

Title: In the World of Big Data, Small Effects Can Still Matter: A Reply to Boyce et al. 2016

Authors: Sandra Matz¹, Joe Gladstone², David Stillwell³

Affiliations:

1 Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, UK.

2 UCL School of Management, London, RM NE3, UK

3 Cambridge Judge Business School, University of Cambridge, Cambridge, CB2 1AG, UK

Correspondence to: sm917@cam.ac.uk; +447858786958; Department of Psychology, Downing Street, Cambridge, CB2 3EB, UK.

Keywords: effect size, happiness, consumption, psychological fit, Big Five personality

We make three points in response to Boyce, Daly, Hounkpatin and Wood (2016). First, we clarify a misunderstanding of the goal of our analyses, which was to investigate the links between life satisfaction and spending patterns, rather than spending volume. Second, we run a simulation study to demonstrate that our results are not driven by the proposed statistical artefact. Finally, we discuss the broader issue of why, in a world of big data, small but reliable effect sizes can be valuable.

Our goal is to study spending patterns, not spending volume.

Boyce et al. argue that our findings do not provide sufficient evidence for the contention that spending *more* money buys happiness, when that money is spent on products and services that fit a person's personality. We agree, because this is not the argument we make in the paper. Our argument is that spending the money one already has on products and services which match one's personality (and thus psychological needs) results in greater happiness. That is, we studied spending *patterns* (what we buy) rather than spending *volume* (how much of it we buy). In fact, we emphasized the relative unimportance of "spending more money" in our measures, analyses, and study designs. First, we considered the fact that many small purchases can result in greater happiness than a few large ones (e.g. Nelson and Meyvis, 2008) in the calculation of the person-basket match variable itself. Rather than weighting all spending categories by the amount spent, we assigned an equal weight to each of them. Second, we highlighted that the coefficients for income and total spending on life-satisfaction are non-significant (Table 3, page 6), suggesting that earning or spending *more* money does not impact life satisfaction and happiness. Finally, we did not vary the amount spent in our experiment, but instead focused exclusively on manipulating the spending pattern (i.e. whether or not the voucher fitted the personality of the recipient).

Our results are not exclusively driven by the proposed statistical artefact.

Boyce et al. (2016) suggest that the effect of the match between customers' shopping baskets and their personality might be driven "by participant personality, which is known to relate to well-being". We also had this concern. Which is why the original analysis controlled for participant personality as well as the overall extremity of profiles and found that the effects of basket-participant match remained stable when controlling for these variables (see Table 3; Matz et al., 2016).

To provide further evidence for our effect's robustness, we ran a simulation on our dataset that is similar to the one reported by Boyce et al. (2016). In 1,000 iterations, we randomly allocated the basket personality calculated for each participant in Study 1 to another participant (i.e., we 'swap' participants' shopping baskets at random). We then calculated the basket-participant match and ran regression analyses including the control variables used by Boyce et al. (Model 1 without total spend) and the original control variables of Model 2 (see online supplementary materials for more details). Out of 1000 iterations, basket-personality match was significant at an alpha level of 0.05 in only 3.5% of cases when including the controls used by Boyce et al., and in only 2.3% of cases when including all control variables of Model 2. The average coefficient of the randomly generated basket-participant match was $B = 0.013$ in Model 1 ($SD = 0.03$; average $\beta = 0.018$; left panel of Figure 1) and $B = -0.004$ in Model 2 ($SD = 0.03$; average $\beta = -0.005$; right panel of Figure 1). This compares to coefficients of $B = 0.06$ and $\beta = 0.10$ reported for real basket-participant match in our original analysis (using the same controls). We take this as strong evidence that the relationship between basket-participant match and life satisfaction is not (or at least not exclusively) driven by the confounding artefacts suggested by Boyce et al. (2016).

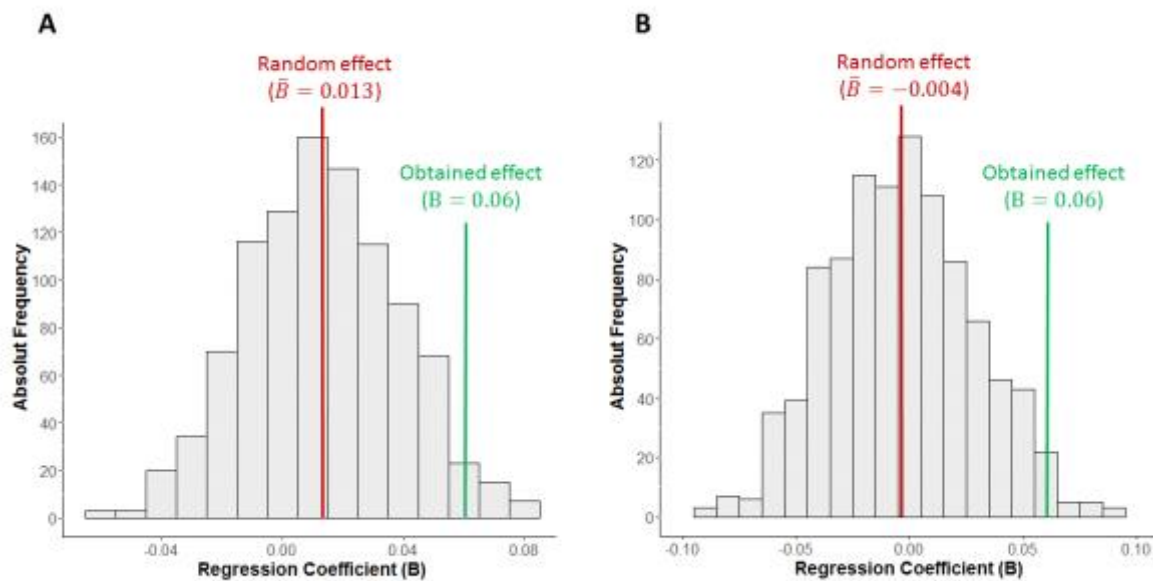


Figure 1. Distribution of regression coefficients (B) for randomly distributed basket personality across 1,000 iterations, using the controls of Boyce et al. (left; original Model 1 without total spend) and the original controls of Model 2 (right).

The importance of small effects.

Boyce et al. (2016) argue that the relationship we report between a consumer's purchases and their wellbeing is of no practical relevance. This raises an important debate about the magnitude of meaningful effect sizes.

Traditional social science research uses small samples (Bertamini & Munafò, 2012; Button et al., 2013), which require large effect sizes to reach statistical significance and be published. Large effect sizes are more likely to be replicable (Open Science Collaboration, 2015), but focusing exclusively on them hinders a nuanced exploration of complex psychological phenomena such as life satisfaction, which are unlikely to be explained by a few strong predictors. Instead of dismissing small effect sizes altogether, researchers should explore new methodologies that enable the discipline of psychology to build a body of small but robust

predictors of behavior. New computational social science approaches (Big Data; see e.g. Kosinski, Matz, Gosling, Popov, & Stillwell, 2015), for example, make it possible to collect behavioral data at an unprecedented scale, providing psychologists with opportunities to identify weak but reliable signals in a complicated world. However, the danger of using these new methods is that even trivial effects will become statistically significant with large enough samples. Therefore, how can we distinguish between small effects which are meaningful, from those which are not?

This requires a scientist's judgement rather than simply following arbitrary cut-offs. A full review of situations when small effect sizes are important would require more space than available in a commentary, but other authors have already discussed some situations. Cortina and Landis (2009), for example, note that small effects can suggest strong support for a given phenomenon if they (1) occur in the context of intentionally inauspicious designs, (2) challenge fundamental theoretical assumptions, and (3) have enormous cumulative consequences. One illustration of the latter is Abelson's Paradox (Abelson, 1985). Investigating the batting performance of major-league baseball players, Abelson showed that historical batting performance was barely predictive of a player's performance each time they stepped up to the batting plate (less than 0.5%). Yet, the cumulative effect over several hundred bats is so meaningful in practice that it leads to some batters being paid 100,000% more than others.

In our paper, one challenge in determining the meaning in effect sizes is that it is difficult to directly compare them across studies with different methodologies, such as self-reported surveys and objective behavioral data. Study 1 is a 'Big Data' field study, where over 76,000 transactions are mined from customers' bank accounts and where the measurement of wellbeing is far removed, in both context and time, from when the bank customers' purchases were being made. Such a procedure is likely to result in smaller – yet probably more realistic -

effect sizes than panel surveys because they eliminate well-documented response biases (e.g., consistency motive, covariation bias, or common method variance). For example, Powdthavee (2008; cited by Boyce et al., 2016), reports a large effect of social contact on subjective wellbeing in the British Household Panel Survey. While social contact is surely important for wellbeing, the reported effect size might be over-estimated because respondents are subjectively reporting both their social contact and their wellbeing at the same time. Respondents who felt negatively about their lives, for example, might be less likely to remember and report short and superficial social contact, such as saying “hello” to a neighbor, than respondents who felt more positive (mood congruence; Bower, 1981). Indeed, the results of our Study 2 underline the importance of considering the methodological context when discussing a given effect size; Study 2’s behavioral field experiment found a medium to large effect size for the personality-fit interaction ($\beta = 0.38$).

A final consideration when deciding whether a small effect may have practical relevance is the extent to which the factor can be manipulated. In their commentary, Boyce et al. (2016) point to other factors that are stronger predictors of wellbeing (in terms of standardized effect sizes) such as relationships (Powdthavee, 2008), personality (Diener & Lucas, 1999) or stable employment (McKee-Ryan, Song, Wanberg, & Kinicki, 2005). Such factors certainly play an important role in predicting life satisfaction. However, personality cannot easily be changed (McCrae & Costa, 1994) and it is often outside the control of an unemployed person to arrange for stable employment. Consumption choices, on the other hand, are usually under the control of the individual and can be changed relatively easily. Even the very poorest groups in the world (i.e., those living on less than \$2 per day), spend substantial amounts on discretionary goods, such as entertainment, celebrations, clothing and tobacco (Banerjee & Duflo, 2007). Given that consumption is such a universal phenomenon, even small changes can have a big impact if they occur on a large scale. A 1% increase in life satisfaction may be

negligible for a single consumer, but if a retail giant like Amazon aimed at making its customers happier by personalizing their product recommendations, a 1% increase in life satisfaction across its 244 million customers (Kline, 2014) could turn a small effect size into a huge social effect.

Psychologists have typically focused on how their findings apply to individuals. However, by providing the opportunity to understand and influence the behaviors of billions of people around the world, the era of big data encourages – and possibly requires - us to think bigger. In this new world, small effects can still matter.

References

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, 97(1), 129.
- Banerjee, A. V., & Duflo, E. (2007). The economic lives of the poor. *The Journal of Economic Perspectives*, 21(1), 141–167.
- Bertamini, M., & Munafò, M. R. (2012). Bite-size science and its undesired side effects. *Perspectives on Psychological Science*, 7(1), 67–71.
- Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36(2), 129. JOUR.
- Boyce, C., Daly, M., Hounkpatin, H., & Wood, A. (n.d.). Money may buy happiness, but so little that it often doesn't matter. *Psychological Science*.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.
- Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Cortina, J. M., & Landis, R. S. (2009). When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 287–308). Routledge New York.
- Diener, E., & Lucas, R. E. (1999). 11 Personality and Subjective Well-Being. *Well-Being: Foundations of Hedonic Psychology*, 213.
- Kline, D. B. (2014). How many customers does Amazon have? Retrieved from <http://www.fool.com/investing/general/2014/05/24/how-many-customers-does-amazon->

have.aspx

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a Research Tool for the Social Sciences. *American Psychologist*, *70*(6), 543–556.

McCrae, R. R., & Costa, P. T. (1994). The stability of personality: Observations and evaluations. *Current Directions in Psychological Science*, *3*(6), 173–175.

McKee-Ryan, F., Song, Z., Wanberg, C. R., & Kinicki, A. J. (2005). Psychological and physical well-being during unemployment: a meta-analytic study. *Journal of Applied Psychology*, *90*(1), 53.

Powdthavee, N. (2008). Putting a price tag on friends, relatives, and neighbours: Using surveys of life satisfaction to value social relationships. *The Journal of Socio-Economics*, *37*(4), 1459–1480.

Supplementary Online Material

Predictors	Simulation Model 1				Simulation Model 2			
	B	SE(B)	β	t	B	SE(B)	β	t
B-P-match	0.01	0.03	0.02	0.46	-0.004	0.03	-0.01	-0.12
Income (log)	0.08	0.05	0.07	1.70	0.05	0.06	0.04	0.77
Gender	0.02	0.07	0.03	0.32	-0.03	0.08	-0.04	-0.40
Age	-0.01	0.00	-0.09	-2.23	-0.01	0.003	-0.12	-2.69
Total spend (log)	-	-	-	-	0.03	0.07	0.03	0.42
Person-O	-	-	-	-	0.04	0.03	0.04	1.16
Person-C	-	-	-	-	<0.001	0.04	0.00	0.001
Person-E	-	-	-	-	0.08	0.04	0.09	2.26
Person-A	-	-	-	-	0.01	0.03	0.01	0.34
Person-N	-	-	-	-	-0.23	0.04	-0.27	-6.09
Extremity	-	-	-	-	-0.05	0.11	-0.02	-0.47
Product-O	-	-	-	-	-0.11	0.07	-0.13	-1.67
Product-C	-	-	-	-	0.07	0.06	0.09	1.18
Product-E	-	-	-	-	0.16	0.09	0.19	1.87
Product-A	-	-	-	-	0.12	0.08	0.13	1.53
Product-N	-	-	-	-	0.04	0.11	0.05	0.41

Note. In Model 1 the B-P match predictor reached significance in 35 out of the 1000 iterations. In Model 2 the B-P match predictor reached significant in 23 out of 1000 iterations