

OPINION

From Assoc. Prof. Dr. Nikolay R. Rachev, Sofia University St. Kliment Ohridski, 3.2. Psychology

On a dissertation for obtaining a doctoral degree

In the professional field 3.2. Psychology

Dissertation topic: “Moral judgment and emotions: Studies using biomarkers”

PhD candidate: Veselina Hristova Kadreva

Supervisor: Prof. Dr. Maurice Grinberg

Consultant: Assistant Prof. Dr. Evgeniya Hristova

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

The dissertation examines the mechanisms behind human moral judgments, and in particular the role of emotions and various factors related to the content of the task. This problem is relevant and significant both from a practical and a theoretical point of view. In practical terms, knowledge of such associations would help optimize the conditions under which people make moral judgments. For better or worse, the field appears to be far from that stage yet. A major reason for this is the state of theory in the field: although there is a dominant theory, the dual-process theory of moral judgment, the results supporting this theory usually have alternative explanations, and there are many results that directly contradict the theory. Probably for this reason, the emphasis of the dissertation is on the importance of the topic in a theoretical sense. The dissertation successfully outlines the theoretical issues and controversies, and offers a strong attempt to answer the questions raised.

2. Justification of the objectives and tasks in the dissertation

Despite the many previous studies in the field – which the author examines in detail – there is no unequivocal evidence on the role of emotions in moral judgment. Factors related to the content of moral dilemmas, such as the presence of physical contact, the inevitability of death, and the instrumentality of the caused harm, are also either understudied or poorly operationalized. In response, the research in the dissertation solves a number of methodological problems of previous research. In view of the outlined deficiencies and contradictions in the literature review, the goals and objectives of the dissertation are fully justified.

3. Correspondence between the chosen research methodology and the set goal and tasks of the dissertation

The purpose of the dissertation is to replicate and extend results from previous research in the field of moral judgments. In order to resolve the contradictions between the previous results, the research in the dissertation introduces a number of new methodological aspects compared to previous research in the field. For example, factors that most often varied unsystematically in previous research (e.g., inevitability of death and instrumentality of harm) are systematically manipulated in the dissertation. All moral dilemmas are composed according to a pre-prepared template to minimize the variation of extraneous factors such as length of the text, presence of specific words and phrases, wording of the question, etc. In Experiments 1 and 2, the magnitude of the emotional response is measured directly by recording skin

conductance for the reading and judgment periods. In Experiment 3, alternative explanations for the potential relationship between emotion and moral evaluation were also explored using eye-tracking. The investigation of response times allows a direct test of the dual-process theory's assumptions that deontological responses are intuitive, i.e. fast, and the utilitarian – thoughtful, i.e. slow. In conclusion, the chosen methodology fully corresponds to the dissertation's goals.

4. Scientific and scientific-applied contributions of the dissertation

The work has numerous contributions, the most pronounced being methodological and empirical. The two most notable methodological contributions are:

- Strict experimental control over factors that have varied unsystematically in previous studies;
- Combining multiple research methods: participants' responses and ratings, response times, skin conductance recording and eye movement tracking.

These aspects largely ensure the validity of the data, which is also related to the empirical contributions of the work:

- Robust findings from previous research have been successfully replicated, for example the influence of physical contact on ratings of the acceptability of the protagonist's actions.
- Expectedly, questionable results from previous research have not been replicated, for instance the role of the incidental experience of disgust on moral judgments. Even the authors of these studies don't seem to believe them now (<https://replicationindex.com/category/disgust/>).

These findings also lend credence to the remaining results, where there is no agreement in the available literature:

- Contrary to the dual-process framework, utilitarian responses are not slower than deontological responses
- No causal relationship has been found between the induction of certain emotions and deontological responses to certain dilemmas
- In a between-group design, even the effect of a recognized stable factor such as physical contact cannot be established. The trend is in the expected direction, i.e. perhaps an effect exists but was too small to be captured with the available sample size. This suggests that studies using a within-group design may overestimate the influence of this factor, compared to more realistic situations such as the between-group design.

The last three results pose many questions to the dominant dual-process framework, which is the main theoretical contribution of the dissertation.

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

Two Bulgarian-language publications on the subject by the author are cited in the dissertation.

6. Citations by other authors, reviews in scientific journals, etc.

I have no information about citations from other authors. A publication in English would make the work and its contributions more visible and would give a chance for more citations than publications in Bulgarian.

7. Opinions, recommendations and notes

I divided my comments into four categories: theory, statistics, planning of experiments, and readability/logic. I hope they are helpful.

Theory

The literature usually distinguishes between several types of two-process theories: default-interventionist (Evans & Stanovich, 2013), parallel-competitive (Sloman, 1996) and hybrid (De Neys & Pennycook, 2019). Is it possible to specify which type the two-process framework for moral judgments falls into? Even after such additional specification, these theoretical frameworks remain rather vague and can easily incorporate results that appear to contradict them. This relative non-specificity and unfalsifiability of the dual-process framework could also be directly addressed: the existence of two systems can neither be directly confirmed nor directly rejected. It would also be of interest to the reader if there is an alternative global explanatory model, i.e. what theoretical framework do critics of the dual-process approach offer.

Statistics

In the literature review, it appears that many of the results are obtained through interaction effects, which are extremely difficult to establish robustly (Rohrer & Arslan, 2021; Wagenmakers et al., 2012). A frequently cited problem is that detecting an interaction requires a sample size many times larger than the sample required to detect a main effect (eg, <https://datacolada.org/17>). In this regard, the work could indicate the sample sizes in the previous studies and, based on this, draw conclusions about the reliability of their conclusions.

Related to the above comment, the lit review could also include the effect sizes in previous research. This would make it possible to specify the expected effect sizes in the author's studies and, accordingly, to justify the sample sizes in the experiments. The current version lacks information on whether the sample sizes are related to a statistical power analysis or is determined by practical considerations such as access to participants.

I personally would be happy to see effect sizes (Cohen's *ds*) in the results sections. This would allow for comparisons not only with previous research, but also between experiments in the dissertation (e.g., between Experiment 1, where the dilemmas are in the first person, and Experiment 3, where they are in the third person).

I noted that statistically non-significant results were not reported as numerical values. In my opinion, there is no reason for such discrimination. We also know that absence of evidence is not evidence of absence. Therefore, frequentist analysis could be supplemented by Bayes factors to indicate the relative support of the alternative versus the null hypothesis (Dienes, 2014).

In data for proportion of "acceptable" responses (e.g., Table 2, p. 75), skin conductance (e.g., Table 4, p. 80), and response times (e.g., Table 6, p. 84), the standard deviations are comparable to the averages, and sometimes larger. This observation suggests that these data are probably not normally distributed. In this situation, it is good to check whether the assumptions behind the ANOVAs are met (e.g., normality of

residuals). If the assumptions are not met, a transformation of the data may be considered to put the data closer to normal distribution.

Multiple tests of statistical significance were performed. This is known to increase the probability of false positive results (the so-called familywise error). It is probably possible to use an adjustment to the p-values that reduces this probability. At least in some cases, it is probably also possible to replace multiple tests with a single test. For example, the six ANOVAs on p. 118 (and also on p. 120) could be replaced by a single linear mixed-effects model (Brown, 2021). Similarly, instead of four chi-square tests, on p. 122, one logistic regression can be used.

The skin conductance results are based on a reduced sample, but I couldn't find out the reasons for this.

In many places in the text, statistical significance analysis precedes descriptive statistics, which I think is counter-intuitive.

Planning of experiments

It is worth describing why, of all possible combinations of factors, these particular manipulations were planned, beyond the fact that they are important and that they have been studied before. How exactly do these manipulations answer the research questions?

I am not familiar with previous research, but it seems to me that the within-group design allows for, and possibly predisposes to, strategic responding in this field of research. If I understand the materials correctly, for example, each participant in Experiment 1 saw an almost identical scenario three times. This enables participants to directly compare between scenarios. Therefore, to me personally, the results of Experiment 3 seem theoretically and practically the most significant ones. According to Kahneman, life is between-subjects design.

Readability/Logic

For a reader who is not well familiar with the literature, it would be more convenient if, when introducing the factors (e.g. imminence of death), author also presents concrete examples from tasks.

The same task is classified as a task with inevitable death (Table 1, p. 70) and a task in which death can be avoided (Appendix 1, p. 160). Overall, I didn't quite get what the "inevitability of death" factor meant.

It will be easier for the reader to navigate through the text if there are direct references to the hypotheses in the discussions. For example, in Experiment 1, Hypotheses 1-7 were formulated, but there are no references to them in the discussion.

In the interim discussions after Experiments 1 and 2, it would be helpful to include sections that indicate exactly how the next experiment adds to the picture and whether it answers questions opened by the current experiment.

References

- Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920960351.
<https://doi.org/10.1177/2515245920960351>
- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28(5), 503–509.
<https://doi.org/10.1177/0963721419855658>

- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, a781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Rohrer, J. M., & Arslan, R. C. (2021). Precise answers to vague questions: Issues with interactions. *Advances in Methods and Practices in Psychological Science*, 4(2), 25152459211007370. <https://doi.org/10.1177/25152459211007368>
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1), 3.
- Wagenmakers, E.-J., Krypotos, A.-M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory & Cognition*, 40(2), 145–160. <https://doi.org/10.3758/s13421-011-0158-0>

8. Conclusion with a clearly formulated positive or negative assessment of the dissertation

My impression of the thesis is highly positive. I congratulate the author for her thorough work and recommend the committee to rate the dissertation positively and highly.

Date 07.08.2023

Signature: