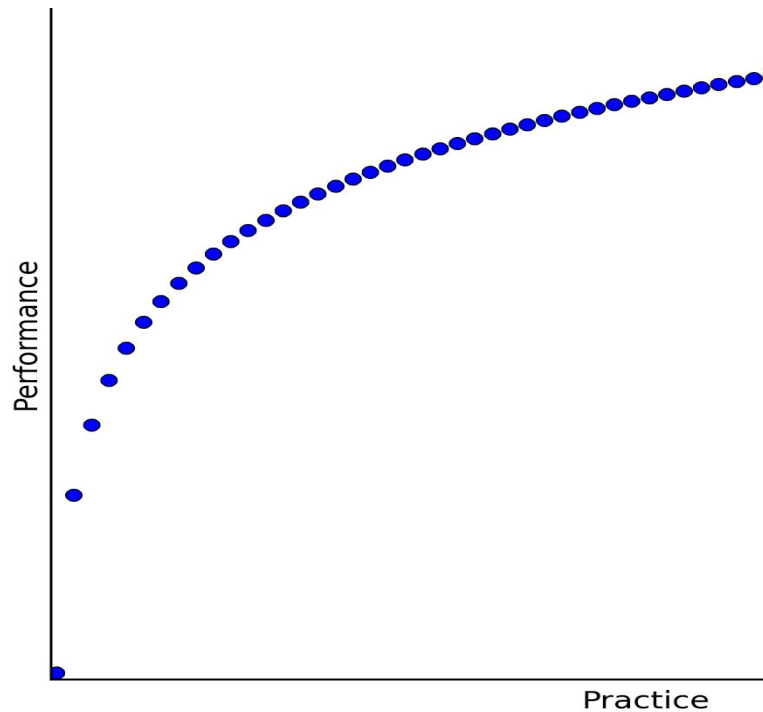
The “Ten Thousand Hours Rule”

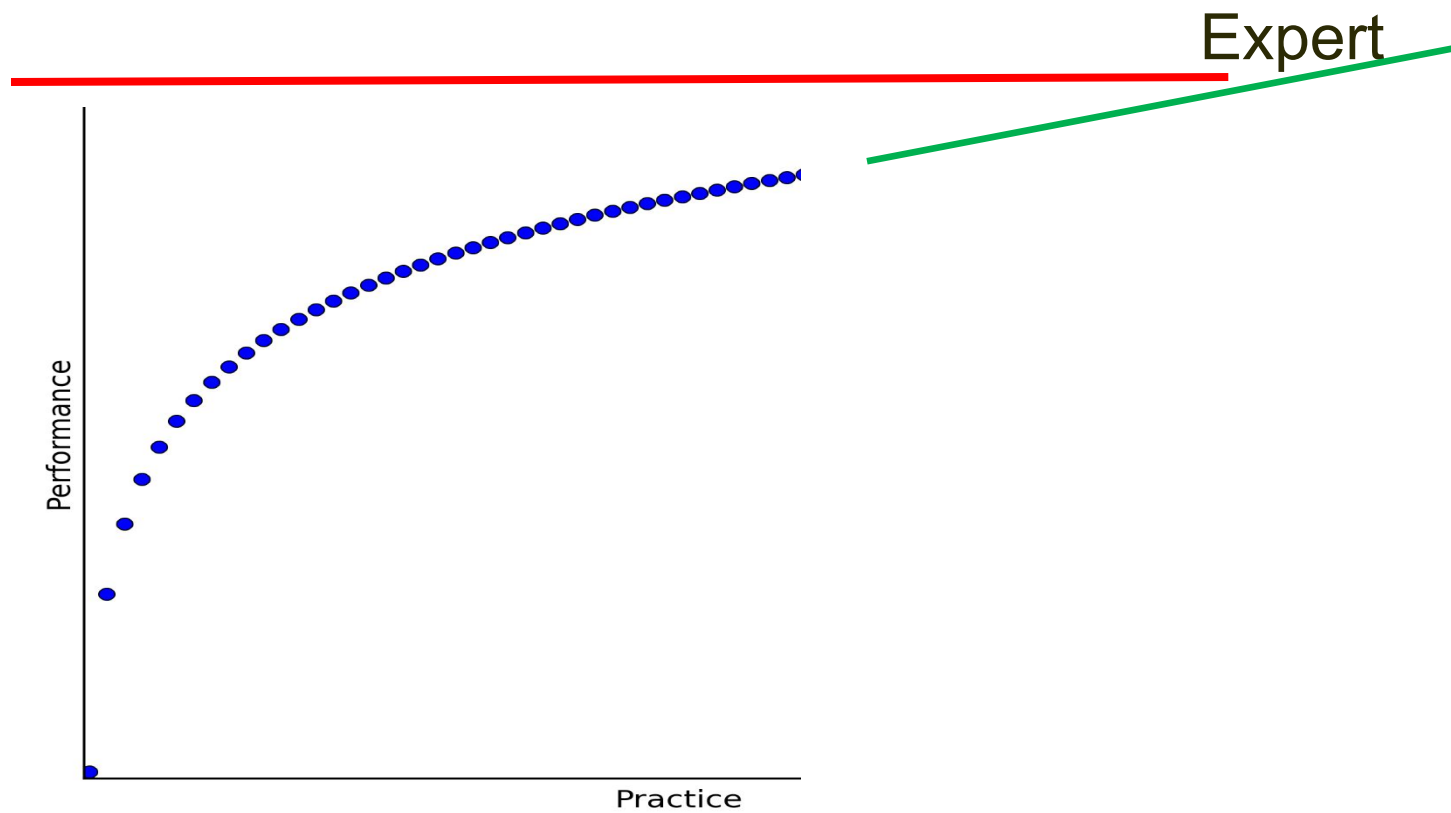


Ericsson, K. A., Krampe, R. T., & Tesch-Rmer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406.

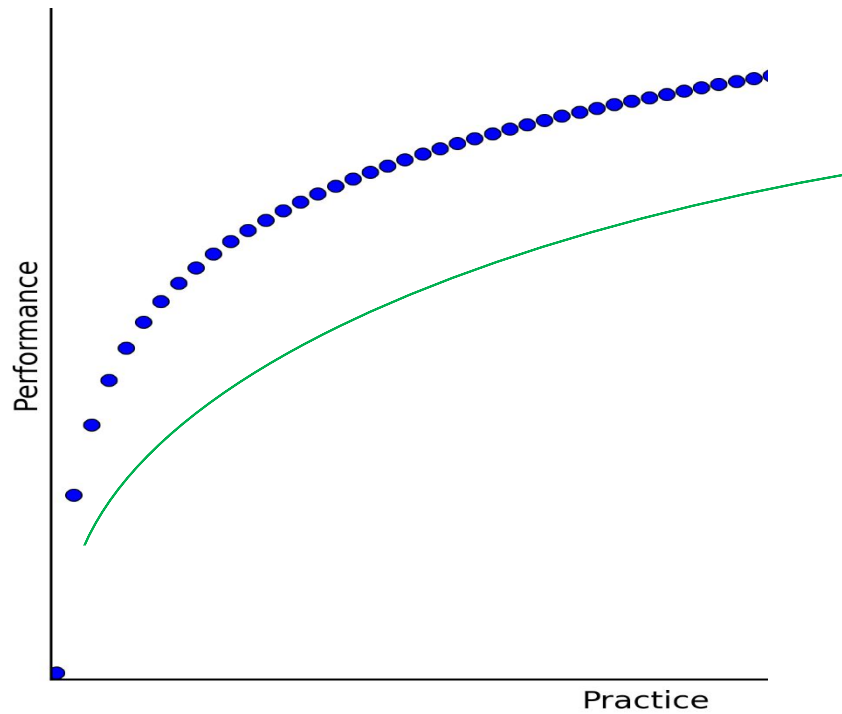
Gladwell, M. (2008). *Outliers: The story of success*. Little, Brown.

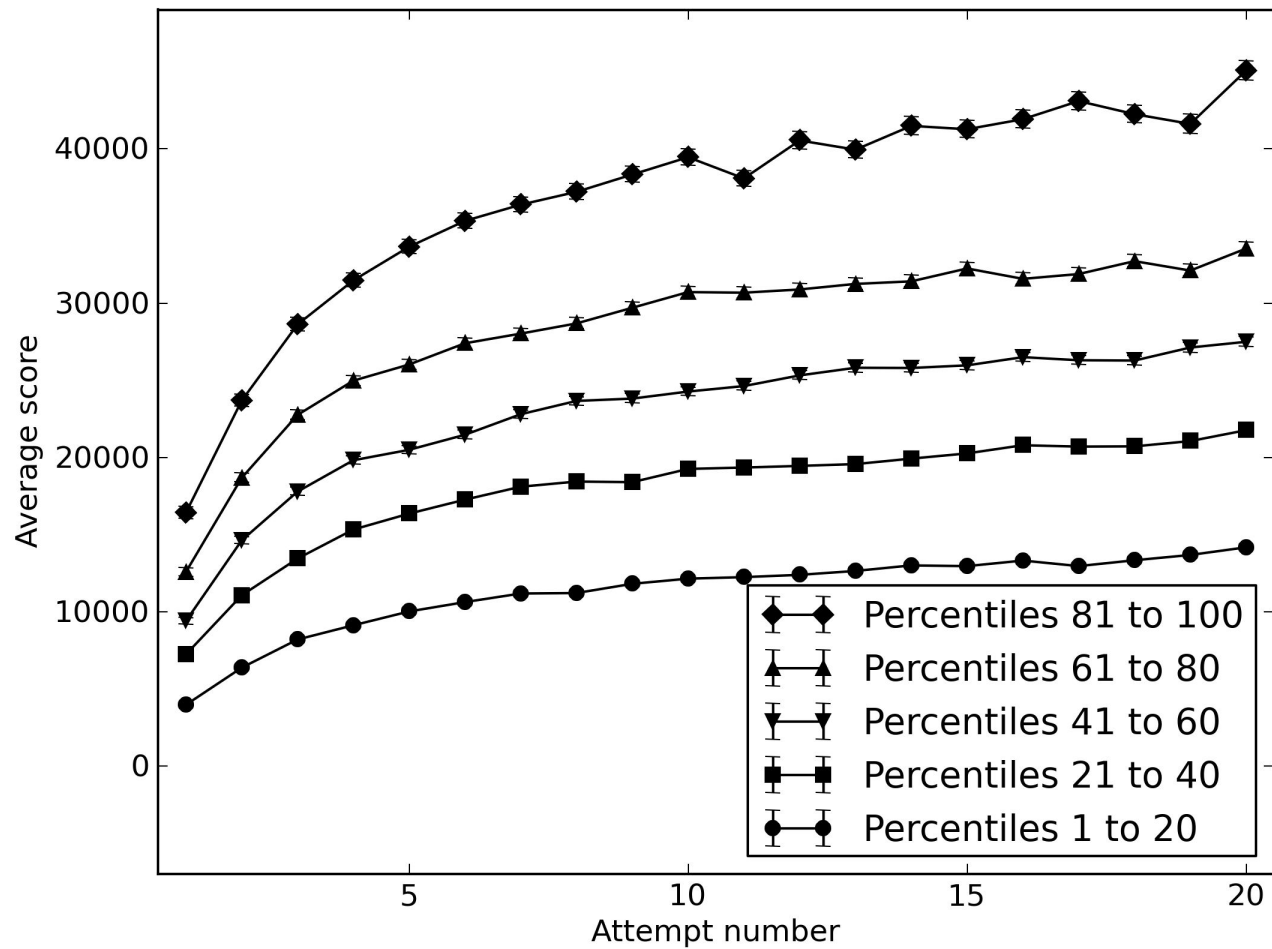
Expert





Expert





2. Beyond effect size

(‘parametric comparison’)

How to learn: Spaced vs blocked practice



The hard way

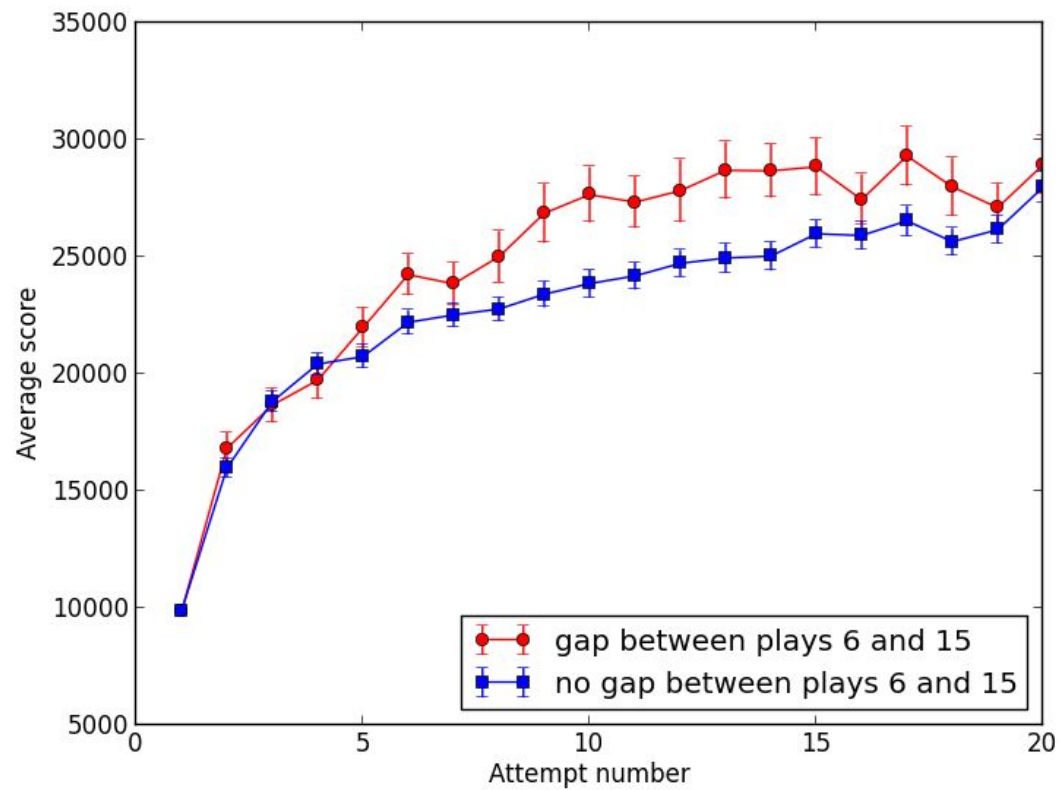


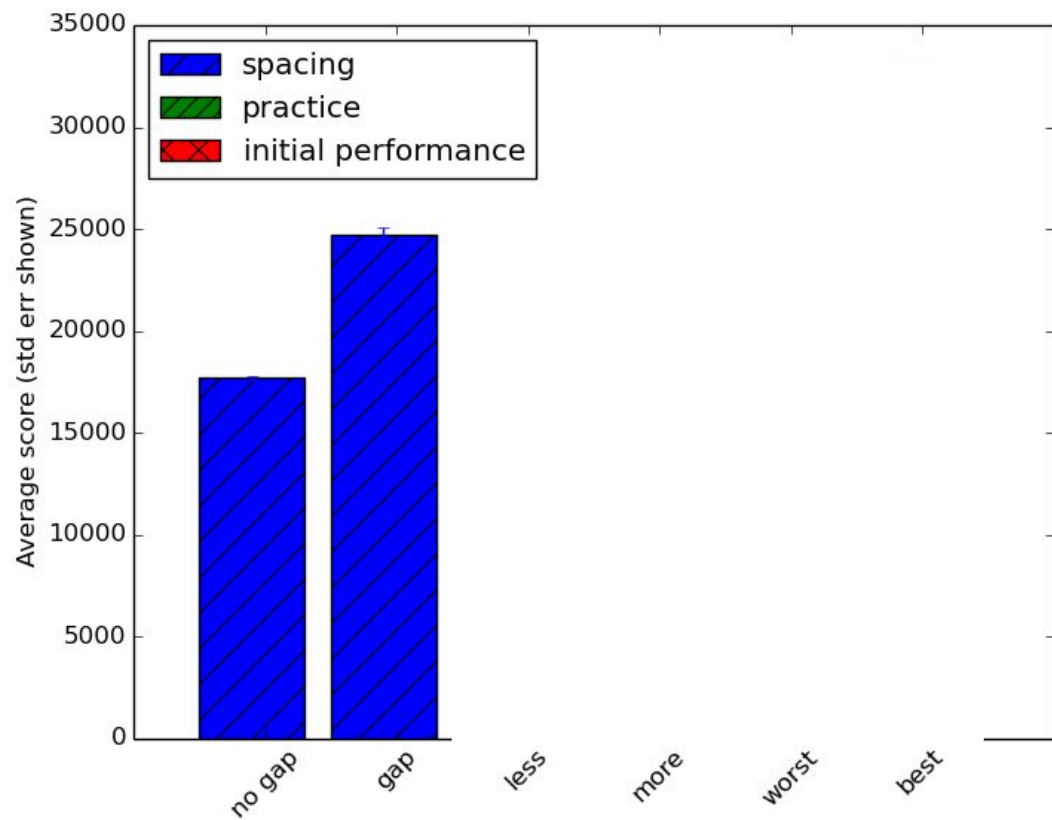
Better

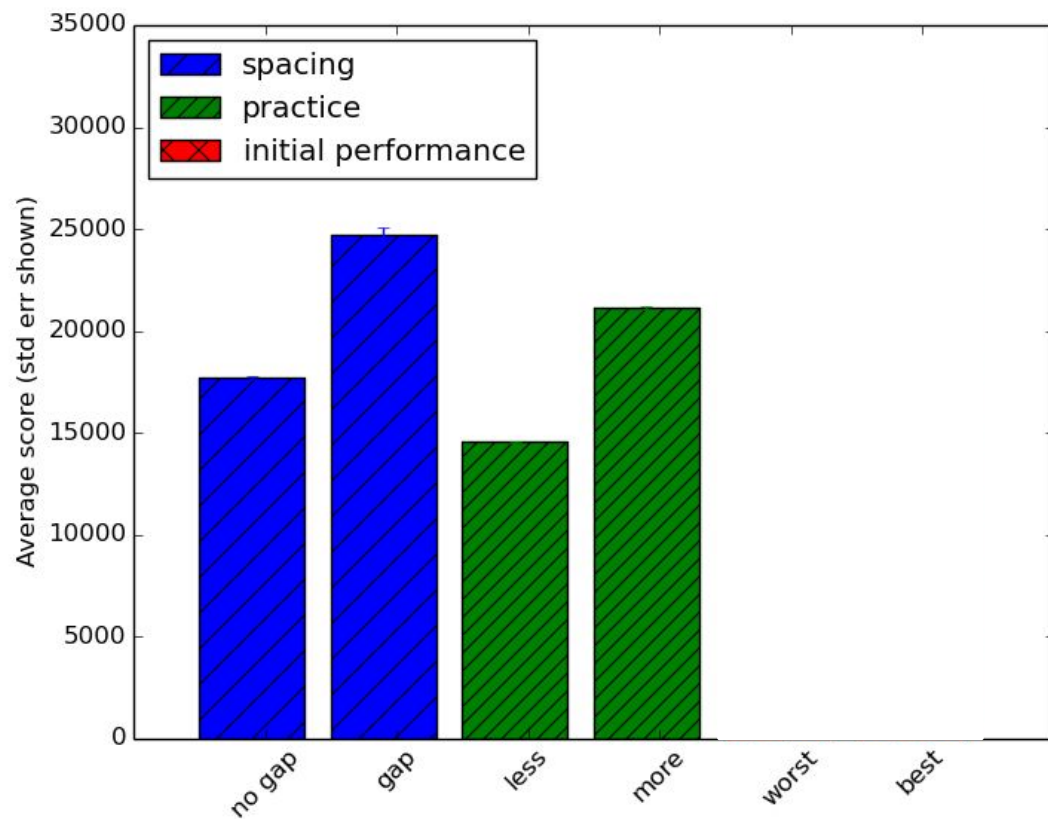


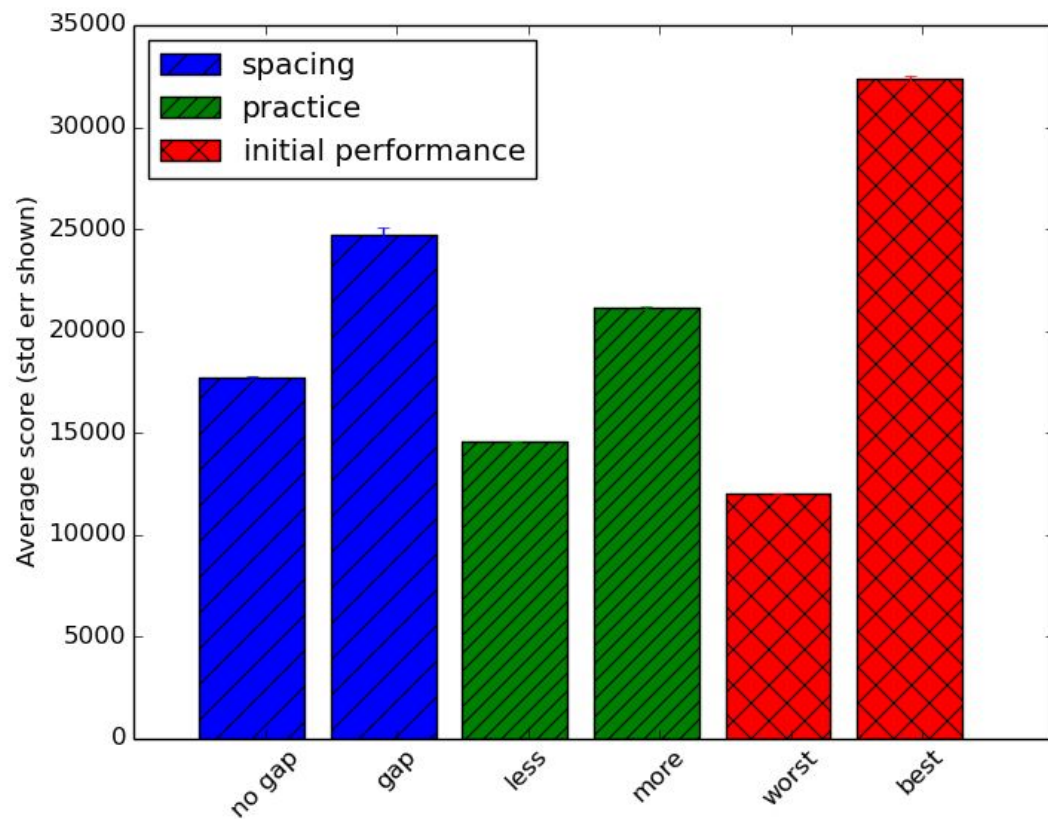
Easier

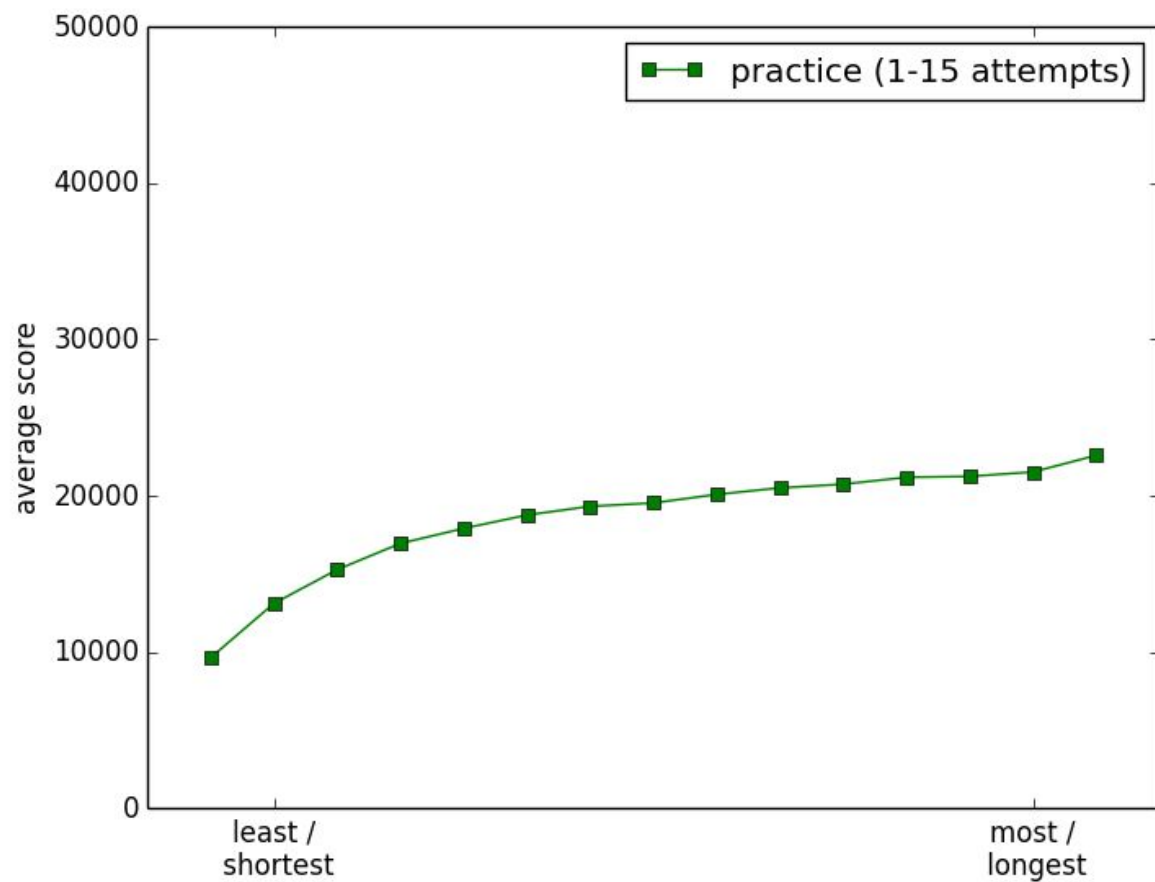
Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, 1, 309-330. doi:10.1007/BF01320097



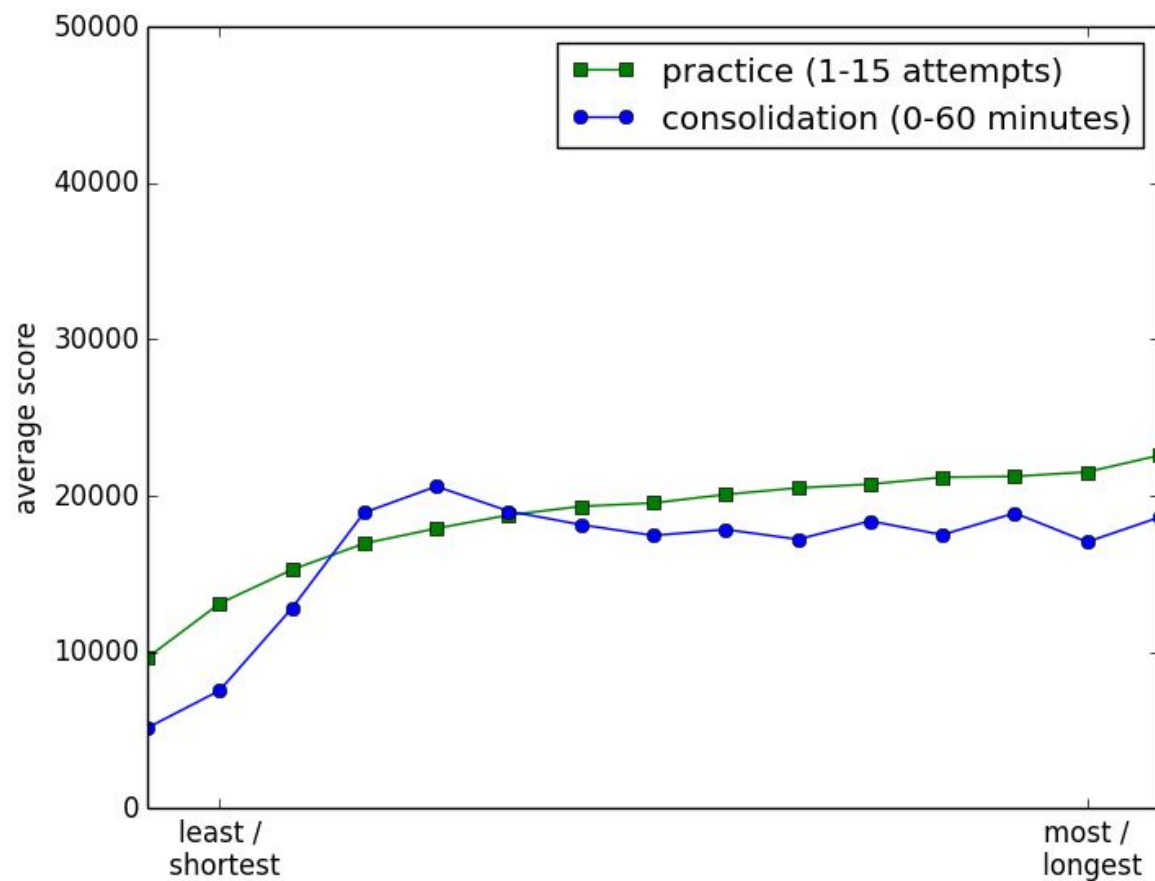




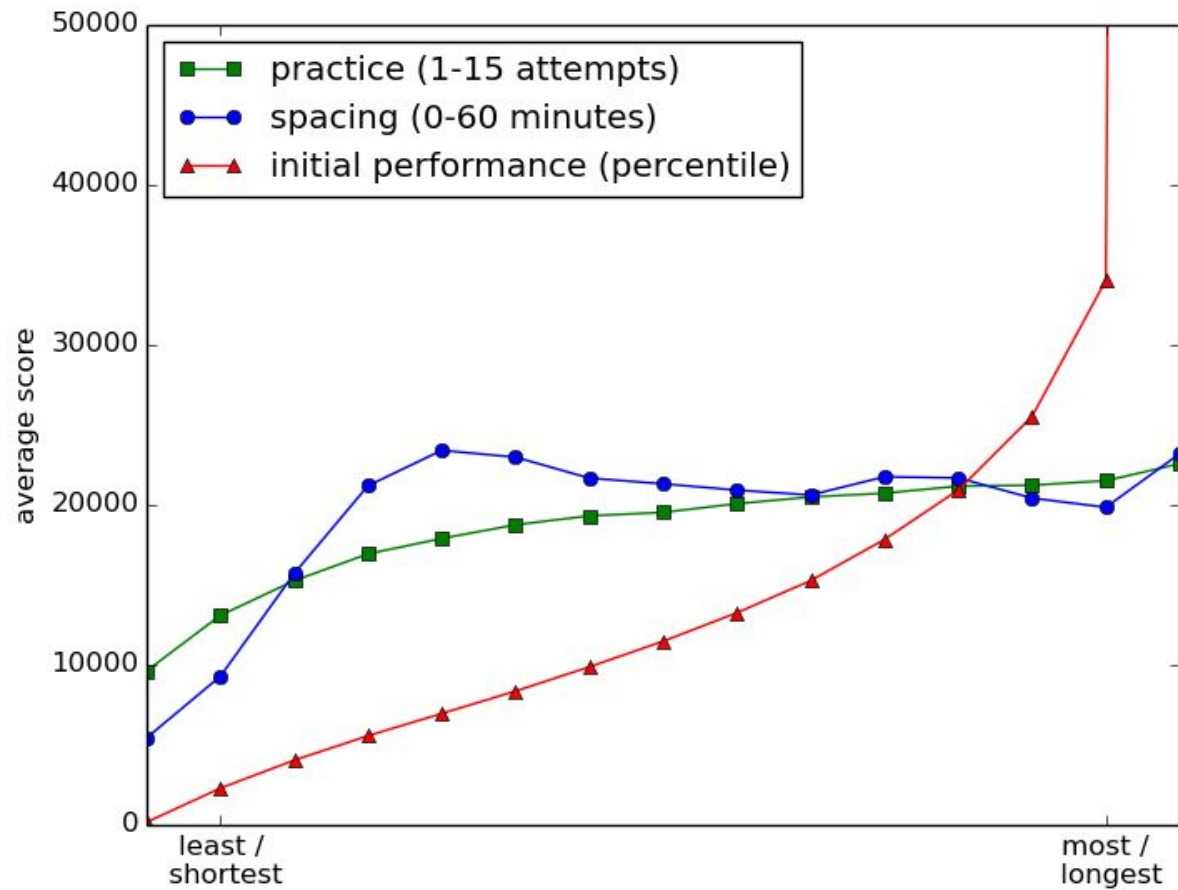




Stafford & Haasnoot (2017)



Stafford & Haasnoot (2017)



Stafford & Haasnoot (2017)

3. Transparency

Open, reproducible workflows

Browser address bar: <https://osf.io/fckq8/>

OSFHOME ▾ Search Support Sign Up Sign In

Skill learning in a... Files Wiki Analytics Registrations Forks

Skill learning in an online game

Public 0

Contributors: [Tom Stafford](#)

Date created: 2015-06-17 12:18 AM | Last Updated: 2017-07-09 06:10 AM

Category: Project

Description: Data from the game Axon and analysis code. See

Stafford, T. & Dewar, M. (2014). Tracing the trajectory of skill learning with a very large sample of online game players. *Psychological Science*, 25(2) 511-518.

and

Stafford, T. & Haasnoot, E. (in press). Testing sleep consolidation in skill learning: a field study using an online game. *Topics In Cognitive Science*. 9(2), 485-496

<https://osf.io/fckq8/>

(Now, things I
worry about)

4. Sample bias

White People

or

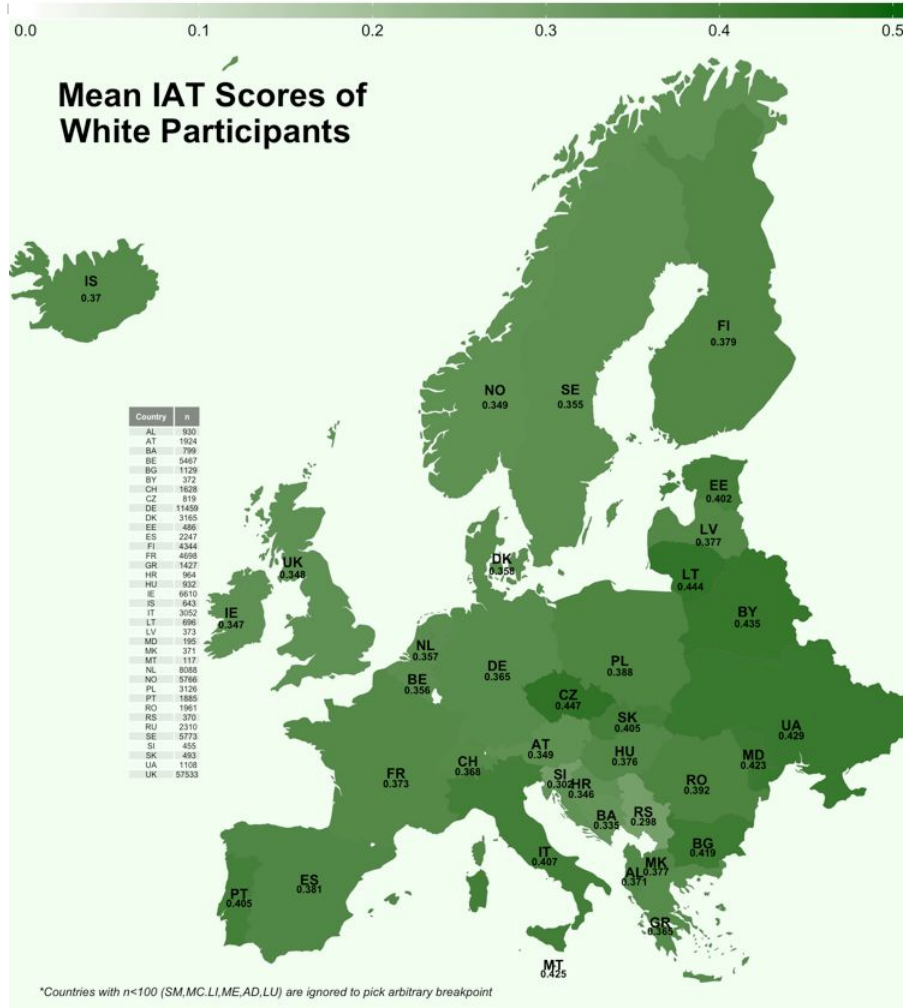
Good

Black People

or

Bad





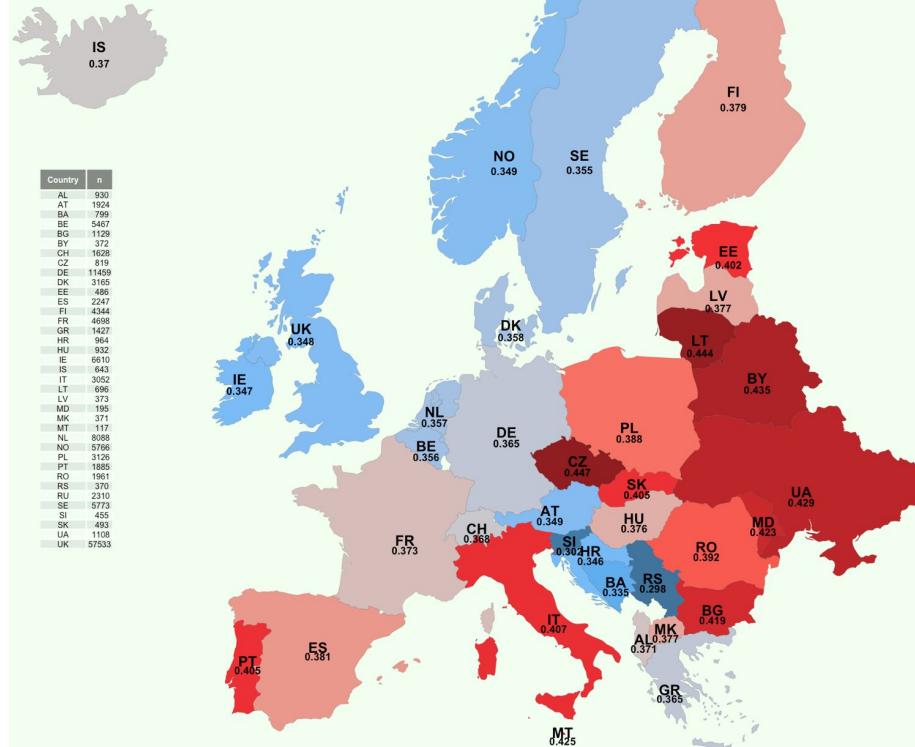
[https://figshare.com/articles/Europe
an_map_of_Implicit_Racial_Bias/47
50588](https://figshare.com/articles/Europe_an_map_of_Implicit_Racial_Bias/4750588)

0.29

0.37

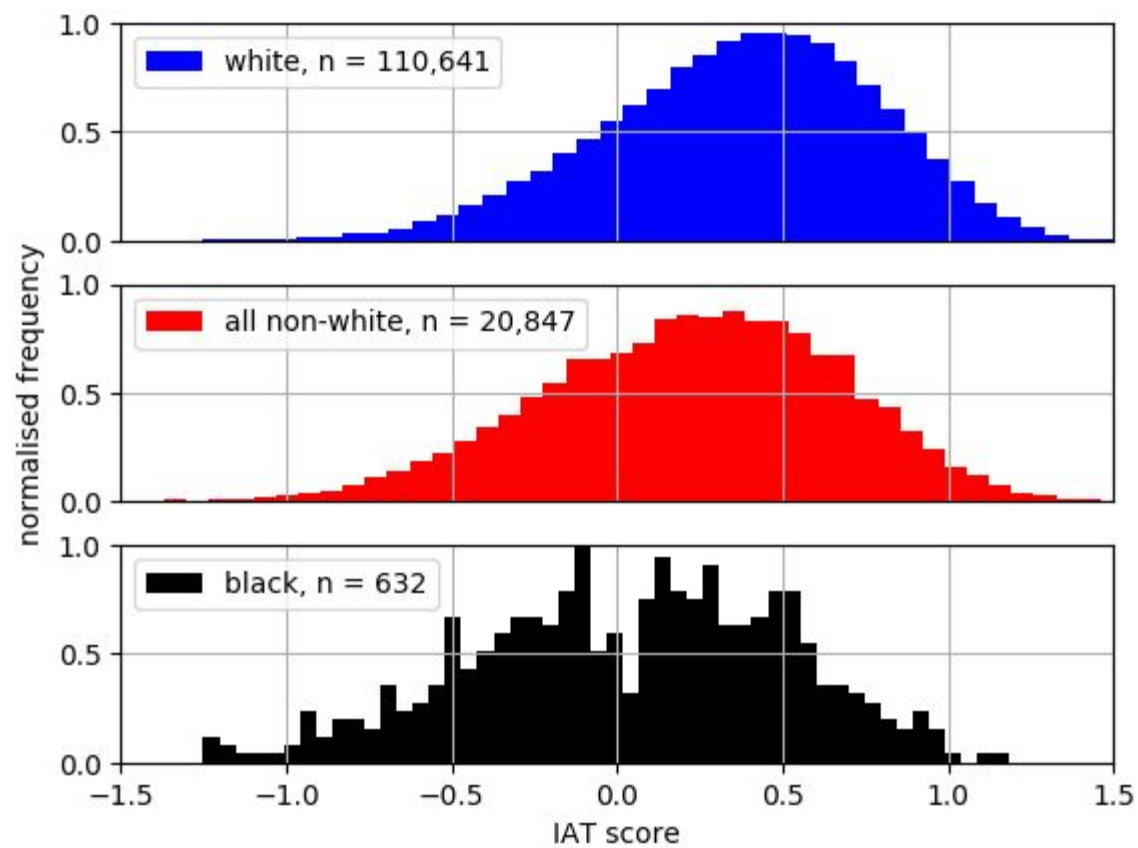
0.45

Mean IAT Scores of White Participants



[https://figshare.com/articles/Europe
an_map_of_Implicit_Racial_Bias/47
50588](https://figshare.com/articles/Europe_an_map_of_Implicit_Racial_Bias/4750588)

*Countries with n<100 (SM,MC,LI,ME,AD,LU) are ignored to pick arbitrary breakpoint



4. Sample bias

-> Triangulate

Destiny

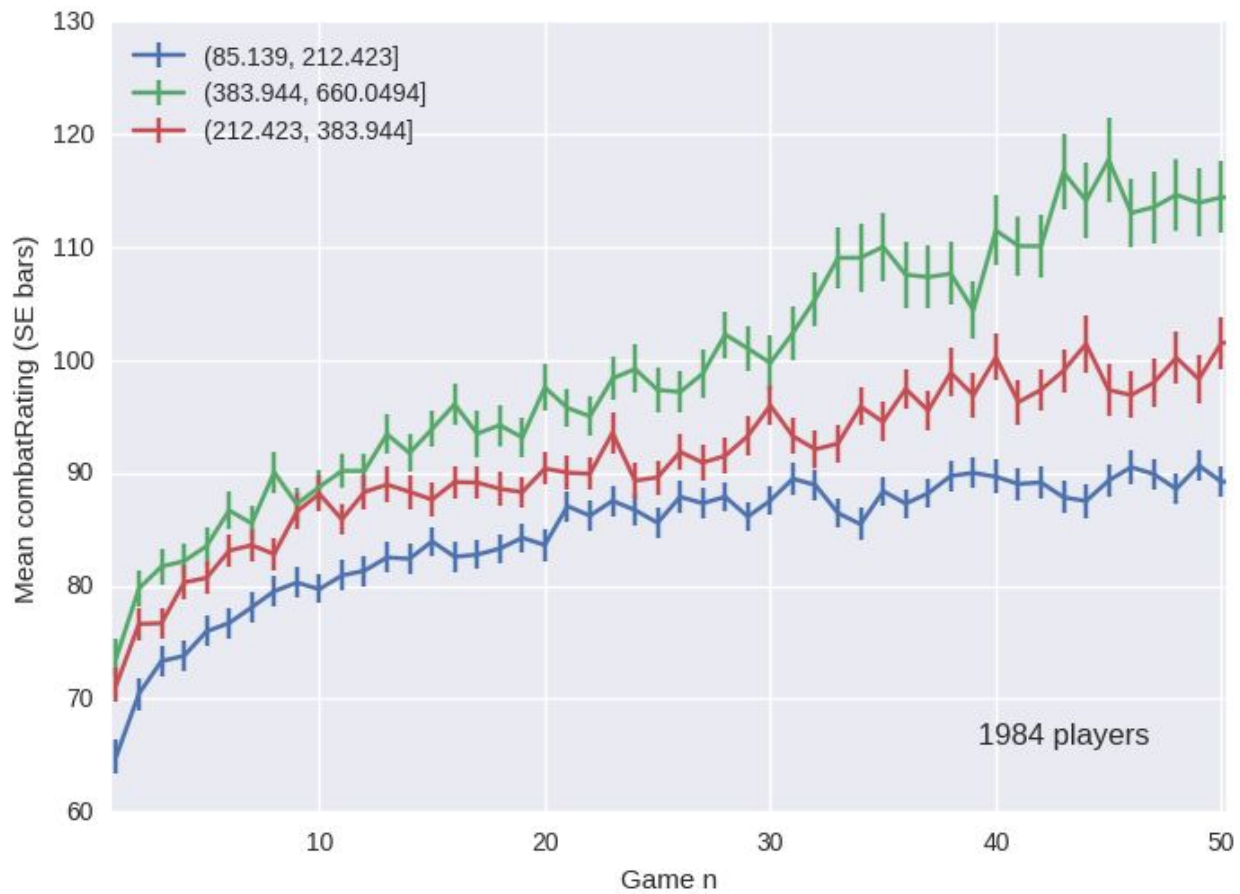


~30 million active players

With Sam Devlin,
Anders Drachen
(University of York),
Rafet Sifa (Fraunhofer
Institute)

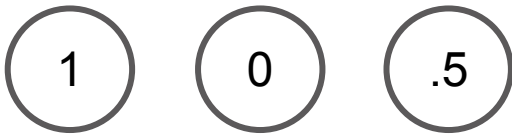
<https://osf.io/dvtsu/>

Exploration and Skill Acquisition in a Major Online Game, Poster Thursday 13:20, CogSci17

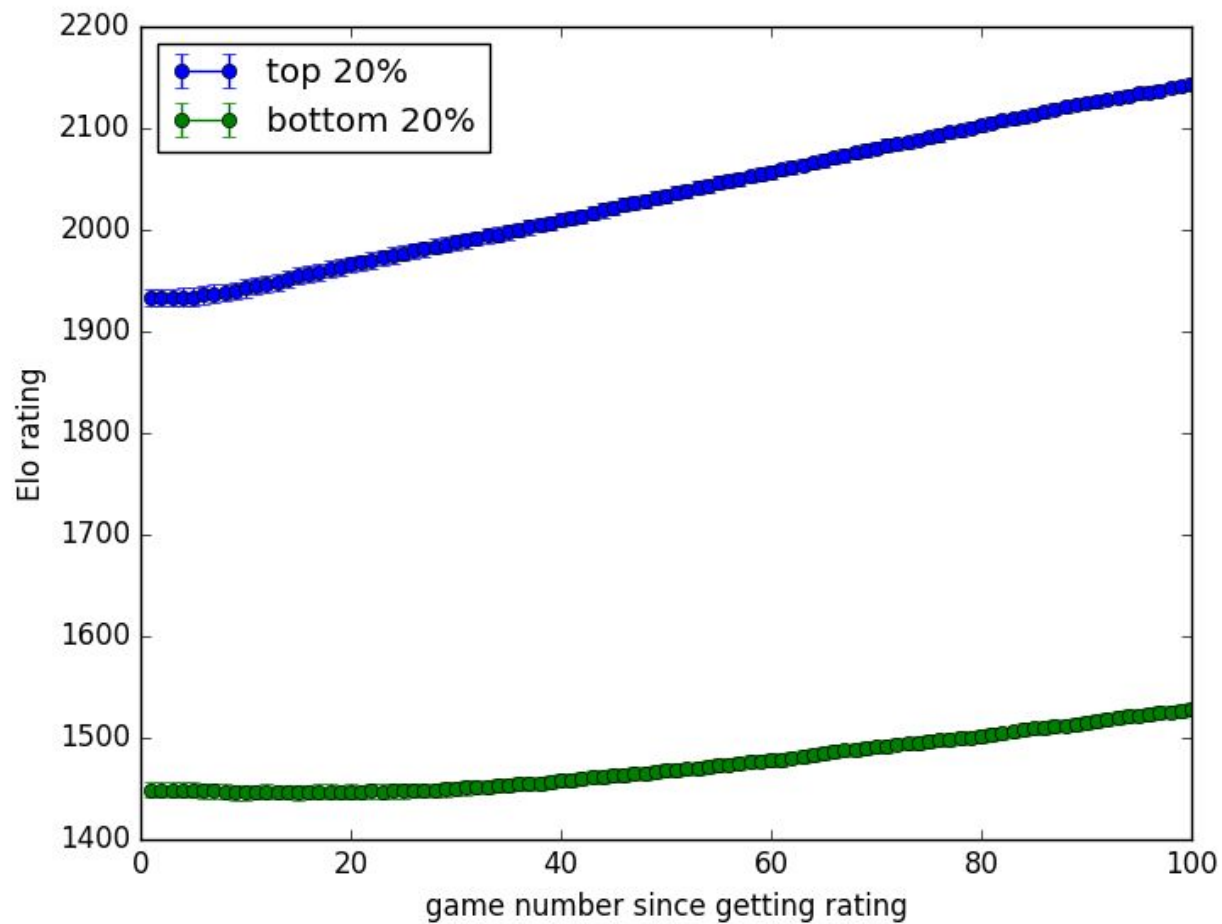


Chess

2008-2015, >11 million FIDE tournament games, via Jeff Sonas



Stafford, T. (in review). *Stereotype lift: female chess players outperform expectations when playing men*
<http://www.tomstafford.staff.shef.ac.uk/?p=464>



Complex - Simple



Axon

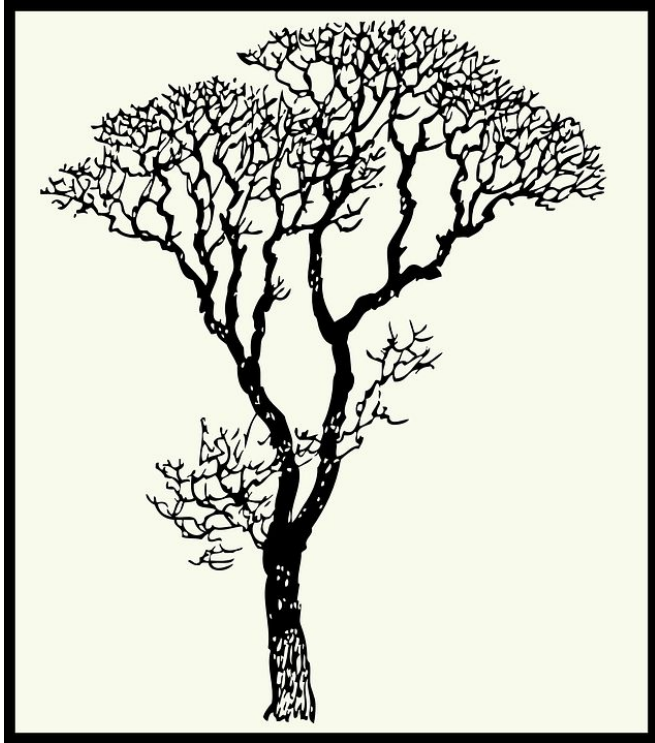
Chess

Destiny



Perceptuo-Motor - Cognitive

5. p-hacking

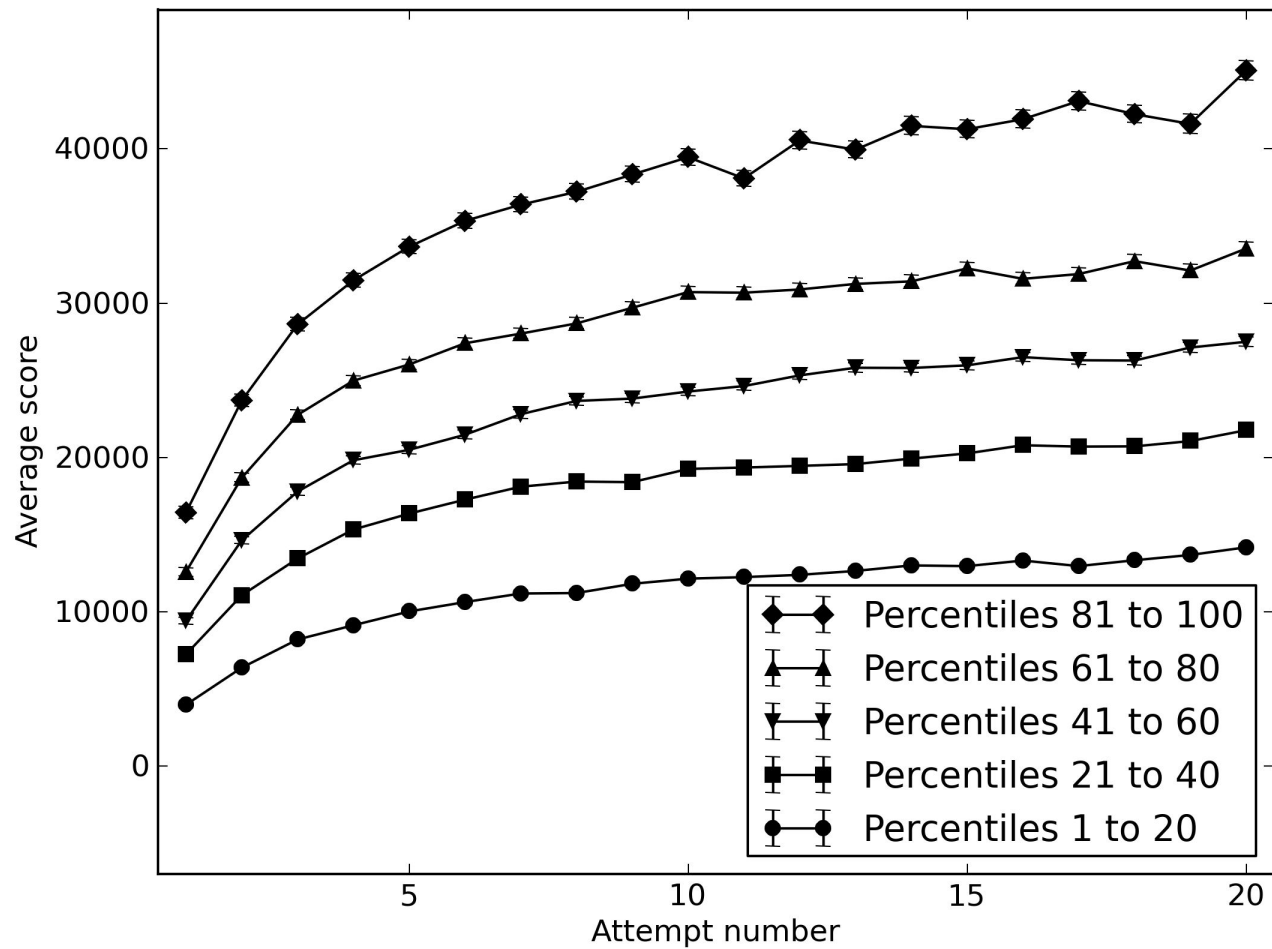


Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). *False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant*. *Psychological science*, 22(11), 1359-1366.

Andrew Gelman & Eric Loken (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time*

5. p-hacking

-> full exploration



5. p-hacking

-> full exploration

-> cross-validation

(Lee: “truly held out data”)

“psychology’s near total focus on explaining the causes of behavior has led much of the field to be populated by research programs that provide intricate theories of psychological mechanism, but that have little (or unknown) ability to predict future behaviors with any appreciable accuracy.”

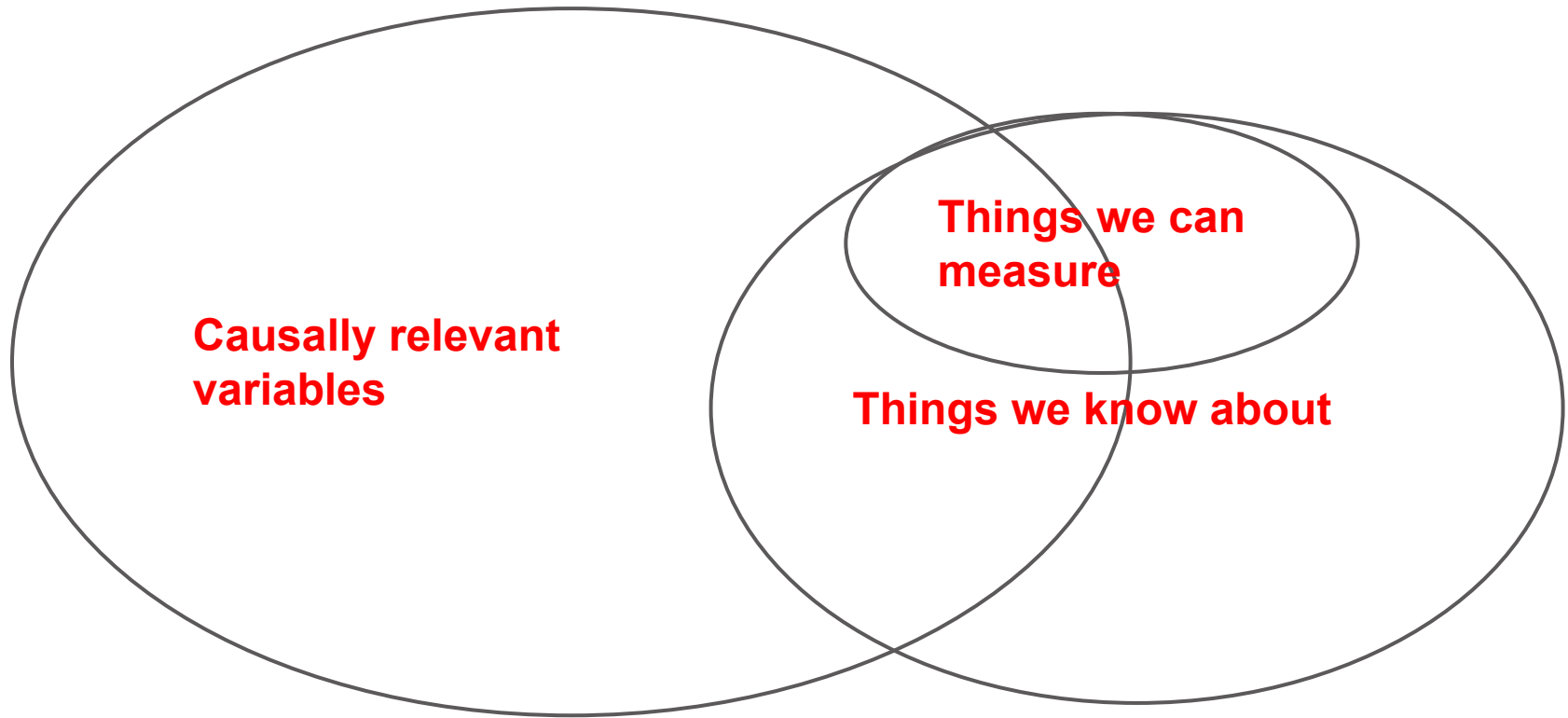
Yarkoni, T. & Westfall, J. (in press). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives in Psychological Science*

6. Confounds

Table 1. We have found 12 papers in which claims coming from observational studies were tested in randomised clinical trials. Many of the trials are quite large. In most of the observational studies multiple claims were tested, often in factorial designs, e.g. vitamin D and calcium individually and together along with a placebo group. Note that none of the claims replicated in the direction claimed in the observational studies and that there was statistical significance in the opposite direction five times

<i>ID no.</i>	<i>Pos.</i>	<i>Neg.</i>	<i>No. of claims</i>	<i>Treatment(s)</i>	<i>Reference</i>
1	0	1	3	Vit E, beta-carotene	<i>NEJM</i> 1994; 330 : 1029–1035
2	0	3	4	Hormone Replacement Ther.	<i>JAMA</i> 2003; 289 : 2651–2662, 2663–2672, 2673–2684
3	0	1	2	Vit E, beta-carotene	<i>JNCI</i> 2005; 97 : 481–488
4	0	0	3	Vit E	<i>JAMA</i> 2005; 293 : 1338–1347
5	0	0	3	Low Fat	<i>JAMA</i> . 2006; 295 : 655–666
6	0	0	3	Vit D, Calcium	<i>NEJM</i> 2006; 354 : 669–683
7	0	0	2	Folic acid, Vit B6, B12	<i>NEJM</i> 2006; 354 : 2764–2772
8	0	0	2	Low Fat	<i>JAMA</i> 2007; 298 : 289–298
9	0	0	12	Vit C, Vit E, beta-carotene	<i>Arch Intern Med</i> 2007; 167 : 1610–1618
10	0	0	12	Vit C, Vit E	<i>JAMA</i> 2008; 300 : 2123–2133
11	0	0	3	Vit E, Selenium	<i>JAMA</i> 2009; 301 : 39–51
12	0	0	3	HRT + Vitamins	<i>JAMA</i> 2002; 288 : 2431–2440
Totals	0	5	52		

Stanley Young, S., & Karr, A. (2011). Deming, data and observational studies A process out of control and needing fixing. *Significance*, 8(3), 116-120.

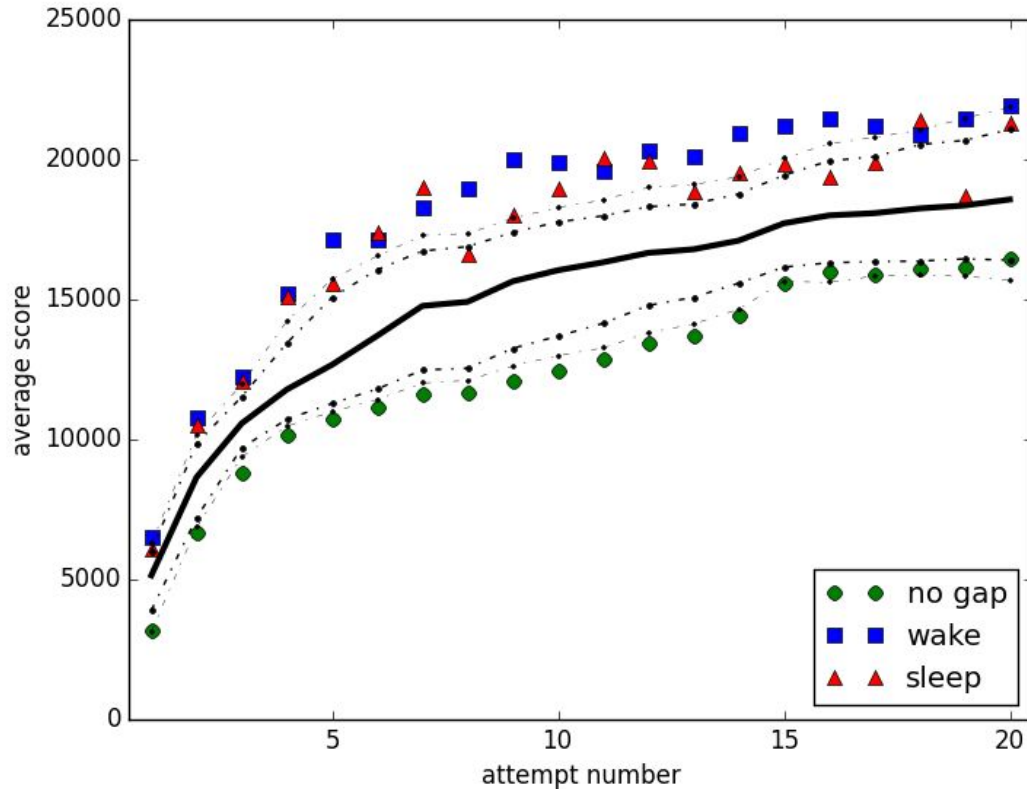


Elwert, F., & Winship, C. (2014). Endogenous selection bias: the problem of conditioning on a collider variable. *Annual Review of Sociology*, 40, 31-53

6. Confounds

-> ~~experiments~~

-> interventions



Stafford, T. & Haasnoot, E. (2017). [Testing sleep consolidation in skill learning: a field study using an online game](#). *Topics in cognitive science*, 9(2), 485-496

Erwin
Haasnoot



Why is sleep consolidation not observable in non-experimental data?

1. The phenomenon isn't robust

Pan, S. C., & Rickard, T. C. (2015). Sleep and motor learning: is there room for consolidation?. *Psychological Bulletin*, 141(4), 812-834.

Rickard, T. C., & Pan, S. C. (2017). Time for considering the possibility that sleep plays no unique role in motor memory consolidation: Reply to Adi-Japha and Karni (2016). *Psychological Bulletin*, 143(4), 454-458.

2. Agents are adaptive

Games for research into Learning

Insight into the 'natural history of learning'

Large data

'Intrinsically motivated'

Unusual individuals? (level of expertise; cf Balota, crowdsourcing to identify usual individuals)

Beyond the lab

1. Test theories in large / unobtainable samples
 2. 'Parametric comparison' / effects in context
 3. Transparent psychological science
 4. Bias? Triangulate
 5. p-hacking/Researcher DoF?
X-validate, full & transparent analysis, ??
 6. Confounds? Combine with expts, predict
-

END

These slides: shoutkey.com/pewter

Email: stafford@sheffield.ac.uk

Twitter: [@tomstafford](https://twitter.com/tomstafford)

<http://www.tomstafford.staff.shef.ac.uk/>

I thought these things might be clues

Theory

A lot of this has been said before

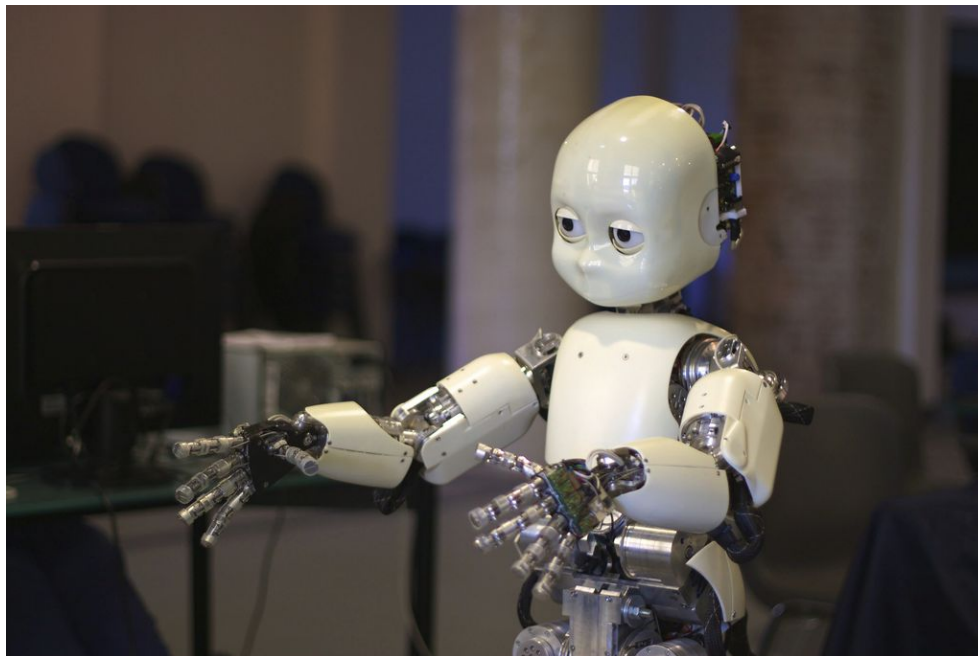
Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring data sets. *Topics in cognitive science*, 8(3), 548-568.

Paxton, A., & Griffiths, T. L. (2017). Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods*, 1-9.

Jones, M. N. (Ed.) (2016). Big data in cognitive science: From methods to insights. New York: Psychology Press.

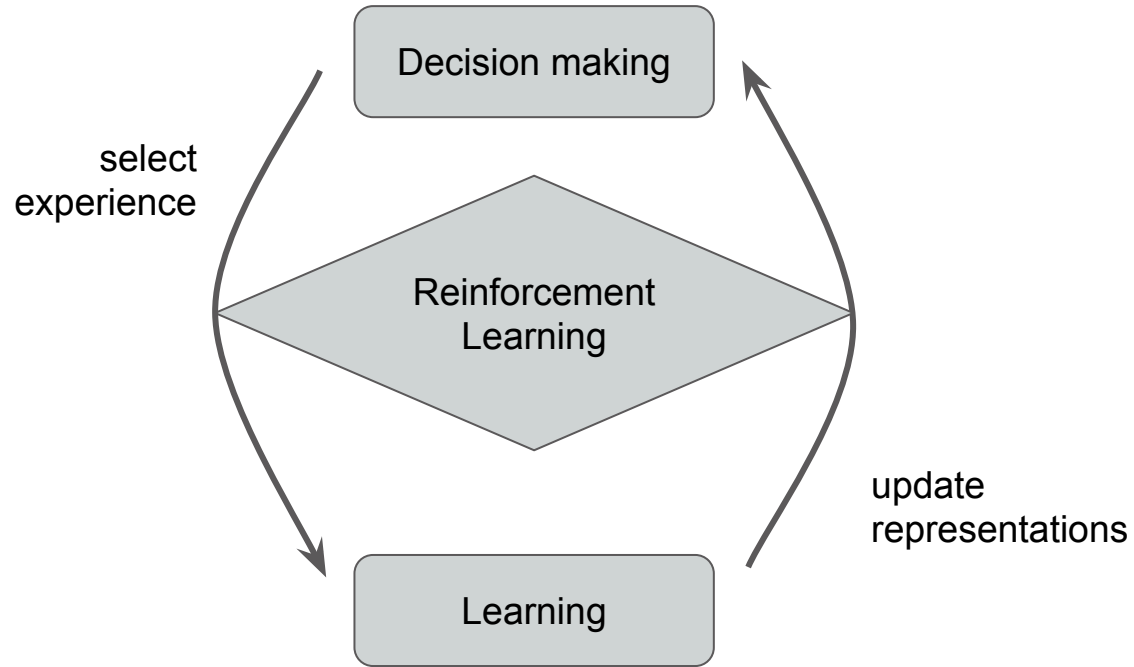


Intrinsically Motivated Cumulative Learning Versatile Robots



[Jiuguang Wang](#), flickr.com, CC-SA-3.0

Baldassarre, G., Stafford, T., Mirolli, M., Redgrave, P., Ryan, R. M., & Barto, A. (2014). [Intrinsic motivations and open-ended development in animals, humans, and robots: an overview](#). *Frontiers in Psychology*, 5(985). doi: 10.3389/fpsyg.2014.00985



Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. Cambridge: MIT press.

Intrinsic motivation and the discovery of novel actions

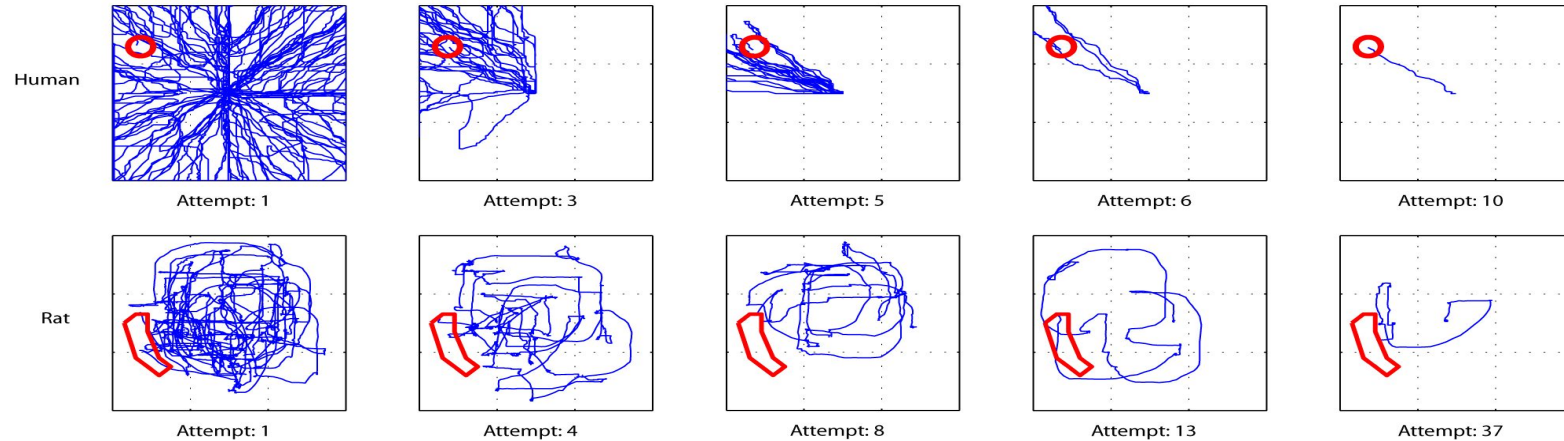
- where do new skills come from?
- study of intelligent movement to reveal principles of intelligence generally

Baldassarre, G., Stafford, T., Mirolli, M., Redgrave, P., Ryan, R. M., & Barto, A. (2014). [Intrinsic motivations and open-ended development in animals, humans, and robots: an overview](#). *Frontiers in Psychology*, 5(985). doi: 10.3389/fpsyg.2014.00985

Gallistel, C. R. (1980). *The organization of action: A new synthesis*. L. Erlbaum Associates.

Rats and humans refine their movements over multiple trials

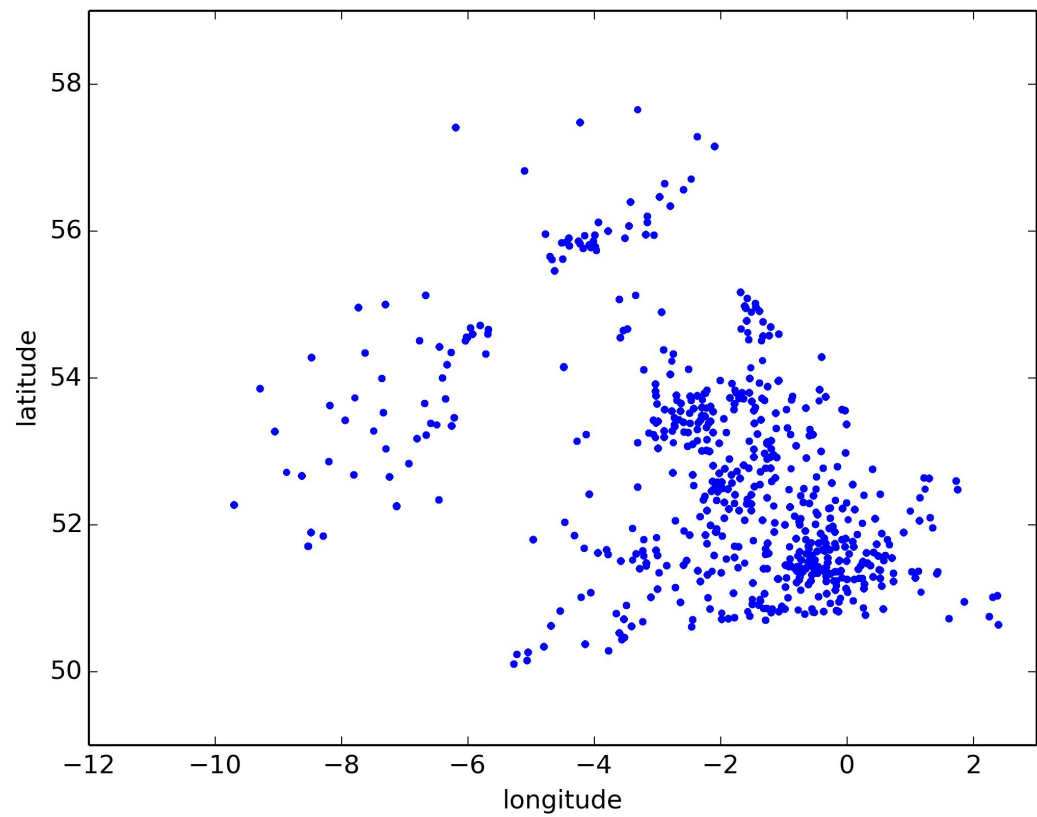
*

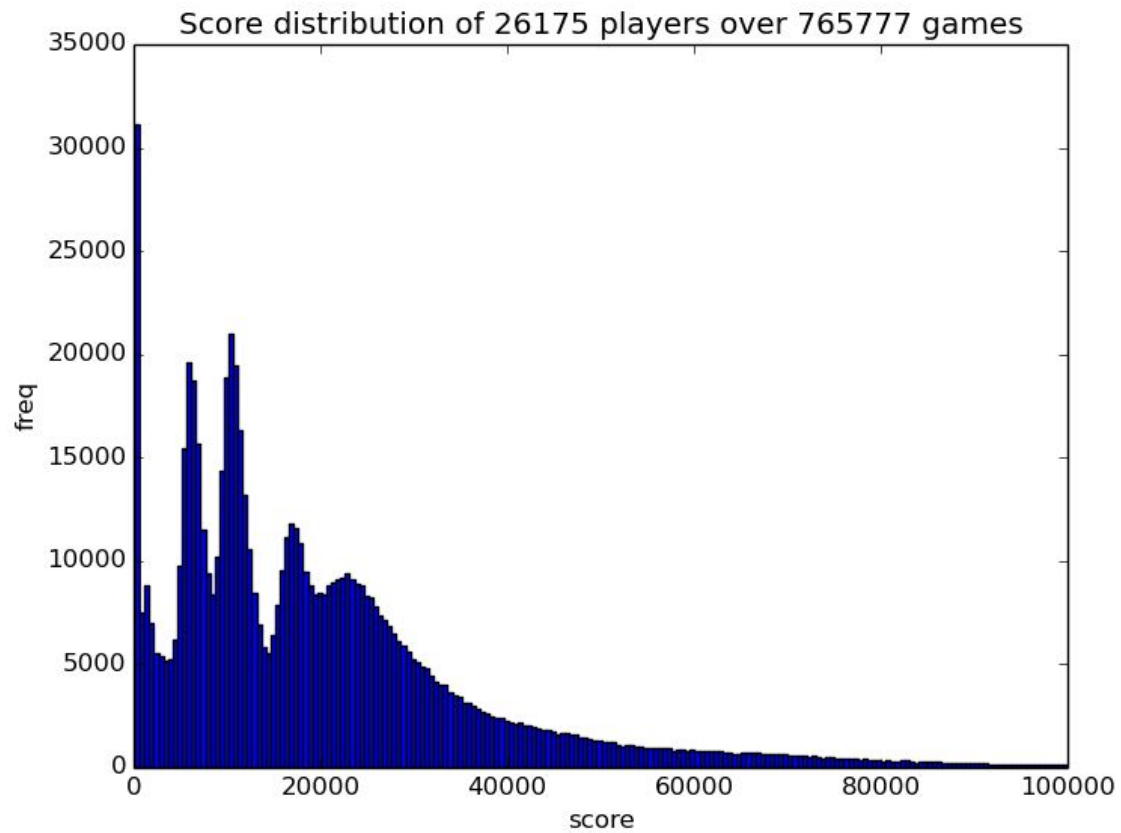


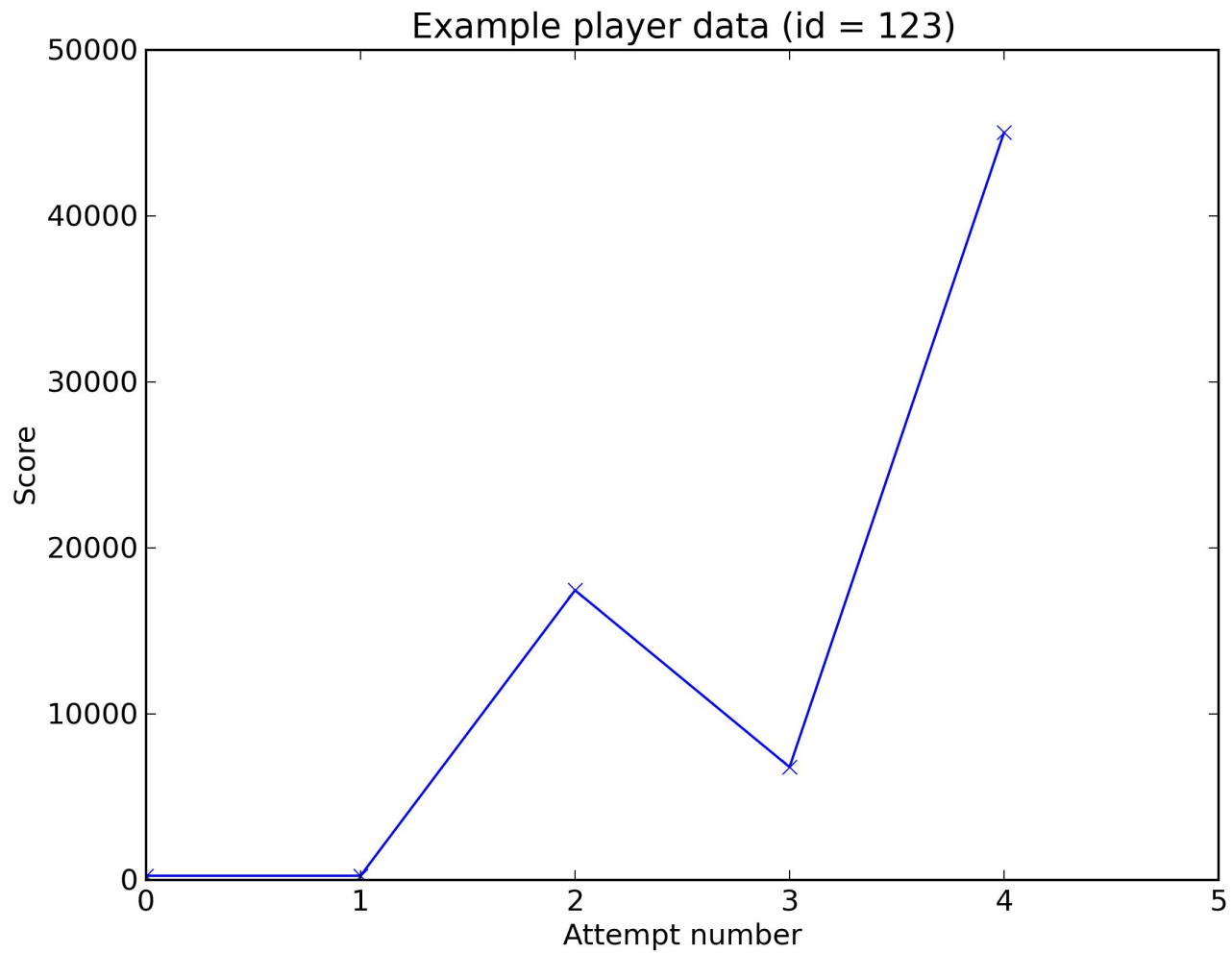
Stafford, T., Thirkettle, M., Walton, T., Vautrelle, N., Hetherington, L., Port, M., Gurney, K.N., Redgrave, P. (2012), [A Novel Task for the Investigation of Action Acquisition](#), *PLoS One*, 7(6), e37749.

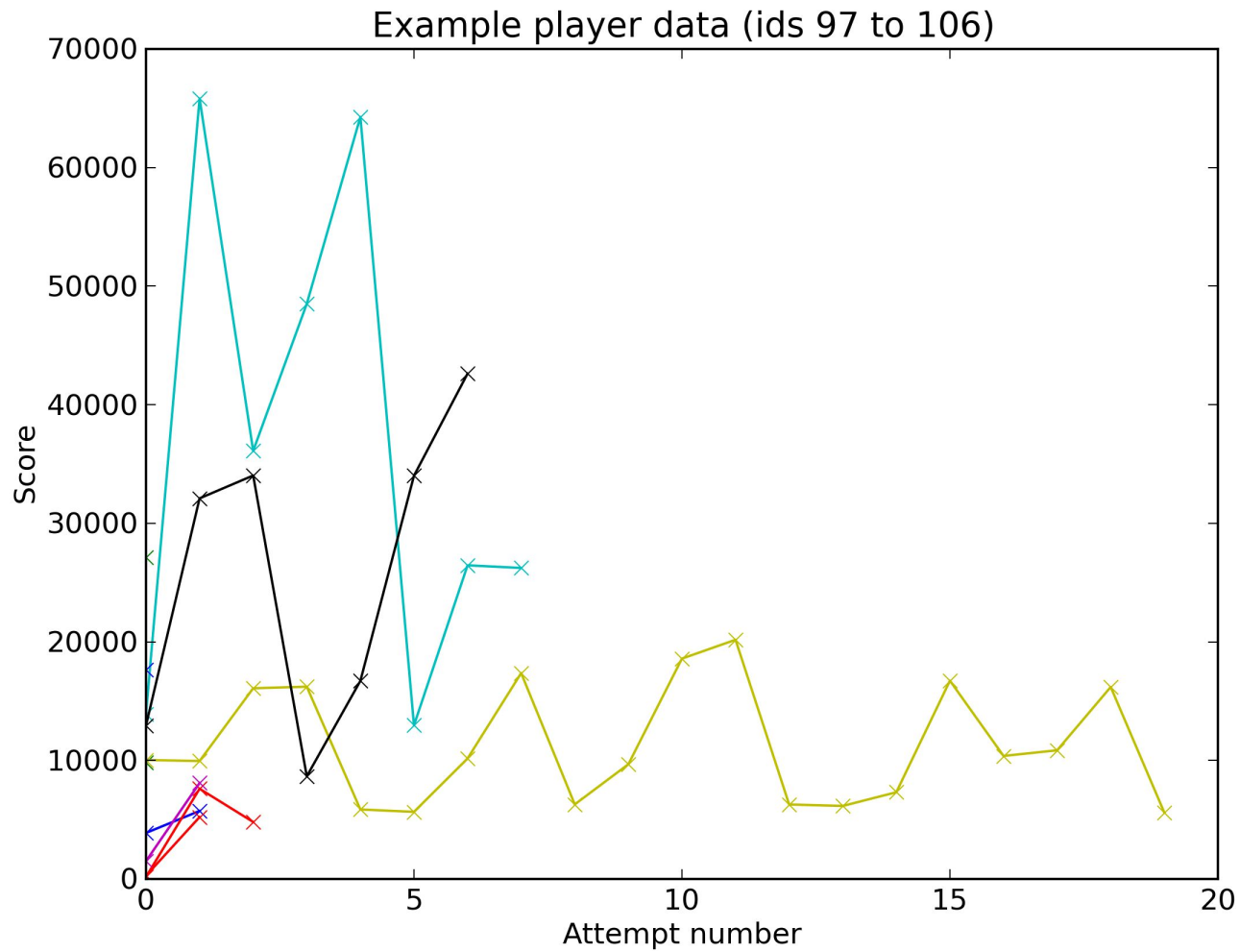
*

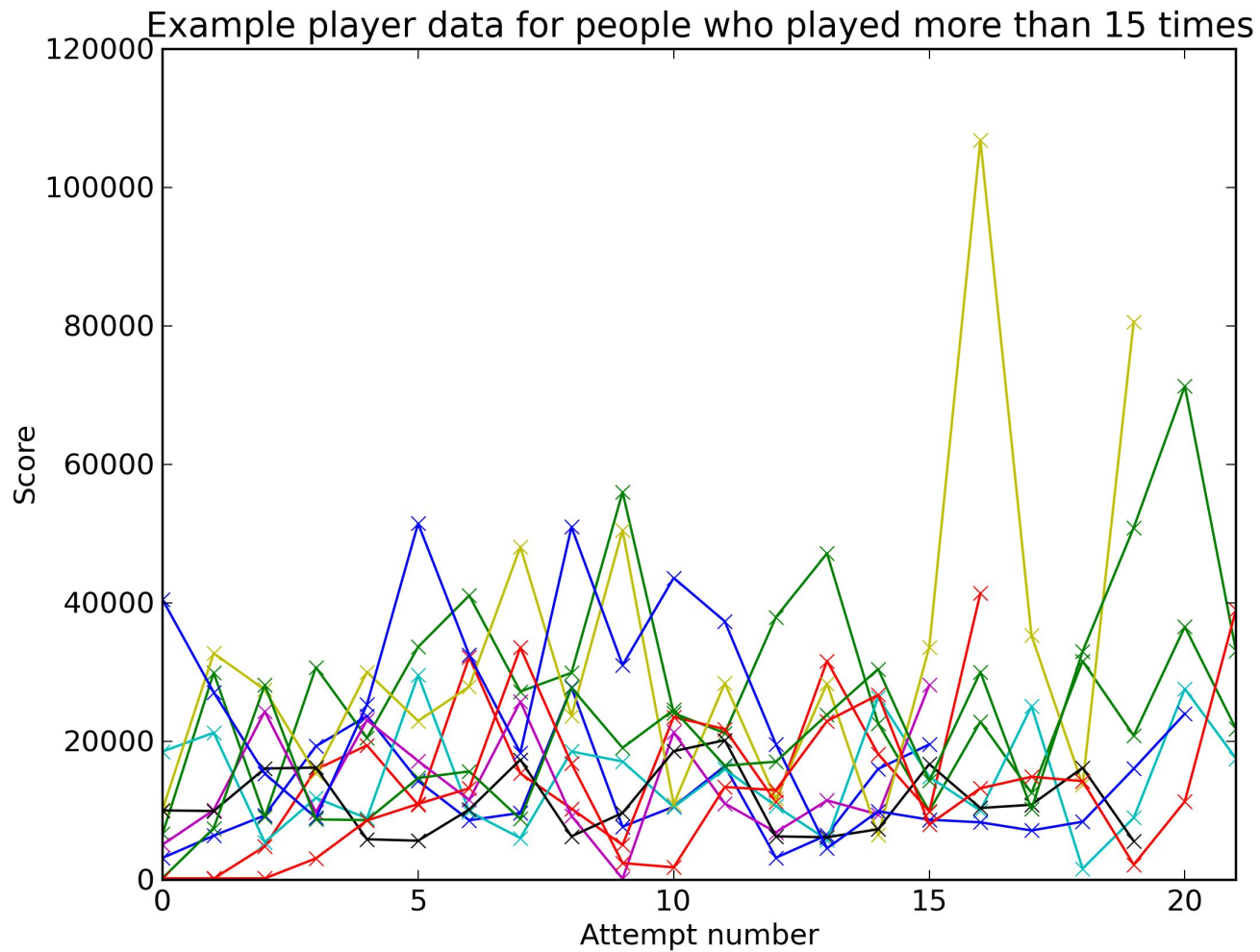
Axon







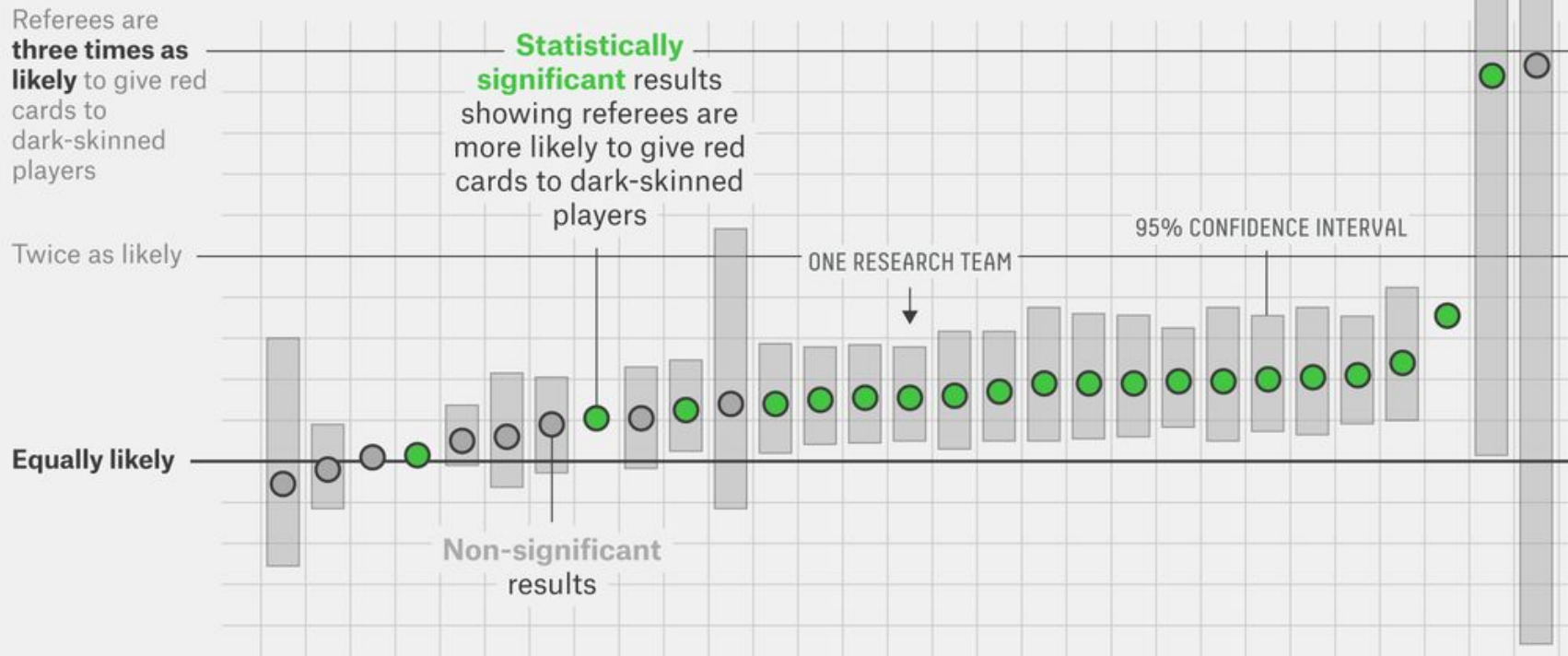




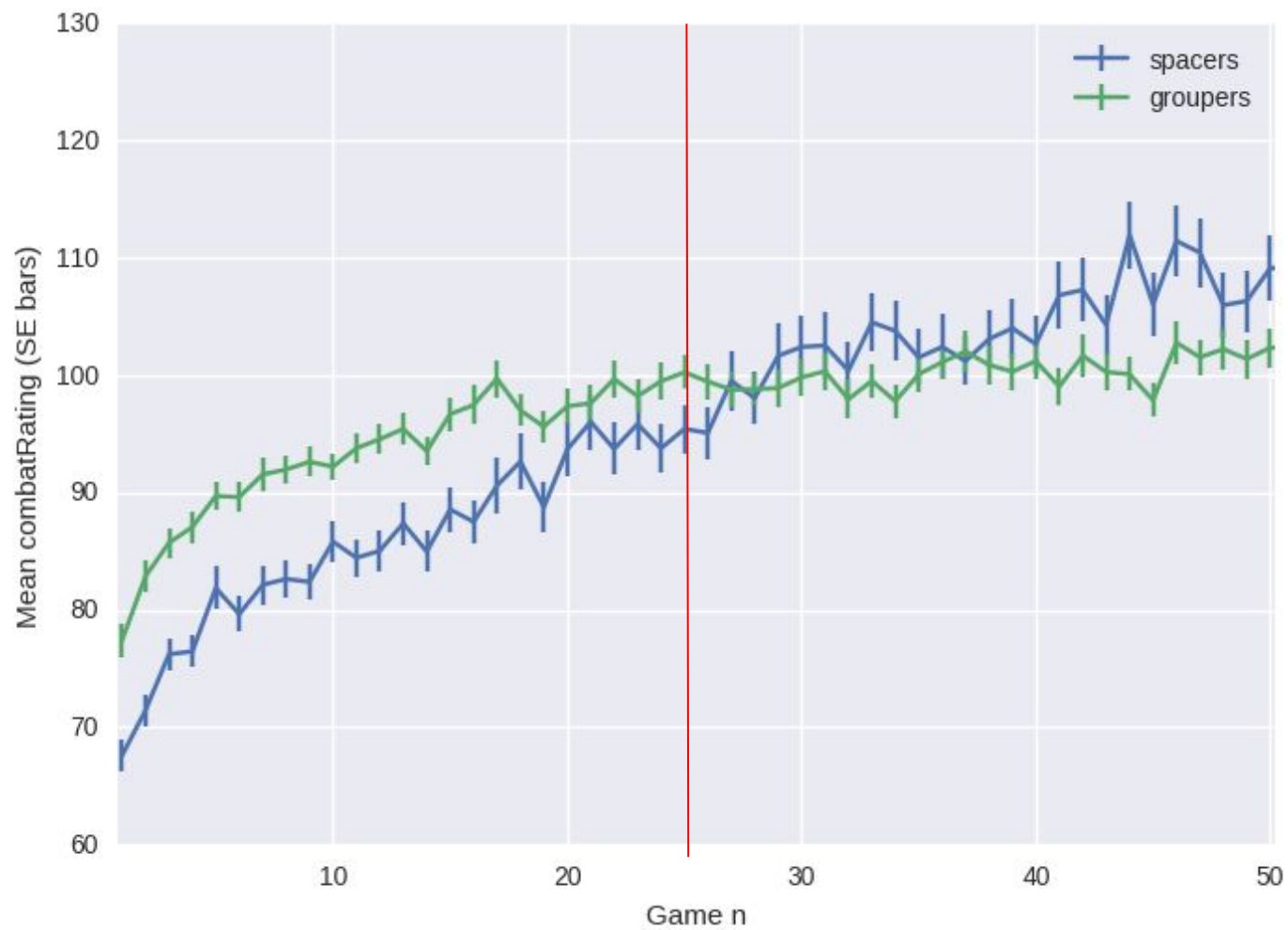
Soccer

Same Data, Different Conclusions

Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.



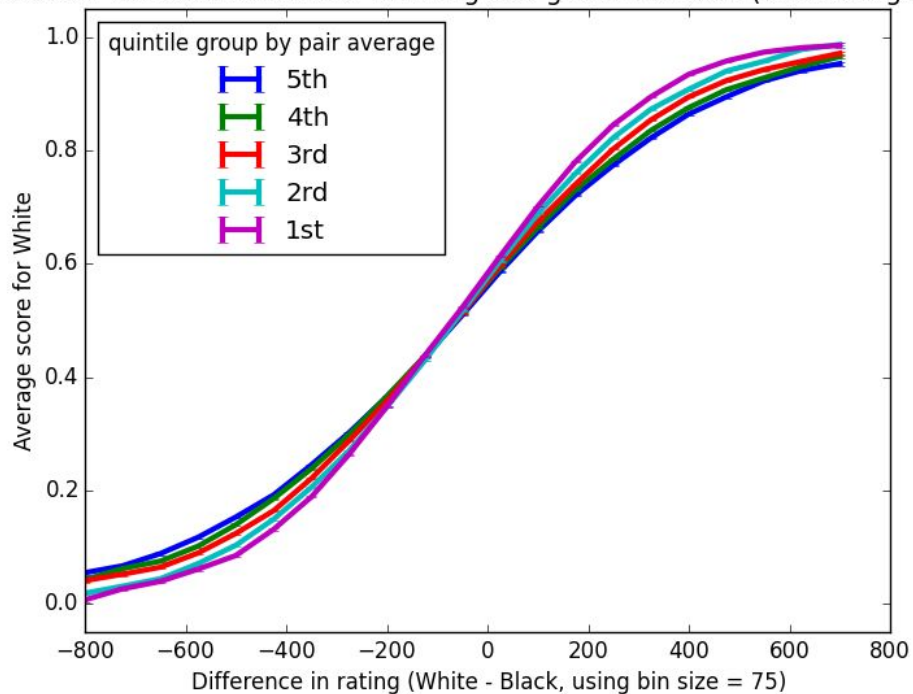
Destiny



Chess

Chess!

Relation between difference in rating and game outcome (6723849 games)



Case study: Stereotype Threat

And mental rotation...

Logic: Stereotype threat one of many potential causes for the observed sex difference.

And chess...

Maass et al (2008) women who believed they were playing men (at internet chess) less likely to win

Rothgerber & Wolsiefer (2014) women less likely to win when playing men, especially if outmatched

Maass, A., D'Ettolè, C., & Cadinu, M. (2008). Checkmate? the role of gender stereotypes in the ultimate intellectual sport. *European Journal of Social Psychology*, 38 (2), 231–245.

Rothgerber, H., & Wolsiefer, K. (2014). A naturalistic study of stereotype threat in young female chess players. *Group Processes & Intergroup Relations*, 17 (1), 79–90.

Robust finding?

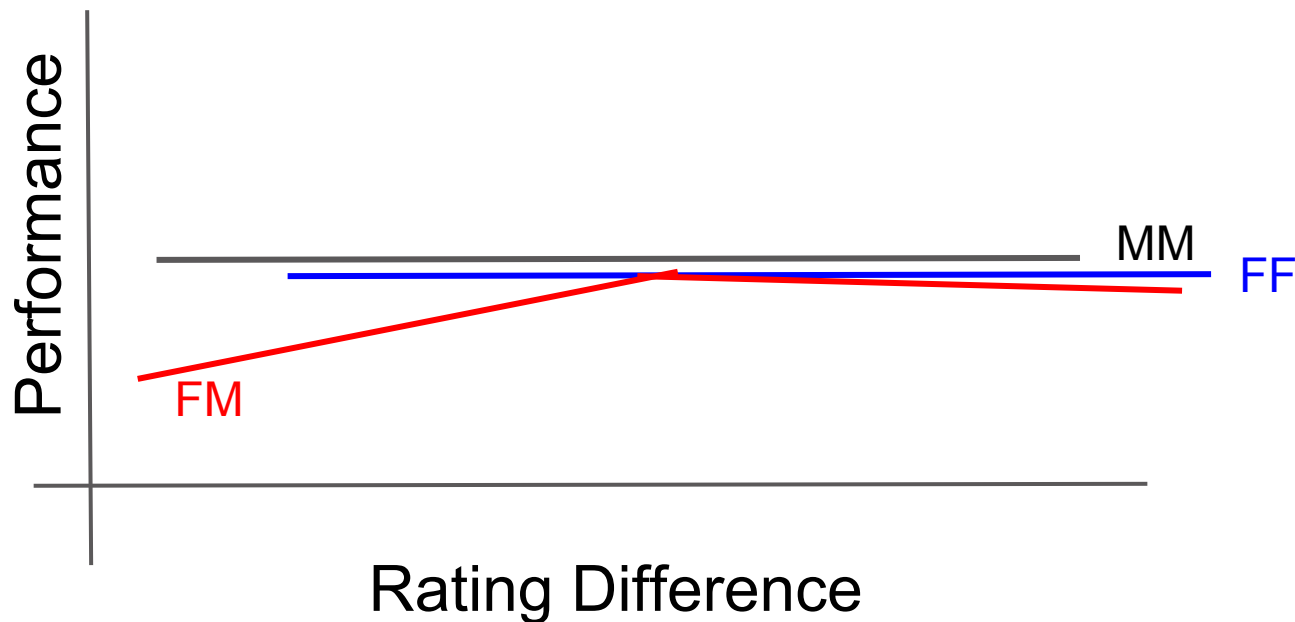
Maass et al (2008) women who believed they were playing men (at internet chess) less likely to win **n = 42**

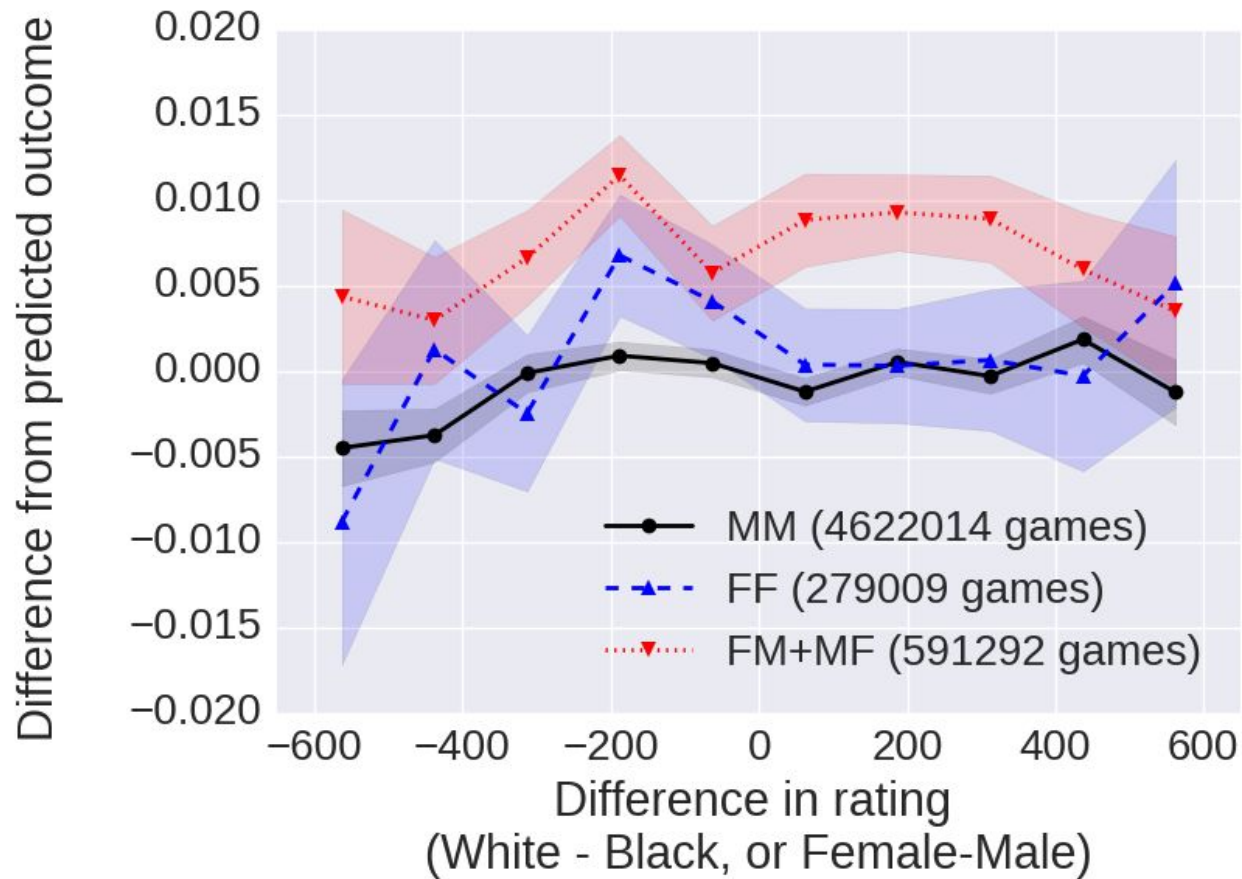
Rothgerber & Wolsiefer (2014) women less likely to win when playing men, especially if outmatched **n = 219**

Publication bias? Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? a meta-analysis. *Journal of school psychology*, 53 (1), 25–44.

FIDE chess tournament data

4.6 million games, n (women) = **12,892**





Stafford, T. (in review) Stereotype lift: female chess players outperform expectations when playing men.
<https://osf.io/aeksv/>