A re-analysis of the relationship between "parasite stress" and authoritarianism

Thomas V. Pollet *

Department of Social and Organizational Psychology, VU University Amsterdam, Netherlands
Correspondence: t.v.pollet@vu.nl

A commentary on

Pathogens and politics: further evidence that parasite prevalence predicts authoritarianism
by Murray, D. R., Schaller, M., and Suedfeld, P. (2013). PLoS ONE 8:e62275. doi: 10.1371/journal.pone.0062275

This paper is a commentary on Murray et al. (2013) where, alongside other findings, data from two cross-cultural samples (nations and cultures), supporting a positive relationship between pathogen prevalence and authoritarianism are presented. This commentary is not a theoretical critique (see the comments on Fincher and Thornhill, 2012, for example, or an alternative: Hackman and Hruschka, 2013). Nor is it a methodological critique of work on pathogen stress (e.g., Currie and Mace, 2012; Pollet, 2013; Pollet et al., in press). The goal of this brief commentary is to re-analyze the data with an alternative technique to examine whether the original conclusions are upheld, specifically the claim that indices of pathogen prevalence are positively related to authoritarianism. To this end, I use a relatively novel technique from machine learning: "Conditional inference trees."
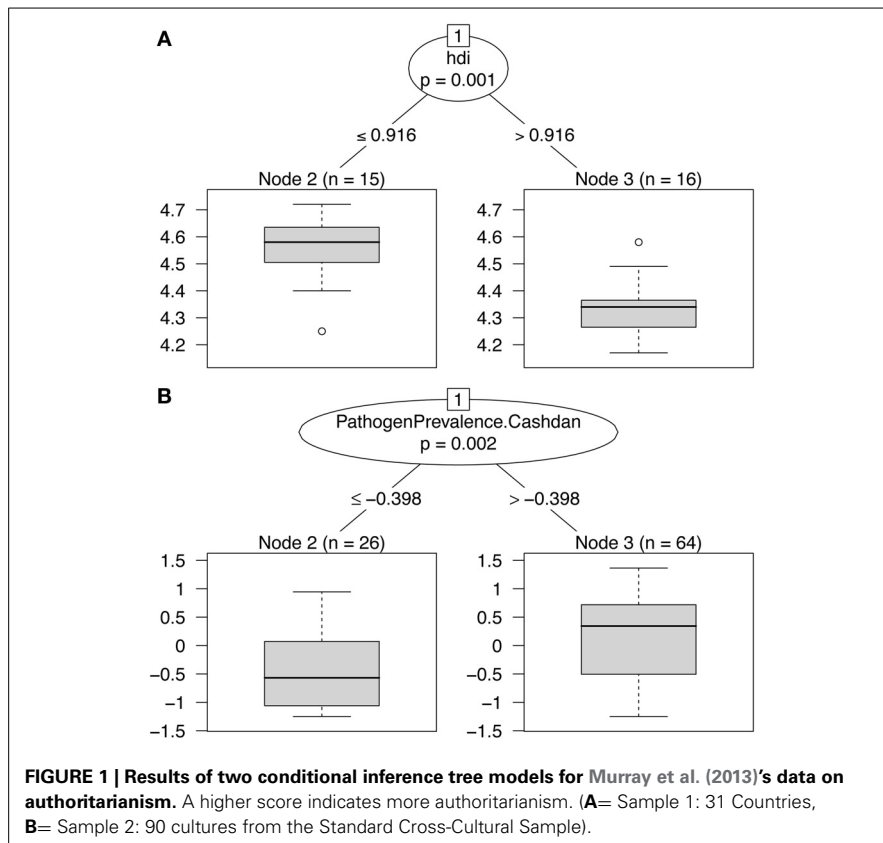
Conditional inference trees are part of "machine learning," a set of algorithms commonly used for data mining (Hastie et al., 2009 for introduction to field). Researchers may be familiar with machine learning algorithms such as "Classification And Regression Trees" (CART; Breiman et al., 1984) and "CHi-squared Automatic Interaction Detection" (CHAID; Kass, 1980; Biggs et al., 1991). These were originally developed to automatically, and "optimally," detect interactions and non-linear patterns in data (reviews in Haughton and Oulabi, 1997; Hastie et al., 2009). Next to other limitations, it is known that these "older" algorithms tend to overfit (Hastie et al., 2009). To this end, researchers have recently developed conditional inference trees, which successfully overcome the limitations of these earlier algorithms (Hothorn et al., 2006, 2010; Strobl et al., 2009; Molnar, 2013). Note that these conditional inference trees can be further nested into "Random Forests" (Breiman, 2001; Liaw and Wiener, 2002; Hothorn et al., 2010). Here I focus on, and limit myself to the "ctree" algorithm. As described by Hothorn et al. (2010:8): "*Roughly, the algorithm works as follows: (1) Test the global null hypothesis of independence between any of the input variables and the response (which may be multivariate as well). Stop if this hypothesis cannot be rejected. Otherwise select the input variable with strongest association to the response. This association is measured by a p-value corresponding to a test for the partial null hypothesis of a single input variable and the response. (2) Implement a binary split in the selected input variable. (3) Recursively repeat steps (1) and (2).*" This algorithm does not overfit and no "pruning" is required (Hothorn et al., 2010). Moreover, it demonstrates good properties for regression and classification modeling. Statistical inference is based on permutation testing, which is non-parametric and relatively free of assumptions (Hothorn et al., 2010; Molnar, 2013). A p-value can be obtained by comparing the observed data, to the (fully) permuted (randomized) data. The "ctree" algorithm has been extensively tested. It seems particularly useful where researchers are faced with a relatively large number of predictors, which could interact or have non-linear effects, and a relatively small sample, as is the case here. For a full description of the algorithm I refer to the papers cited above.

I apply the "ctree" algorithm from "party" to the datasets provided by Murray et al. (2013). I expanded the first dataset containing countries to incorporate their geographical regions (United Nations, 2013) and their economic development classification (United Nations Development Programme, 2012). Note that this dataset also contained the Human Development index (HDI), which was included in the dataset, but does not seem to be part of the original analyses. The only key change made to second dataset, based on the Standard Cross-Cultural Sample (Murdock and White, 1969), is reversing the sign of "authoritarianism.composite," which seemed to not have been reverse-coded. I used *all* possible predictors for the algorithm ($n = 14$ and $n = 25$ respectively) and the same outcome variables as Murray et al. (2013). Analyses were run in R version 3.0.1 (R Development Core Team, 2008) in duplicate with different starting seeds. A significance level of 0.05 (corrected for multiple testing) was used. The data used and script are in Supplementary Material.

For the first sample (**Figure 1A**), the only variable selected was the HDI ($p = 0.001$). The algorithm partitions the data nearly equally, with those countries scoring low on the HDI having high authoritarianism. The pathogen stress variable

FIGURE 1 | Results of two conditional inference tree models for Murray et al. (2013)'s data on authoritarianism. A higher score indicates more authoritarianism. (A= Sample 1: 31 Countries, B= Sample 2: 90 cultures from the Standard Cross-Cultural Sample).

was not selected by the algorithm as a significant predictor. For the second sample (**Figure 1B**), the algorithm selected Cashdan and Steele's (2013) pathogen prevalence variable ($p = 0.002$). It implemented a split whereby a cluster ($n = 26$) of countries with low levels of pathogen prevalence also have low authoritarianism. The majority of these cultures ($n = 15$) were North American.

The current analyses thus provide mixed support for a positive association between pathogen prevalence and authoritarianism. For the first sample, the conclusion is fundamentally different than that of Murray et al. (2013). Only the Human Development Index is found to be a significant predictor of authoritarianism. For the second sample, as in Murray et al. (2013) historical pathogen prevalence seems to be a key predictor. However, it seems that the pattern is non-linear. Rather than an overall linear positive effect of pathogen stress on authoritarianism, as argued by Murray et al. (2013), it appears that there is a cluster of cultures with very low pathogen stress and very low authoritarianism, consisting mainly

of North American cultures. This cluster could account for the findings from their correlational and regression analyses.

This alternative approach via conditional inference trees therefore questions the original findings. This alternative technique has some benefits [e.g., good performance with small sample size and multiple (correlated) predictors], but does not overcome some of the limitations inherent to cross-cultural data (e.g., measurement equivalence). Further research examining the effect of pathogen prevalence on politics remains necessary, both at an individual as well as at a cultural level, preferably using advanced statistical techniques. For now, it seems that the purported effects of pathogen stress on authoritarianism at the level of a nation or culture may well hinge upon the analytical strategy, such as correlations and/or ordinary least squares regressions, employed by researchers.

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIAL

REFERENCES

Biggs, D., De Ville, B., and Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *J. Appl. Stat.* 18, 49–62. doi: 10.1080/02664769100000005

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. New York, NY: CRC press.

Cashdan, E., and Steele, M. (2013). Pathogen prevalence, group bias, and collectivism in the standard cross-cultural sample. *Hum. Nat.* 24, 59–75. doi: 10.1007/s12110-012-9159-3

Currie, T. E., and Mace, R. (2012). Analyses do not support the parasite-stress theory of human sociality. *Behav. Brain Sci.* 35, 83–85. doi: 10.1017/S0140525X11000963

Fincher, C. L., and Thornhill, R. (2012). Parasite-stress promotes in-group assortative sociality: the cases of strong family ties and heightened religiosity. *Behav. Brain Sci.* 35, 65–79. doi: 10.1017/S0140525X11000021

Hackman, J., and Hruschka, D. J. (2013). Fast life histories, not pathogens, account for state-level variation in homicide, child maltreatment, and family ties in the U.S. *Evol. Hum. Behav.* 34, 118–124. doi: 10.1016/j.evolhumbehav.2012.11.002

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The Elements of Statistical Learning*. New York, NY: Springer.

Haughton, D., and Oulabi, S. (1997). Direct marketing modeling with CART and CHAID. *J. Interact. Mark.* 11, 42–52. doi: 10.1002/dir.4000070305

Hothorn, T., Hornik, K., Strobl, C., and Zeileis, A. (2010). *Party: a Laboratory for Recursive Partytioning*. Available online at: http://cran.r-project.org/web/packages/party/party.pdf

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 651–674. doi: 10.1198/106186006X133933

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Appl. Stat.* 29, 119–127. doi: 10.2307/2986296

Liaw, A., and Wiener, M. (2002). Classification and regression by randomforest. *R News* 2, 18–22.

Molnar, C. (2013). *Recursive Partitioning by Conditional Inference*. Available online at: https://github.com/christophM/overview-ctrees

Murdock, G. P., and White, D. R. (1969). Standard cross-cultural sample. *Ethnology* 8, 329–369. doi: 10.2307/3772907

Murray, D. R., Schaller, M., and Suedfeld, P. (2013). Pathogens and politics: further evidence that parasite prevalence predicts authoritarianism. *PLoS ONE* 8:e62275. doi: 10.1371/journal.pone.0062275

Pollet, T. V. (2013). Much ado about p. What does a p value mean when testing hypotheses with aggregated cross-cultural data in the field of

evolution and human behavior? *Front. Psychol.* 4:734. doi: 10.3389/fpsyg.2013.00734

Pollet, T. V., Tybur, J. M., Frankenhuis, W. E., and Rickard, I. J. (in press). What can cross-cultural correlations teach us about human nature? *Hum. Nat. Interdiscip. Biosoc. Perspect.*

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* 14, 323–348. doi: 10.1037/a0016973

United Nations. (2013). *Composition of Macro Geographical (continental) Regions, Geographical Sub-regions, and Selected Economic and Other Groupings*. Available online at: https://unstats.un.org/unsd/methods/m49/m49regin.htm

United Nations Development Programme. (2012). *Statistical Annexe: World Economic Situation and Prospects*. Available online at: http://www.un.org/en/development/desa/policy/wesp/wesp_current/2012country_class.pdf

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