The Science Behind the Magic? The Relation of the Harry Potter "Sorting Hat Quiz" to Personality and Human Values

Lea Jakob[1], Eduardo Garcia-Garzon[2], Hannes Jarke[3], & Fabian Dablander[4]

[1] Assessment Systems International, Prague, Czech Republic
[2] Universidad Autónoma de Madrid, Madrid, Spain
[3] RAND Europe, Cambridge, UK
[4] Department of Psychological Methods, University of Amsterdam, The Netherlands

Author note

Correspondence concerning this article should be addressed to Lea Jakob, E-mail:

lea.jakob.psy@gmail.com

Appendix 2

Crysel, Cook, Schember, and Webster (2015) hypothesized that Gryffindor would score higher on Extraversion, Ravenclaw higher on Intellect, Hufflepuff higher on Agreeableness, Emotional Stability, and Conscientiousness, and Slytherin higher on Machiavellianism, Narcissism, and Psychopathy compared to the respective mean of the other Houses. In order to test such set of hypotheses, the authors conducted classical ANOVAs using Helmert contrasts to test whether, for example, Gryffindor is higher on Extraversion compared to the *mean of the other Houses*. Sticking to this example for ease of exposure, Crysel and colleagues' approach does not only test this contrast, but also whether Slytherin is higher on Extraversion than Hufflepuff and Ravenclaw, and whether Hufflepuff is higher on Extraversion as Ravenclaw. The authors hypothesized that Gryffindor would score higher on Extraversion, Ravenclaw higher on Intellect, Hufflepuff higher on Agreeableness, Emotional Stability, and Conscientiousness, and Slytherin higher on Machiavellianism, Narcissism, and Psychopathy compared to the respective mean of the other Houses.

Testing whether, for example, Gryffindor scores higher on Extraversion than the *mean of the other Houses* strikes us as inappropriate. In fact, it is perfectly possible for that to be the case even when Ravenclaw and Slytherin have higher scores than Gryffindor—Hufflepuff just needs to score really low on Extraversion (see Appendix 1). Thus, ANOVAs with Helmert contrasts do not test whether Gryffindor is the most extraverted House, which is arguably what we want to know.

We expect, similar to the original paper, that:
- o Gryffindor will score *highest* on Extraversion
- o Ravenclaw will score *highest* on Intellect
- o Hufflepuff will score *highest* on Agreeableness, Emotional Stability, and Conscientiousness
- o Slytherin will score *highest* on Machiavellianism, Narcissism, and Psychopathy.

Using Extraversion as an example, our example hypothesis can be formally stated as

$$H_{\text{Extraversion}}: \mu_{\text{G}} > (\mu_{\text{S}}, \mu_{\text{R}}, \mu_{\text{H}})$$

where μ_x indicates the mean of the respective House. We assume a constant error variance across Houses which leads to an ANOVA model. A classical approach would be to conduct three t-tests: $\mu_{\text{G}} > \mu_{\text{S}}$, $\mu_{\text{G}} > \mu_{\text{R}}$ and $\mu_{\text{G}} > \mu_{\text{H}}$ —but again it is not clear what one would make

from a non-significant result in one of these tests. Therefore, we opted for a different statistical approach. To better understand it, we briefly sketch the main ideas behind Bayesian inference.

Bayesian Inference

Bayesian statistics entails quantifying one's uncertainty about the world using probability. Statistical inference reduces simply to applying the rules of probability theory. For an introduction to Bayesian inference from this angle, see Etz and Vandekerckhove (2018).

The method dictated by probability theory for Bayesian hypothesis testing is the Bayes factor (Kass & Raftery 1995). Let the null hypothesis be instantiated by a restricted model, M_0, while the alternative hypothesis is specified by M_1. Parameters within the models—here denoted by —are estimated using Bayes' rule

$$p(\mathbf{\mu}|\mathbf{y}, M_0) = \frac{p(\mathbf{y}|\mathbf{\mu}, M_0)p(\mathbf{\mu}|M_0)}{\int p(\mathbf{y}|\mathbf{\mu}, M_0)p(\mathbf{\mu}|M_0)\mathrm{d}\mu}$$

An equivalent expression exists for M_1. Note that, using the *sum rule* of probability, the denominator can be written as $p(\mathbf{y}|M_0)$, which is sometimes called the *marginal likelihood*.

The ratio of marginal likelihoods for two models yields the Bayes factor

$$\underbrace{\frac{p(M_0|\mathbf{y})}{p(M_1|\mathbf{y})}}_{\text{Posterior odds}} = \underbrace{\frac{p(\mathbf{y}|M_0)}{p(\mathbf{y}|M_1)}}_{\text{Bayes Factor}} \cdot \underbrace{\frac{p(M_0)}{p(M_1)}}_{\text{Prior Odds}} .$$

The Bayes factor denotes the shift in the models' beliefs given the data. Because the marginal likelihood is an integration over the likelihood with respect to the prior, the Bayes factor instantiates an automatic Ockham's razor—complex models can predict more data patterns than simple models, i.e., their prior distribution is spread out over a larger space, leading to decreasing weights multiplied with the likelihood, resulting in an overall reduced marginal likelihood (Vandekerckhove, Matzke, & Wagenmakers, 2015). More generally, (weakly informative) priors help guard against overfitting. We use the BayesFactor R package for testing the order-constraints (Morey, Rouder, & Jamil, 2014).

Bayesian Order-constrained Inference

Let M_0 denote the null model, M_f denote the full model in which each mean is free to vary, and M_r denote the order-constrained model. Formally

$$M_0 : \mu_G = \mu_S = \mu_R = \mu_H$$

$$M_f : \mu_G, \mu_S, \mu_R, \mu_H$$

$$M_r : \mu_G > (\mu_S, \mu_R, \mu_H)$$

Therefore, our problem reduces to finding the most plausible model among those three. Note that the Bayes factor is transitive, such that

$$\frac{p(\mathbf{y}|M_r)}{p(\mathbf{y}|M_0)} = \frac{p(\mathbf{y}|M_r)}{p(\mathbf{y}|M_f)} \cdot \frac{p(\mathbf{y}|M_f)}{p(\mathbf{y}|M_0)}.$$

Therefore, given the three models above, Bayes factors can be computed for each model comparison. To compute the Bayes factor of M_f against M_0, we use the BayesFactor R package (Morey et al., 2014) which uses so-called 'default priors' (see Morey, Wagenmakers, & Rouder, 2016; Rouder, Morey, Speckman, & Province, 2012), that is, an uninformative Jeffreys' priors over, in this case, the nuisance parameter σ^2 and a Cauchy prior with scale parameter r over the effect of *House*. The scale parameter indicates the width of the Cauchy prior; the higher r, the wider the distribution, the more we expect large effect sizes a priori. We report a sensitivity analysis varying r from .1 to 1.5 using a step size of .05.

To compute the Bayes factor of M_r against M_f, note that the Bayes factor is the ratio of the posterior odds and the prior odds (to see this, refactor the equation above). Uninformative prior odds for order-restricted models are simply one over the number of possible order restrictions. For four means, the number of possible order restrictions is 4! = 24. The posterior odds can be estimated by the proportion of the samples of the posterior distribution of the parameters (in this case, the s for the respective *Houses*) that are in line with the order-restriction (Klugkist & Hoijtink, 2007; Klugkist, Laudy, & Hoijtink, 2005). Note that this Bayes factor is bounded by 24.

References

Crysel, L. C., Cook, C. L., Schember, T. O., & Webster, G. D. (2015). Harry Potter and the Measures of Personality: Extraverted Gryffindors, Agreeable Hufflepuffs, Clever Ravenclaws, and Manipulative Slytherins. *Personality and Individual Differences*, *83*, 174–179. doi:10.1016/j.paid.2015.04.016

Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review, 25*(1), 5-34.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Klugkist, I., & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, *51*(12), 6367–6379. doi:10.1016/j.csda.2007.01.024

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, *10*(4), 477. doi:10.1037/1082-989X.10.4.477

Morey, R. D., Rouder, J. N., & Jamil, T. (2014). BayesFactor: Computation of bayes factors for common designs. *R Package Version 0.9*, *8*.

Morey, R. D., Wagenmakers, E.-J., & Rouder, J. N. (2016). Calibrated Bayes factors should not be used: A reply to Hoijtink, van Kooten, and Hulsker. *Multivariate Behavioral Research*, *51*(1), 11–19. doi:10.1080/00273171.2015.1052710

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. doi:10.1016/j.jmp.2012.08.001

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford library of psychology. The Oxford handbook of computational and mathematical psychology* (pp. 300-319). New York, NY, US: Oxford University Press.