Do Individual Effects Reflect Quantitative or Qualitative Differences in Cognition?

ABSTRACT

Rouder and Haaf (2020) posed the important question if there are some individuals whose behavior is not in accordance with well-established experimental effects and whether these individual differences are quantitative or qualitative in nature. In our commentary, we discuss the distinction between quantitative and qualitative individual differences and between individual and average causal effects and come to the conclusion that this is not a new question, but in fact one that has already been discussed by Gordon W. Allport (1937) and Donald B. Rubin (1974, 1978). Moreover, we critically examine their proposed rule of thumb to collect about 100 trials per experimental condition to reliably measure individual differences in typical experimental effects. Based on simulation results, we suggest to not rely on any general rule of thumb, but to use simulation studies and the convenient quid function provided by the authors to make more informed decisions regarding trial numbers for specific experimental designs.
Psychological science aims to identify laws that govern human thinking, feeling, and behavior. Unlike the natural sciences, however, we cannot directly measure psychological constructs. Cognitive psychologists study how certain experimental manipulations designed to induce processing demands for these constructs affect performance in an experimental task. Typically, most individuals behave in a manner consistent with well-established experimental effects, while a minority does not show the expected effect or even shows a reverse effect. If this were itself a robust phenomenon – if certain individuals behaved in ways inconsistent with “universal” psychological laws – it would have serious and far-reaching implications for theory building. In physics, for example, a far greater research effort is now devoted to the study of deviations from than confirmations of theoretical predictions. In their paper, Rouder and Haaf (2020) pose the question if these individual deviations from experimental effects are substantive or if they can be accounted for by measurement error. They propose a hierarchical model to distinguish these substantive deviations from measurement error and conclude that there are certain experimental effects from which individuals deviate in a qualitative manner by showing effects in the other direction. We appreciate the elegance of the statistical solution proposed by Rouder and Haaf (2020) and the accessibility of the accompanying R functions. In our commentary, we will focus on two more conceptual parts of their paper: The distinction between quantitative and qualitative individual differences and their proposed rule of thumb for trial number considerations.

**QUANTITATIVE VS. QUALITATIVE INDIVIDUAL DIFFERENCES AND INDIVIDUAL VS. AVERAGE EXPERIMENTAL EFFECTS**

The question of quantitative vs. qualitative individual differences has a long tradition in psychology. Already Gordon W. Allport (1937) made a clear distinction between qualitative and quantitative individual differences. He suggested that an “individual trait” is peculiar to a person and as such, has no metric and no population distribution and thus is truly qualitative in nature (c.f. idiographic approach). In contrast, a “common trait” is an aspect of personality that has a metric and that allows a numerical comparison between individuals and thus is quantitative in nature (c.f. nomothetic approach). From this perspective, not much of the present authors’ proposal is “qualitative” because they ground their “qualitative individual differences” on negative values of an effect size measure, which allows a comparison between individuals, has a population distribution, and is therefore quantitative in nature (it corresponds clearly to the common trait in Allport’s terminology).

To expand on this issue, the authors did not distinguish between the level of measurement and the level of theoretical interpretation of the measure. The general and individual effect is something that can be measured and is therefore quantitative (elsewise, it would not have been possible to place individuals along a common dimension denoted as “effect” in Figure 1). Perhaps the theoretical interpretation of certain effect size values must shift from one theory to another, depending on the range of effect size values, allowing for a qualitative difference in the interpretation of the values. However, this does not make the values qualitative in nature, they still are quantities. In this context it must be seen that the authors ground their decision about “qualitative individual differences” solely on the outcome of statistical model testing, irrespective of the measured constructs or theoretical considerations. Taking this stance literally, one could measure the temperature in a sample of villages in January and February and find that it increased in most villages but decreased in others. Does this indicate that there are qualitative differences between the villages in terms of theoretical explanation of temperature, i.e. do we need another physics in some of the villages? The answer is certainly no.

We also question that “laterality” is a good example for “obvious qualitative individual differences” (p. 11). Handedness comes in grades, ranging from extreme right-handedness over ambidexterity to extreme left-handedness. As a matter of fact, only few individuals use only one hand exclusively for all one-handed activities (e.g., Springer & Deutsch, 1993). Therefore, handedness questionnaires typically allow to locate a given person on a continuum between extreme left-handedness through equal use of both hands to extreme right-handedness (e.g., Oldfield, 1971). Hence, handedness is not a qualitative but a quantitative trait. The same case can be made for “preferences”, which the authors use as another example for “qualitative” individual differences.
We would also like to highlight that Rubin (1974) published an authoritative paper on the analysis of causal effects, where he made a clear distinction between “individual causal effects” and the “average causal effect”. This paper also provides a discussion of the consequences of the homogeneity vs. heterogeneity of individual causal effects, including outliers, on the interpretation of the average causal effect. He used this distinction between individual and average causal effects for an extended treatment of a Bayesian analysis of randomized studies (Rubin, 1978). Later, Steyer (2005) translated Rubin’s ideas into a contemporary structural equation modeling framework, which allows to estimate the variance as well as the numerical values of individual causal effects (needless to say, random noise is controlled via latent variable modeling). We believe that this work is highly relevant to Rouder’s and Haaf’s (2020) considerations.

**Figure 1** Observed (blue) and model-estimated (red) effects for lower degrees of interindividual variability.
In their introduction (p. 2), the authors stated: “Rather than ask an on average question, we advocate a new question: Are there qualitative individual differences?” It is our impression that this question is not new but has been extensively treated in the literature. The target article does not do this previous work justice by ignoring the important milestones made by Alport (1937), Rubin (1974, 1978), and Steyer (2005).

**TRIAL NUMBERS**

Rouder and Haaf (2020) emphasized the need to collect sufficient data to reliably measure interindividual differences in experimental effects. As a rule of thumb, they proposed to collect data from at least 100 trials per experimental condition. We ran a simulation study to illustrate
how different factors (number of trials, trial noise, and the degree of interindividual variation) affected the visibility of individual differences. For this purpose, we used the R package \textit{flankr} (Grange, 2016) to simulate behavioral data from the dual-stage two-phase (DSTP) model (Hübner et al., 2010), which is a mathematical model of attention and decision processes in the Eriksen flanker task (Eriksen & Eriksen, 1974). We realized four simulation factors: Number of trials per condition (25, 50, 100, 200 vs. 400), trial noise (0ms, 50ms, 100ms, 200ms, 400ms), and interindividual variability of the stimulus-selection drift rate $v_{\text{StimSelection}}$, which is the model parameter that describes the speed and efficiency with which individuals focus their attention on the target stimulus (0.00, 0.10, 0.20, .40).

As expected, observed and model-estimated individual experimental effects converged when there was either a high number of trials or a low degree of trial noise (see \textbf{Figures 1} and \textbf{2}). As soon as there was some true interindividual variation in the latent cognitive process, at least 200 trials per condition were needed to compensate for the effects of moderate-to-high trial noise on the measurement precision of the observed individual experimental effects (see \textbf{Figures 1B} and \textbf{2}). This result confirms our hunch to not rely on a rule of thumb of collecting at least 100 trials per condition. Instead of replacing it with another arbitrary rule of thumb, we urge to use simulation studies and the convenient \textit{quid} function provided by the authors of the target article to make more informed decisions regarding trial numbers for specific experimental designs whenever possible.

**COMPETING INTERESTS**

The authors have no competing interests to declare.

**AUTHOR CONTRIBUTIONS**

A.L.S. and D.H. contributed equally to the manuscript. A.L.S. and D.H. discussed the outline of the commentary. A.L.S. wrote the introductory section and the section on trial numbers. D.H. wrote the section on qualitative vs. quantitative individual differences. A.L.S. and J.G. ran the simulation study. J.G. visualized results from the simulation study. All authors reviewed and commented on the manuscript.

**AUTHOR AFFILIATIONS**

Anna-Lena Schubert  \orcid.org/0000-0001-7248-0662
Institute of Psychology, University of Mainz, Mainz, Germany

Dirk Hagemann  \orcid.org/0000-0002-1982-3248
Institute of Psychology, Heidelberg University, Heidelberg, Germany

Jan Göttmann  \orcid.org/0000-0002-3487-4262
Institute of Psychology, University of Mainz, Mainz, Germany

**REFERENCES**

Allport, G. W. (1937). Personality. A psychological interpretation. New York: Henry Holt and Company.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception & Psychophysics, 16(1), 143–149. DOI: https://doi.org/10.3758/BF03203267

Grange, J. A. (2016). \textit{flankr}: An R package implementing computational models of attentional selectivity. \textit{Behavior Research Methods}, 48(2), 528–541. DOI: https://doi.org/10.3758/s13428-015-0615-y

Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. \textit{Psychological Review}, 117(3), 759–784. DOI: https://doi.org/10.1037/a0019471

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. \textit{Neuropsychologia}, 9(1), 97–113. DOI: https://doi.org/10.1016/0028-3932(71)90067-4

Rouder, J. N., & Haaf, J. M. (2020) Are There Reliable Qualitative Individual Difference in Cognition? \textit{Journal of Cognition}, 4(1): 46, 1–16. DOI: https://doi.org/10.31234/osf.io/3ezmw

Rubin, D. B. (1974). Estimating causal effects of treatments in randonimized and nonrandomized studies. \textit{Journal of Educational Psychology}, 66(5), 688–701. DOI: https://doi.org/10.1037/h0037350

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. \textit{Annals of Statistics}, 6(1), 34–58. DOI: https://doi.org/10.1214/aos/1176344064
