

CSC2552: Review 4, Paper 2

Due on February 13

Ashton Anderson

498 words

Thomas Hollis

Paper 2

This paper, by Dietvorst, Simmons and Massey is a *controlled digital lab experiment* whose original research question is to explain when people use human forecasters rather than superior algorithms and why people have algorithm aversion. The main result of this paper is the revelation that users tend to experience algorithm aversion when they are shown the algorithm's mistakes.

The main weakness of this paper stems from its five small-scale experiments. Since candidate numbers ranged from 1,036 (study 3b) to only 206 (study 2), we are dealing with very small sample sizes exacerbated by further subdivision into separate groups, yielding batches with as few as 67 candidates. An alternative approach could have been the *partner-with-the-powerful* approach which would have allowed a much stronger statistically significant experiment at much larger scale. This is a common limitation of the DIY approach, however this compromise seems to have been made to allow more customisability and more qualitative data collection. Another notable weakness of this paper is the lack of identification of a consistent underlying *mechanism*. An alternative to this approach would have been to have an exit survey asking users to explain why they chose humans even if the algorithm was the most performant. This would have allowed the authors to factually demonstrate the underlying mechanism behind algorithmic aversion which was only hypothesised in this paper. Unfortunately this will require compromise as this approach takes more time for both subjects and designers of the experiment. Fortunately, later publications in this academic field have picked up on this limitation and a paper was subsequently published by the same authors attempting to explain such a mechanism [1]. In [1], experiments revealed that humans trust algorithms they can slightly modify, hence the principal underlying *mechanism* behind algorithm aversion is 'lack of control'.

Conversely, a significant strength of this paper is the *armada strategy* methodology used, as explained in Bit By Bit [2]. Such a strategy leverages multiple smaller experiments rather than one large one which, in this case, is mostly beneficial as it allows this paper to increase its *external validity* by demonstrating generalisations across multiple tasks and multiple reward systems in two different experimental conditions. In addition, effort was made in this paper to demonstrate *construct validity* by providing external examples of unreasonable desires for algorithmic perfection. The Nate Silver example has now been proved particularly well chosen as Nate Silver's 2012 algorithm failed catastrophically in the subsequent election forecast of 2016 [3].

The implications of this paper were key in coining the term 'algorithm aversion' and stimulated numerous subsequent papers [1]. My interpretation of these results does however differ from those of the authors. I believe that telling candidates the algorithm was 'statistical' rather than using a less rigid term, like 'machine-learning', significantly impacted the results. I also believe these results are likely to no longer hold in the modern era of the AI boom. This lack of longitudinal analysis is nonetheless a common limitation of *digital lab experiments*.

[1] Dietvorst, B.J., Simmons, J.P. and Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), pp.1155-1170.

[2] Salganik, M. J. (2017). *Bit By Bit: social research in the digital age*. Princeton University Press.

[3] Silver, N. (2016). 2016 Election Forecast. FiveThirtyEight. Available at: projects.fivethirtyeight.com/2016-election-forecast/ (Accessed: 06/02/2019)