Peer Review of Reflection on 'Trial and Error (-Related Negativity)'

Stefan Gaillard¹, Sean Devine²

¹ Reviewer 1

² Reviewer 2

This is the originally submitted version of
Reflection on 'Trial and Error (-Related Negativity)'
with peer review comments attached.

Original author(s):

Maarten Derksen.

Published version:

<https://doi.org/10.36850/r2>.

How to read:

Sections which have peer review comments are marked by ▼ or [links](#),
clicking which directs you to said comments.

The reviews contain links referring back to the original section.

This paper is about error on three levels. First of all, it deals with research into the ERN, error-related negativity. The ERN is a negative deflection in the EEG signal, which tends to occur within 100 milliseconds of making an error. The authors hypothesize that physical pain can be considered as a bodily signal that a type of error has been committed: there is a "discrepancy between the actual and optimal/targeted state", as the authors put it (p.1). This raises the question whether the ERN is also associated with pain and the avoidance of pain, and if so how. More specifically the authors want to know whether people with an elevated ERN are more prone to avoidance behaviour, which in turn can lead to chronic pain.

I am an historian of psychology with philosophical interests and have no expertise in clinical neuropsychology, so I will not comment on this hypothesis. But the paper also deals with error in two other ways, which I do feel able to reflect on. The authors describe their attempts to develop an experimental paradigm for the study of the role of the ERN in pain avoidance. In these attempts they make errors which they then try to correct in a further attempt, six task versions in total. This is the second way this paper deals with errors -- those of the experimenters themselves. But there is a third level too: it is crucial for the experimental task to induce the participant to make the right number of errors -- not too many, not too few. The second and third aspects are obviously related: the errors of the experimenters concern, among others, the number of errors the participants make.

The authors describe their challenge as an interdisciplinary one: they had to combine elements of neurophysiology (ERN) with clinical psychology (pain avoidance). Specifically: they had to somehow induce an ERN in the participants, and elicit and measure some type of avoidance behaviour at the same time. Moreover, to determine what each participant's average ERN is, they needed at least six ERN measures per participant, and thus a minimum of six errors. The errors, finally, had to be "inhibition errors", not errors due to

lack of knowledge or skill. It wasn't clear to me why this was important, or what an inhibition error is in the first place, but this is no doubt due to my own lack of knowledge in this field.

All in all, the specifications of the task were narrow and demanding: not any type of error would do (only inhibition errors); the errors had to produce a proper ERN; a minimum of six was needed; the participants had to be aware of their error (otherwise they would not show avoidance behaviour); and of course there had to be pain associated with the errors, but not so much pain that the ethics committee would reject the pilot study, or the subjects would refuse to participate. [What followed was a kind of dance, or rather a series of dances, with the experimenters leading the participants, successively trying different choreographies in an attempt to get their partners to make the right moves and not step on their toes.](#)

In a recent article, Brenninkmeijer, Rietzschel, and I reported on a series of interviews we had conducted with researchers in psychology (Brenninkmeijer et al., 2019). We had asked them about their informal research practices, that is to say, those practices that are not made explicit in the method section of a paper but are nevertheless considered important. We saw two themes in their answers. The first is a strong concern with professionalism, expressing itself a.o. among others in an orderly lab and a smoothly running experiment, and respectful conduct towards the participants. The second theme is a focus on producing good data by managing the performance of the participant. This second theme is relevant with regard to this paper on 'Trial and Error'.

Much of the work that is done in the lab, including the kind of informal, unwritten work that we explored in our interviews, is geared towards eliciting the right kind of behaviour from the participant. What happens in the lab has a [theatrical](#) quality: a particular performance is expected from the participants, and they are guided to it by the staging and scripting (or choreography) of the experiment and the conduct of the experimenter. Our interviewees mentioned

the importance of the instructions (clear, fool-proof, not too long), of keeping participants focussed and motivated (particularly challenging in cognitive psychology, where tasks are often boring and repetitive), and of creating a certain psychological realism, so that the artificiality of the laboratory situation is less salient and the stimuli and tasks more life-like (this was especially a concern for social psychologists).

What makes this work so difficult is that despite the staging and the scripting, the participant's behaviour must be spontaneous and natural. Or rather: the spontaneity is elicited and facilitated by the staging, the instructions, the props, the stimuli, and the conduct of the experimenter. A psychological experiment aims to create natural behaviour artificially, aims to produce spontaneity (see also Derksen, 2001 about psychological tests). This paradoxical task demands that the experiment is crafted with subtlety and tact. The experimental situation and the experimenter must be forceful, yet unobtrusive. If the staging and the scripting become too prominent the participant risks become recalcitrant, and their behaviour is no longer natural and spontaneous (Derksen, 2017; Lezaun, 2007). In designing an experiment, the management of the participants' awareness is therefore often a major concern: they must be attentive to the stimulus, but not to the stimulus *as* stimulus, as part of an experiment in an artificial laboratory situation, aimed at probing their responses. They must not become reflexive but remain natural and spontaneous.

Interestingly, the authors of this paper designed their new experimental paradigm with the help of their twelve participants. (It would not have been out of place to mention them in the Acknowledgements.) Although they are described in the paper in the usual impersonal, technical terms -- "twelve participants (8 females)", "mean age of $M = 29.25$ ($SD = 10.64$)" (p.6) -- the participants nonetheless had an active role in the pilot, almost as collaborators, specifically by supplying the researchers with their experience of the experiment. "(T)hey were asked whether they were able to perceive when they made errors, how difficult they found the task as well as initiating the avoidance response, and any additional comments they would like to provide." (p.8-9)

As noted in this quote, it was vital that the participants were aware they had made an error, because otherwise they would not 'initiate an avoidance response'. The experimenters had to hit a very narrow target here. On the one hand, participants had to be induced to make a sufficient number of errors, so the task had to be difficult enough that participants did not always know the right response. On the other hand, once they had responded they had to be aware of having made an error, and thus *did* have to know the right response. This transition from not knowing to knowing had to occur within 1300 ms, being the fixation period plus the time allowed for an avoidance response. This required precision engineering of the awareness of the participants.

It quickly became clear that the basic task -- determining where on their lower back they felt a vibration from a 'tactor' -- did not induce enough errors. To make the task more difficult the researchers decided to distract the participants. The tool that they used to modulate the awareness and attention of the participants was an extra tactor, the wonderfully named distractor tactor. The researchers tried various ways to distract the participants, aiming at just the right amount of distraction to induce errors, without distracting the participants so much that they no longer were aware of their errors. Their first attempt was creative, but unsuccessful. I suspect that working out which song is playing

merely from the beat is difficult enough in normal circumstances, but if the beat is transmitted by vibrations on one's lower back it becomes well-nigh impossible. I'm not surprised that the one participant in this task version simply gave up trying, and focused entirely on the main task of locating the main stimuli. Here we have an example of participant recalcitrance, in this case not in response to the fact of being manipulated, but in response to the difficulty of the distractor task. The participant refused to be distracted. Moreover, this participant discovered they could hear the beat as well as feel it: the distractor tactor worked as a little speaker, and the ear plugs that the experimenters had given the participant didn't stop the sound. In other words, the distractor task was at the same time too difficult and too easy: it was too difficult to perform in the intended way (by attending to the vibrations of the distractor tactor), and at the same time it was too easy to not be distracted by it, and too easy to cheat. The experimenters had failed to tactfully usher the participant into the right state of distraction, and instead met resistance. *The participant did not play the game as intended.*

The researchers, therefore, tried a different tack and used the distractor tactor to vibrate simultaneously with the main tactors in order to make locating the stimuli more difficult. This too required careful calibration of the difficulty of the task. In their 'different intensities task' the researchers set the third, distractor tactor at a higher intensity than the two main tactors in an attempt to increase the number of errors. Indeed, the participants did make more errors, but they were less aware of having made them. Yet, although the 50% coactivation task that the researchers tried next reduced the number of errors to an average of 18.5, participants still "reported having difficulty recognizing error commission" (p.13). Ultimately, the 100% coactivation task, which the experimenters had tried first, was the only one that met the requirements of a sufficient number of errors combined with awareness of having made an error after responding.

With regard to the other part of the experimental paradigm, the avoidance response, matters were even more complicated. To get a sufficient number of avoidance responses the experimenters not only had to make sure that participants were aware of having made errors, but were also sufficiently motivated to avoid the punitive electrocutaneous stimulus (e-stim) that would follow an error. While in the error-induction part of the task the attention of the participants had to be modulated in such a way that they made errors but were also aware that they made errors, with regard to the avoidance response a different balance had to be struck. The dynamic of control and resistance that is so typical of psychological experiments is quite prominent here.

First of all, the e-stim had to be calibrated for each individual participant, in such a way that it was "painful and demanding some effort to tolerate" (p.8). The researchers, in consultation with the participants, had to find the golden mean between not enough pain (and therefore no motivation to avoid) and too much pain (which would lead participants to reject the experiment altogether, if it would even pass the Ethics Committee). However, despite the fact that the participants had been allowed to choose their preferred level of punishment, the researchers were nonetheless afraid that participants would subvert the experiment by always performing the avoidance response (pressing the space bar on the keyboard) as a precaution, whether or not they thought they had made an error. Thus, there had to be a cost to avoidance. At the same time,

non-avoidance should be costly too, otherwise the researchers would not get the kind of responses they were interested in.

To direct their participants to this happy medium between too much and too little avoidance the experimenters employed threats. Participants were told that they would be punished for avoiding punishment: every time they pressed the space bar to prevent the e-stim extra trials would be added to the experiment. This ploy failed: participants resisted by simply foregoing avoidance and accepting punishment. The authors imply that participants had made a strategic calculation: if each avoidance response would lead to extra trials the experiment would take longer to finish. Moreover, participants realised that extra trials would also mean more chances to make errors, which would either be punished with the e-stim, or, if avoided, would lengthen the experiment even more. In other words, the threats of the experimenters unwittingly focused the attention of the participants on the boundaries of the experiment, rather than on the task *within* the experiment. The experiment became visible to the participants as a social situation that they were in, with a beginning and an end, and they were eager not to stay in it too long.

To "balance the trade-off" (p.21) and make non-avoidance more costly the experimenters again issued a threat: occasionally a punishing stimulus with "a slightly higher intensity" (p.21) would be delivered after an error. This failed to strike sufficient fear into the participants and did not result in more avoidance responses. The authors are surprised and do not offer an explanation, but I suspect the reason may be similar to why the first threat failed. Participants perhaps reasoned that a 'slightly higher' intensity would still be bearable, bearing in mind that the Ethics Commission sets conservative limits on how much pain a participant may be made to bear. Participants did not attend to the task alone, but were aware of the boundaries of the experiment (in this case the limits set by an Ethics Commission) and took these into consideration in the way they performed the task.

Above I wrote that, in psychology, the experimental situation and the experimenter must be forceful yet unobtrusive. It seems to me that, in as much as the pilot failed to elicit the right number of avoidance responses and at the same time increase their cost, it was due to the situation becoming too prominent relative to the task. The experimenters' threats made the participants attend to the experimental situation as such, as well as to the task within it, and their responses became coloured by their thoughts about the boundaries of the situation they were in. It is not clear whether the researchers asked the participants about their experiences with the threats, as they said they did regarding other aspects of the task. It would have been interesting to learn their reflections and see whether they confirm mine.

Discussing the results of their pilot, the authors argue that the low-cost avoidance responses that they ended up with are nonetheless clinically relevant, since that kind of avoidance behaviour, like always carrying pain medication, occurs in real life too. Given that they also emphasise the "considerable costs" (p.20) of avoidance behaviour in real life this argument seems unconvincing. I'm sure they would have preferred high-cost avoidance in their experiment. Perhaps this is an inherent limitation of studying this phenomenon experimentally. First of all, it is difficult for ethical reasons to make responses in an experiment really costly. Many if not most participants will realise this and understand that the cost they are threatened with is illusory, as indeed it was here.

Secondly, if the cost of avoidance in real life concerns people's "social life, physical functioning and personal well-being", as the authors write (p.20), an experiment may simply be too limited a situation; too small and brief to cause that kind of cost. The authors themselves note that "the operationalization of avoidance behavior as a single button press may be considered simplistic." (p.20) But it is, I think, not just the simplicity of a button press compared to the complexity and variety of real-life avoidance that is the problem, but also the brief duration of an experiment. As became clear in this pilot study, participants know there is an end to the experiment and take it into account in their response strategy. But, if I may be permitted one remark on the topic of study itself, it seems to me that the cost of avoidance behaviour in real life builds up gradually over time. It has a history, and one that is longer than the duration of an experiment. The failure of this pilot study to elicit high-cost avoidance behaviour may well be due to the fact that [an experiment is simply too limited an event to successfully emulate the development of pathological avoidance behaviour in real life](#). As such, however, it is an informative failure.

Peer Reviews

Reviewer 1 Stefan Gaillard

- 0) Perhaps write [this](#) out the first time.
- 1) Very interesting to perhaps expand on [how participation of humans in experiments is constructed](#).
- 2) It would perhaps be interesting in comparing the [game analogy](#) with the choreography analogy and whether they supplement each other, or whether it is more a choreography for the experimenters and more of a game for the participants and how that relates to each other.
- 3) Interesting to expand [upon](#)! Would there be other methodologies which are better suited (perhaps even outside psychology and more inside the scope of for example the humanities)?

Editor comments

- I thoroughly enjoyed reading this reflection article, especially the parts about the choreography of participant studies and the perspective of the participants (i.e. of the study as a game to be 'rigged'). I think it would strengthen the article if you could expand on these aspects.
- In addition, at the end of the article you briefly mention your doubts about the ecological validity of the experiment at hand. I was wondering whether you could expand on this by suggesting (abstractly) what types of research would be better suited? And whether these types of research should be sought in psychology at all or rather in other academic disciplines?
- Finally, I was wondering whether you have anything to say on the tutorial section? Specifically, do you think such a tutorial could help the researchers reading the article from preventing similar mistakes? Can the tutorial be improved? Or do you have any thoughts on the concept of publishing tutorials in general?

Reviewer 2 Sean Devine

- 0) Fascinating [point](#).
- 1) [focused](#)
- 2) Perhaps simply terminological, but does an experiment not aim to capture natural behaviours in a controlled, artificial setting, rather than [create](#) them? Again, this might just be semantic, but it strikes me as a place where much ink may be spilled debating the generalizability of psychological phenomena from the lab to the "real world" (cf. Yarkoni, 2019, The Generalizability Crisis)

Editor comments

- I thank the author for his reflection piece. It was interesting and a pleasure to read. I do, however, have some comments that might improve the reflection article.
- First, I find that too much time is spent summarizing the findings of the original paper. Presumably, readers who are interested in this reflection will have read the original article and thus do not need the point-by-point in such detail. I believe such a trimming-down would leave room for my second recommendation.
- Second, I believe more could be said about the "[theatrical](#)" component of research and how the difficulties of eliciting the proper response(s) in this paper highlight the artificiality of lab experiments. Throughout the reflection article, there are many elucidating moments where the psychology of psychological experiments is hinted at. However, in the manuscript's current state, I do not find that these insights materialize into a more fully formed theory. It would be interesting to hear your developed thoughts on how the "narrowness of the target" the authors are trying to hit might speak to the inherent limitations lab research proposes. In other words, does the necessity of strict experimental control preclude researchers from making ecologically valid claims from their results? If so, do not the challenges of even calibrating such an experiment as in this one speak to this fundamental disconnect? Again, it is my opinion that the answers to these questions are within the text already, but are either not structured or developed enough to leave the reader with a strong impression of what they are.
- Structurally, both the above points may be helped by incorporating subheadings and clearly delineating summary from interpretation.
- Overall, I found this to be an extremely enjoyable read and thoroughly thought-provoking. I would be very excited to see these ideas fleshed out and to hear the author's thoughts on the points raised above.

License

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020