

REVIEW

By Assoc. Prof. Dr. Nikolay Rumenov Rachev, SU "St. Kliment Ohridski, 3.2. Psychology

On a dissertation for obtaining a doctoral degree

In the professional field 3.2. Psychology

Dissertation Topic: "Critical Conditions for Observing the Inverse Base-Rate Effect"

PhD candidate: Yolina Petrova

Research supervisor: Assoc. Prof. Dr. Penka Hristova

Aspects of the review

1. Significance of the researched problem in scientific and scientific-applied terms.

The dissertation addresses an important issue in current research on categorization: the inversed base-rate effect (IBRE). This puzzling effect has engendered much research in the last 30 years, yet there is no definite answer regarding its underlying mechanisms.

2. Precisely formulated aims and tasks of the dissertation.

The aims of the dissertation are clearly stated: to critically test the two dominant accounts of the IBRE: the association-based account (Kruschke, 1996) and the rule-based account (Juslin et al., 2001).

3. Evidence that the dissertation work was developed independently and does not literally repeat the topic and a significant part of the content of the work presented for the acquisition of the educational and scientific degree "doctor".

Not applicable

4. Degree of knowledge of the state of the problem and relevance of the used literature.

The author shows deep knowledge of the topic and an expert-level ease in using and explaining the complex notions involved. The literature covered is broad (around 100 sources) but highly relevant.

5. Correctness in citing a representative number of authors.

The citations correspond to the relevance of the sources, with the most relevant sources being the most often cited ones. The citations are correctly accomplished according to the APA guidelines.

6. Availability of a justified and developed theoretical model of the research.

The research tests not one but two established theoretical models: the association-based account (Kruschke, 1996) and the rule-based account (Juslin et al., 2001). The rationale for each account is well elaborated upon in Chapter 3.

7. Relevance of the chosen methodology and research methodology to the set goal and tasks of the dissertation work.

The research, consisting of six experiments, uses both well-established and novel methods in investigating IBRE. All the methods are highly relevant to the research goals, a conclusion that follows naturally from the author's argumentation.

8. Availability of own contribution in the collection and analysis of the empirical data.

The collection and analysis of the data represent an original contribution by the author.

Description of Contributions

1. Brief description of the nature and assessment of the credibility of the material on which the contributions of the dissertation are based.

The data that the research findings are based on is rigorously collected in lab settings or online and, in my opinion, lives up to the highest quality standards.

2. Description of the applicant's contributions and their classification.

I agree with the author that the contributions can be divided into methodological, empirical, and theoretical. Among these, the most important are:

Methodological contributions

- Using various manipulations across the experiments while keeping the materials constant
- Consistent statistical analyses across the experiments

- Introduction of novel methods into the topic, especially the verbal reports in Experiments 1 and 2.

Empirical contributions

- Findings that challenge the association-based approach: IBRE is observed without learning; IBRE is observed among bad learners
- Findings that challenge the rule-based approach: the frequent category is not tested first; higher-level reasoning processes might not be critical for the effect

Theoretical contributions

- The most important empirical findings also have larger-scale theoretical implications, by challenging the relevant theoretical models they were designed to test.

a) The dissertation must demonstrate that the candidate possesses in-depth theoretical knowledge of the relevant specialty and the ability for independent research.

The first four chapters of the dissertation clearly demonstrate the expert knowledge of the author on the researched topic. Chapters 5 to 11 also demonstrate her highly developed experimenter skills.

b) The dissertation must contain theoretical summaries and solutions to scientific or applied scientific problems, which represent an original contribution to science.

Not applicable

3. Evaluation of the degree of personal participation of the dissertation student in the contributions.

As long as this can be evaluated from the dissertation, the author is the main researcher, so she has the largest share in the contributions of this work.

4. Evaluation of the compliance of the autoreferat with the main points and contributions of the dissertation work.

The autoreferat is a relevant summary of the whole dissertation.

Dissertation impacts on the external environment.

1. Evaluation of the publications of the dissertation work: number, nature of the editions in which they were published.

In the references, I detected two publications by the author related to the dissertation. One is a co-authored conference paper published in 2021 in the Proceedings of the International Conference on Recent Advances in Natural Language Processing. The other is a co-authored journal article published in 2019 in *Frontiers in Psychology*. I wish the author a successful publication of more of this exciting work in international journals.

2. Use and citation by other authors, reviews in the scientific press, etc.

According to Google Scholar, the 2021 conference paper has been cited once, and the journal article from 2019 has been cited three times.

Personal qualities of the author (if the reviewer knows him).

I've known Yolina Petrova for several years now. We've met during summer and winter schools of the NBU Cognitive science and psychology department. During our informal meetings and conversations, I've been able to form an impression of Yolina as someone of high intellectual potential and personal motivation, a person of both technical skill and disposition to critical thinking, a person who is at the same time a team player and someone who is able to express and defend her point of view. All these are valuable features in academia and beyond.

Opinions, recommendations and notes.

Although I read the dissertation thoroughly, I am not an expert in the field. Accordingly, take my comments as the best effort of an outsider to make sense of this interesting and specific topic. I divided my comments into four categories: theory, readability/logic, statistics, and technical details. I hope they are helpful.

Theory

1. An important question is how exactly IBRE is defined: is it the "rare" preference in the Combined test case or is it the larger pattern, represented by a decreasing "frequent" preference going from Common to All together to Combined? From the introduction, it seems that the effect is about the pattern. However, on p. 98 the author states that "the IBRE is associated with a preference for assigning specific ambiguous examples to less prevalent outcomes", suggesting that it is about the "rare" preference. I suppose this ambiguity is inherited from the literature but I think the author could be more explicit about it and reiterate more clearly which definition she uses. A clearer definition might

also help the reader understand why results from Experiment 5 are treated as supporting IBRE despite a lack of preference reversal.

2. It is not clear why each of the dominant approaches emphasize either learning or test phase at the expense of the other phase? Why these approaches are more popular than approaches that would consider both phases, and that would be more consistent with the data presented in the dissertation? A similar question arises in respect to the debate between the associative vs. rule-based nature of the effect: why approaches emphasizing on one aspect have dominated approaches considering both?
3. I found it hard to realize how IBRE and the base-rate neglect (Kahneman & Tversky, 1973) stem from the same mechanisms, as argued on pp. 101-102. The base-rate neglect rests on the existence of compelling specific information which clashes with the base rates. There is no evident analogue to this conflict in IBRE.
4. In the light of the individual differences the author found in IBRE, it would be interesting to further investigate if these are related to individual differences in base-rate neglect and in cognitive capacity. One possibility is that those of lower WMC would be more prone to both base-rate neglect and IBRE. Interestingly, previous research in individual differences found that different types of base-rate tasks might be either positively or negatively related to WMC and intelligence (Burgoyne et al., 2021).
5. On p. 100, the author posits that “IBRE relies on some kind of exemplar-based reasoning ... or exemplar-based learning”. How can we reconcile this suggestion with the lit review showing that the exemplar approach fails to account for IBRE?

Readability/Logic

1. I believe that some of the theoretical points raised above could be resolved by elaborating on the competing theoretical frameworks early on in Chapter 3. In the present version of the text, many details seem to come late. For instance, in Chapter 4, discussing the rationale of the studies, it is clarified that IBRE takes place during the learning phase, according to the association-based approach, or during the test phase, according to the rule-based approach. This important distinction might be made earlier, in Chapter 3, which introduces the competing approaches to IBRE. Also, in the interim discussion after Experiment 2 (p. 55), the author states that the rule-based explanation expects an IBRE in both classification and inference learning. This would be better stated before rather than after the results are known.

2. Given the importance of simulation-based approaches, like RoleMap and GPT3, I believe they should be introduced early in the lit review chapters.
3. Some experiments (e.g., Experiment 6) lack interim discussion altogether, so their contribution to the whole picture is not made clear until later in the text.
4. On p. 36, the rationale for choosing novel stimuli in investigating IBRE can go beyond mere generalizability, and state if any theories are directly affected depending on whether IBRE is found or not: which theories would be supported. In addition, the ecological validity of the new materials can be discussed, e.g. does this task make as much intuitive sense to participants as does a task involving disease and symptoms?
5. In Experiment 2, the author concludes that inference learning leads to symmetric representations by focusing on the proportion of definitions using both features. However, she omits to mention that the proportion of reported definitions using a unique feature in this task is still larger for the rare than for the frequent category. In Experiment 1, using a classification task, the ratio of using “unique only” definitions for the rare relative to frequent category was 33.61 to $14.75 = 2.28$. In Experiment 2, using an inference task, the corresponding ratio was 15.18 to $8.93 = 1.73$, smaller but still comparable to that in Experiment 1. Doesn't this suggest that some form of asymmetry can still play a role in this task?
6. In Experiment 1 the proportion of definitions of the rare category that use both features (41%) is still larger than the proportion of definitions that the unique feature only (34%). Doesn't this contradict the author's conclusion on p. 53 that the findings are “in favor of the associative-based view, emphasizing that ... the rare categories are represented mainly by their unique features”?
7. Given the number and complexity of the findings, the reader would benefit from a table summarizing the main findings and assessing whether they support or contradict either of the dominating approaches.

Statistics

1. Related to my theoretical point #1: If we define IBRE as a pattern, should we not use an analysis, where the null hypothesis is an even distribution of the three types of responses, rather than comparing each proportion separately to a 50:50 null hypothesis?
2. Maybe also related to the definition issue, I couldn't get how the difference in magnitude of the IBRE is estimated using a chi-square test (e.g., p. 52).

3. Throughout all the Results sections, I could not get what the CIs stand for. If these are confidence intervals, they are not related to the statistic that precedes them. For instance, on p. 40, “ $t(61) = 6.68, p < .001, d = 0.864$ with 95% CI [0.17, 0.31]” The CI cannot possibly refer to d . Similarly, in “ $\chi^2(1, N=237) = 7.09, p = .008, \phi = .173$ and 95% CI [83, 114] and [124, 154] for the frequent and the rare outcomes respectively” the CIs cannot refer to ϕ . Please specify what these numbers refer to.
4. The author tends to interpret null results as evidence for absence. For instance, on p. 54, “ $\chi^2(2, N=112) = 4.85, p = .089, w = .21$ ” is interpreted as “no association in relation to whether a rare category is reported with its unique feature prioritized or not, and the generalization preference towards the Combined test types including that category”. However, absence of evidence is not evidence of absence. The study might simply not have enough power to find a genuine effect. To establish evidence for absence one might use Bayesian statistic, advocated by John Kruschke (Kruschke & Liddell, 2018), among others.
5. In reporting the sensitivity power analysis, the author should explicitly explain how the numbers obtained (e.g. $N = 237$ in Experiment 1) refers to the number of participants (e.g., 62 in Experiment 1). Reporting the exact estimate of sensitivity would also be better than saying that something “has a sensitivity higher than 0.8”.
6. The chi-square test rests on the assumption that the data points are independent. However, this assumption is violated in the present studies, where each participant provided at least 20 responses. Maybe it is possible to think of analyses that take into account the data structure?

Tech Details

1. Page 16: “The conditional probabilities of the four features are .2, .3, .4, and .6 for the frequent category (and .6, .4, .3, and .2, respectively, for the rare category).” I couldn’t get why the conditional probabilities sum up to 1.5, not to 1.0. Maybe a technical explanation can be added into an appendix?
2. Also on page 16: “3x0.2 probability for the frequent category and 1x0.6 for the rare one”. Isn’t it $\frac{3}{4} * 0.2$ and $\frac{1}{4} * 0.6$?
3. Page 18: “ $p(\text{Unique to F} \cdot \text{Unique to R} | \text{F}) = p(\text{Unique to F} \cdot \text{Unique to R} | \text{R}) = 1$ ” Assuming the point stands for the logical AND operator, I did not understand why any of these conditional probabilities would be equal to 1, that is, to certainty.
4. Page 40 – reference to Appendix A should most probably be reference to Appendix B

References

Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., Hambrick, Z., & Engle, R. W. (2021).

Understanding the Relationship Between Rationality and Intelligence: A Latent-Variable Approach. PsyArXiv. <https://doi.org/10.31234/osf.io/ns9ky>

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective.

Psychonomic Bulletin & Review, 25(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>

Conclusion

My opinion of the dissertation is highly positive and I am excited to recommend the committee to rate the it positively and highly.

Date 11.08.2023

Signature: