International Journal of Psychophysiology 205 (2024) 112438

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Theories and hypotheses: The forgotten plane of the multiverse

Michael Richter [a,b,*], Guido H.E. Gendolla [c,d]

[a] Effort Lab, School of Psychology, Faculty of Health, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
[b] Research Centre for Brain and Behaviour, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK
[c] Geneva Motivation Lab, FPSE, Section of Psychology, University of Geneva, 40, Bd du Pont-d'Arve, 1205 Geneva, Switzerland
[d] Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

ARTICLE INFO

Keywords:
Multiverse analysis
Research methods
Theory
Comparative analysis

ABSTRACT

Multiverse analyses—the systematic examination of the effects of decisions that researchers can take over the course of a research project—became more common in recent psychophysiological research. However, multiverse analyses in psychophysiology almost exclusively focus on methodological and statistical decisions that can have a considerable impact on the findings. The role of the conceptual multiverse regarding theory-related research decisions is largely ignored. We argue that the choice of a theory that guides hypotheses, study design, measurement methods, and statistical analyses is the first plane of the psychophysiological multiverse. Depending on the chosen theoretical framework, researchers will choose different methods, and statistical analyses will emphasize specific aspects. We illustrate this process with a research example studying the effects of task difficulty manipulations on cardiovascular effects reflecting effort. We argue in favor of an approach that explicitly acknowledges the various theoretical accounts that can constitute the background of a study and demonstrate how a comparative analytical approach can provide a comprehensive multiverse without increasing type I error due to mere exploration.

1. Theories and hypotheses: The forgotten plane of the multiverse

Multiverse analysis refers to reporting and comparing the effects of all reasonable decisions that researchers can take over the course of a research project. Like in other scientific disciplines (e.g., El Bahri et al., 2022; Engzell and Mood, 2023; McBee et al., 2021), multiverse analysis (Clayson, 2024; Steegen et al., 2016) and the related specification curve (Simonsohn et al., 2020) and vibration of effects analyses (Patel et al., 2015) have become more frequent in psychophysiology over recent years (e.g., Lewis et al., 2023; Sadus et al., 2024; Sjouwerman et al., 2022). A multiverse approach acknowledges that researchers have to make a multitude of decisions when conducting a research project and that this introduces variability in the research outcome because researchers vary in the specific decisions that they make (Breznau et al., 2022). Every decision that a researcher can take leads to a new universe in which the final research outcome potentially differs from the research outcome in another universe. It is of note that pre-registrations and registered reports—valuable tools to encourage researchers to explicitly state their decisions *before* beginning a research project—cannot prevent the variability that multiverse analysis addresses: Even if researchers pre-register their projects, there will be variability in the decisions that different researchers take and decisions that one and the same researcher takes in the context of different projects.

1.1. The methodological multiverse

In psychophysiology, researchers have amongst others demonstrated the variability of research results as a function of data-related decisions on outlier treatment (Bloom et al., 2022), data exclusion criteria (e.g., Bloom et al., 2022; Lewis et al., 2023), EEG reference schemes (Klawohn et al., 2020), applied filter frequencies (Sadus et al., 2024), skin conductance response quantification (Sjouwerman et al., 2022), model covariates (Bloom et al., 2022), and study sample composition (Lewis et al., 2023)—to name a few. Depending on the analyzed multiverses and outcome variables, some of these multiverse analyses provided evidence for a considerable influence of researchers' decisions on the reported findings. For instance, Lewis et al. (2023) observed that the difference in extinction retention in a fear conditioning paradigm between individuals with posttraumatic stress disorder (PTSD) and trauma-exposed individuals who did not suffer from PTSD varied considerably as a function of extinction retention index, sample sex,

* Corresponding author at: Effort Lab, School of Psychology, Faculty of Health, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK.
E-mail addresses: michael.richter@ljmu.ac.uk (M. Richter), guido.gendolla@unige.ch (G.H.E. Gendolla).

https://doi.org/10.1016/j.ijpsycho.2024.112438
Received 25 April 2024; Received in revised form 15 August 2024; Accepted 7 September 2024
Available online 10 September 2024

exclusion criteria, and outlier trial removal—they reported a mean Cohen's *d* of −0.29 with a standard deviation of 0.36.

Most of these study examples—and most multiverse analyses in general—focused on the impact of differences in the processing and analyzing of data in individual, original empirical studies. However, multiverse analysis is not limited to this. For instance, Voracek et al. (2019) applied the multiverse idea to meta-analysis demonstrating how the various decisions that can be made in the context of the analysis of secondary data influence the final result. Harder's (2020) work constitutes another example that the multiverse approach is not limited to data processing and analysis. They showed that decisions at the initial data collection stage can create a multiverse, as well. Amongst other effects, they demonstrated that the time that researchers allow participants to respond to the target in the first-person shooter task developed by Correll et al. (2002)—a measure of racial bias—strongly influences the likelihood of finding a bias in shooting errors and response times. Longer response windows around 850 ms make it more likely to find a bias on response times whereas shorter windows around 650 increase the likelihood of finding a bias on response error. The purpose of the present paper is to extend the multiverse perspective even further by illustrating that decisions about the theoretical framework and scientific hypotheses—i.e., conceptual decisions that have to be made *before* the decisions about data collection and data analysis are taken—can also lead to considerable variability in the research outcome and thus constitute another plane of the multiverse. Surprisingly, this conceptual plane of the multiverse has been largely neglected to date.

1.2. The conceptual multiverse

When planning their studies, researchers are guided by theories and hypotheses. These theories and hypotheses can be *explicit* and clear like in the context of our own research in the context of Brehm and Self's (1989) motivational intensity theory (e.g., Bouzidi and Gendolla, 2024; Brinkmann et al., 2021; Falk et al., 2024; Framorando et al., 2023; Gendolla et al., 2019; Richter et al., 2016) or they can be *implicit*, latent beliefs about variables and how these variables should be related to one another (Fried, 2021). However, even the most exploratory research is inevitably guided by an implicit theory or hypothesis that tells the researcher which topic to explore and what to manipulate and measure. Independent of whether theories and hypotheses are explicit or implicit, the theory that researchers adopt and the hypotheses that they decide to examine determine the study that they conduct and thus the results that they can obtain. The sample that is recruited, the factors that are manipulated, and the variables that are assessed are amongst the many aspects that are determined by the chosen theory and hypotheses.

Given that there is frequently more than a single theory that can be adopted to study a phenomenon and more than a single hypothesis that can be formulated and tested, decisions about theories and hypotheses are part of the multitude of decisions that researchers have to make when conducting a research project. As any other research-related decision, considered and selected theories and hypotheses can introduce variability in the research outcome given that researchers differ regarding the decisions that they take. Decisions about theories and hypotheses thus create an additional—conceptual—plane of the multiverse. If primary data are collected in the context of a project, the conceptual and methodological planes are tightly connected, as discussed in the preceding paragraphs. In contrast, if secondary data are collected, both planes can become disconnected given that researchers may decide to analyze theories and hypotheses that are different from the ones that drove the original data collection. It is of note that variability in the hypotheses that researchers examine is not necessarily the consequence of the adoption of different theories. It can also be the result of researchers selecting different hypotheses from one and the same theory (Loken and Gelman, 2014). According to our view, psychophysiologists have largely ignored the conceptual theory-and-hypotheses plane of the multiverse, so far. This is probably due to psychophysiologists traditionally working with only a single theory and testing only a single hypothesis for a particular outcome variable. Psychophysiologists thus frequently miss out on many universes that are created by decisions about theories and hypotheses.

It may seem that a conceptual plane multiverse analysis is qualitatively different from the traditional multiverse analysis on the methodological plane. This is because comparing the effects of the selection of different theories and hypotheses may suggest that different questions are analyzed whereas the traditional multiverse analysis focusses on the effects of different methodological decisions in the context of a single question. However, this is not the case. On both the conceptual and methodological planes, multiverse analyses compare the effects of researchers' decisions on outcome variability in the context of a single question or research aim. In methodological multiverse analyses, the questions are more specific, frequently corresponding to a single hypothesis. For instance, the multiverse analysis of the impact of different analytic methods on the strength of fear conditioning effects by Lewis et al. (2023) described in a preceding paragraph was guided by the hypothesis that individuals suffering from PTSD and individuals not suffering from it differ regarding the effectiveness of fear conditioning. In a multiverse analysis on the conceptual plane, the questions and aims are more general, but there is still a single question or aim guiding the analysis. For instance, the two analyses that we describe in Section 1.4 are driven by the general question about the specific relationship between task difficulty and effort mobilization.

Examining the conceptual multiverse created by theory choice can be difficult given that different theories may lead to completely different studies. In contrast to a multiverse analysis of data analysis-related decisions where research outcomes of alternative universes can easily be created by re-analyzing one original data set using different processing and analytical strategies, it is obviously impossible to examine the research outcome of a study that has not been conducted. However, if two or more theories can be applied to generate hypotheses for one and the same study, it is possible to examine the effect of theory choice on research outcomes by testing the hypotheses generated by the different theories and comparing the results. The same strategy can be adopted to examine the variability created by the choice of one of multiple hypotheses offered by one and the same theory.

1.3. An example: The conceptual plane of the Richter et al. (2008) experiment

We will now illustrate the effect that the adoption of different theories and hypotheses has on research outcomes using a psychophysiological experiment that we published a while ago (Richter et al., 2008). In that study, participants performed a short-term memory task randomly allocated to one of four task difficulty conditions—labelled "low difficulty", "moderate difficulty", "high difficulty", and "impossible difficulty". Task-related reactivity of participants' cardiac pre-ejection period[1] (PEP) constituted the main outcome variable to assess the intensity of mental effort during task performance. The first source of variability in this study results from the various hypotheses that our guiding theoretical framework—motivational intensity theory (Brehm and Self, 1989)—offers for a manipulation of task difficulty across four levels. Motivational intensity theory suggests that effort—which we operationalized as PEP reactivity (Kelsey, 2012; Wright, 1996)—is a direct function of task difficulty if task success is possible and if the

[1] Pre-ejection period (PEP) is a measure of the heart's contractile force and assessed as the time interval (in ms) from the onset of the electrical depolarization of the left heart ventricle to the beginning of the ejection of blood from the ventricle into the aorta. PEP reflects beta-adrenergic sympathetic nervous system impact on the heart and becomes shorter with increasing contractile force. Pre-ejection period reactivity refers to the change in pre-ejection period from a baseline period to a task performance period.

required effort is justified by the importance of successfully completing a task. If success is not possible—for instance, because the necessary effort exceeds a person's ability—or if the required effort is not justified, individuals should disengage from the task and no effort should be invested. The specific hypothesis that the theory predicts for a study involving a manipulation of task difficulty across four levels depends on two decisions that a researcher has to make: A decision about the difficulty level where disengagement is expected because of the task being too difficult—either because the required effort is not justified by success importance or because task success is impossible—and a decision about the relative difference between the predicted effort in the low-difficulty condition and the conditions where disengagement is expected.

The first decision to be taken determines the range of difficulty levels where an increase in effort and thus stronger PEP reactivity is expected with increasing task difficulty and leads to five different hypotheses: (1) A researcher assuming that success importance is high enough to justify the required effort in all four conditions and additionally assuming that success is also possible in the condition that we labelled "impossible" will predict an increases in PEP reactivity with increasing task difficulty across all four levels. (2) A researcher assuming that success was indeed impossible in the impossible-condition or that success importance was not high enough to justify the extreme effort required in the impossible-condition, will predict an increase in PEP reactivity with increasing task difficulty across the first three levels and disengagement—that is, no or very weak reactivity—in the impossible-condition. (3) A researcher assuming that the required effort is only justified in the two easiest conditions, will expect an increase in PEP reactivity across these two conditions and disengagement in the other two conditions. (4) A researcher assuming that the required effort is only justified in the easiest task condition, will expect increased PEP reactivity only in this condition but not in the other three conditions. (5) A researcher assuming that success importance is so low that it does not justify the required effort in any of the four difficulty conditions will predict disengagement and no PEP reactivity in all four conditions. The last hypothesis is probably not very meaningful given that it makes not much sense to conduct a study where the single manipulated variable is expected to have no effect. We will therefore not consider this latter hypothesis in the following discussion.

The second decision to be taken is about the difference between the effort predicted in the low-difficulty condition and the condition(s) where disengagement is expected. From a purely theoretical perspective, one would expect more effort and stronger PEP reactivity in the low-difficulty condition—if one assumes that the required effort is justified by success importance in this condition—than in the disengagement conditions. Even if only very little effort is required to successfully perform the easy task, this would still be more than no effort, which is expected if someone disengages. However, researchers may be reluctant to make such a prediction given that the difference in effort between the low-difficulty task and disengagement may be so small that it may be difficult to empirically observe it. Researchers thus have two choices: predicting higher effort in the low-difficulty condition than in the disengagement conditions or predicting equal effort in these conditions.

Summing up, the two decisions that researchers have to make to use motivational intensity theory (Brehm and Self, 1989) to predict effort in a study with a manipulation of task difficulty across four levels like in our Richter et al. (2008) experiment lead to the six meaningful hypotheses displayed in Fig. 1. However, following the standard approach in psychophysiology, researchers would normally only test the hypothesis that they derived from motivational intensity theory as a function of their own specific decisions. We followed this standard approach in our article and tested the single hypothesis that we had formulated—Hypothesis 2 in Fig. 1—which was significant with $p < .001$. It is, however, not difficult to complement this single test with a multiverse analysis. Given that the study had a between-persons design, the reported summary statistics are sufficient to conduct tests for the other five meaningful hypotheses. Using the reported information about means, standard deviations, and number of participants in each condition, one can create a surrogate data set (Larson, 1992)—an artificial data set that leads to the same single-factor analysis of variance summary statistics than the original data set—and then test the other hypotheses with the same planned a priori contrast method (Rosenthal and Rosnow, 1985) that we used to conduct the test presented in the paper.[2] The contrast weights that can be used to test the hypotheses as well as the *p* values of the associated planned contrasts for all six hypotheses are indicated in Table 1. The R script that was used to conduct the analyses described in the paper can be found at https://osf.io/7teuv/?view_only=857d4287b84c411eb1e4cc51f2d91ef5.

As Table 1 reveals, the statistical tests were significant for some hypotheses, but not for others. This demonstrates the variability of the research outcome as a function of the expected effects and consequently as an effect of the two decisions that researchers are forced to make when applying motivational intensity theory to a study like the Richter et al. (2008) experiment. Assuming that task difficulty in the impossible-difficulty condition was indeed impossible and that effort in this—disengagement—condition would not be different from the effort invested in the low-difficulty condition, we predicted Hypothesis 2 in the paper and found a significant results. A researcher applying motivational intensity theory's disengagement prediction more strictly by expecting higher effort in the low-difficulty condition than in the impossible-difficulty condition would have formulated Hypothesis 5 and also obtained a significant planned contrast result. However, researchers making any of the other meaningful predictions would not have found a significant effect.

In addition to the variability in research outcome that is introduced by examining different hypotheses resulting from one and the same theory (Loken and Gelman, 2014), using different theories can also lead to different hypotheses and corresponding variability in the outcome. For instance, some authors suggested that the relationship between task difficulty and effort is best described by an inverted-U shape (Fairclough and Ewing, 2017; Mallat et al., 2020; van der Wel and van Steenbergen, 2018)—and not the sawtooth pattern suggested by motivational intensity theory (Brehm and Self, 1989). This theory would lead to the prediction of high effort in the moderate-difficulty and high-difficulty conditions and low effort in the low-difficulty and impossible-difficulty conditions. Hypothesis 7 in Fig. 1 displays this prediction. Another example of a theory that leads to different predictions for the Richter et al. (2008) experiment is the postulate that task difficulty and effort-related sympathetic activity are related in an exponential, quadratic manner (Slade et al., 2021)—and not in the linear manner assumed by many researchers who relied on Wright's (1996) psychophysiological adaption of motivational intensity theory. This alternative postulate assumes that the differences in the amount of sympathetic activity (and thus PEP reactivity) induced by adjacent difficulty levels are not stable but increase with increasing task difficulty. That is, the difference in effort-related sympathetic activity between the easy-difficulty condition and the moderate-difficulty condition should be smaller than the difference between the moderate-difficulty condition and the high-difficulty condition. This hypothesis is presented as Hypothesis 8 in Fig. 1.

Testing the hypotheses derived from these two alternative theories—the associated contrast weights are indicated in Table 1—leads to

[2] Please see Richter, M. (2016). Residual tests in the analysis of planned contrasts: Problems and solutions. *Psychol Methods*, *21*(1), 112–120. https://doi.org/10.1037/met0000044, for an introduction on how to translate theoretical hypotheses into contrast weights and how to conduct the associated planned a priori contrast. Please note that a priori contrasts take into account all cells/conditions of a design and thus differ from tests of pairwise comparisons that are sometimes also called planned contrasts.

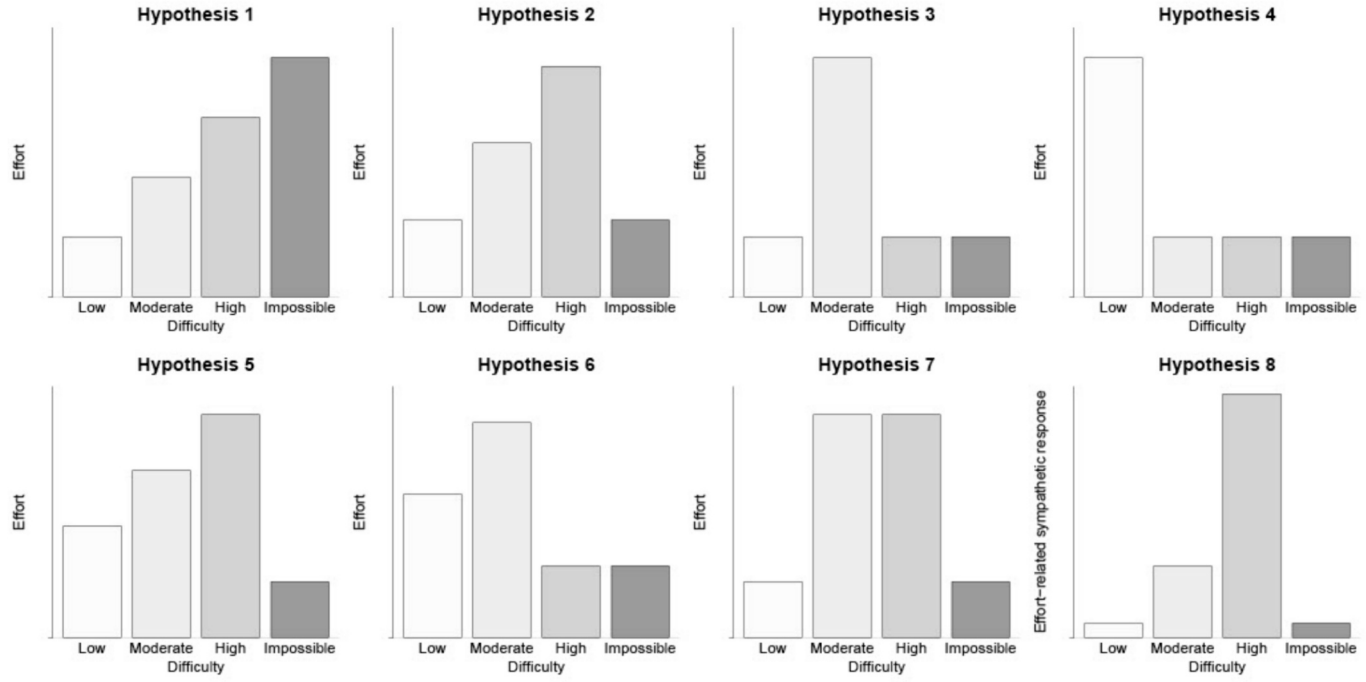

Fig. 1. Meaningful Hypothesis for the Richter et al. (2008) Study.
Notes. Hypotheses 1–6 are derived from motivational intensity theory (Brehm and Self, 1989). Hypotheses 2–4 reflect predictions that are based on the decision to predict equal effort in the low-difficulty and disengagement condition(s). Hypotheses 5 and 6 are based on the decision to predict less effort in the disengagement condition(s) than in the low-difficulty condition. Hypothesis 7 is based on the prediction of an inverted U-relationship between task difficulty and effort. Hypothesis 8 reflects the assumption of an exponential relationship between increases in task demand and effort-related sympathetic activity if the required effort is justified and success is possible.

Table 1
Contrast weights and p values of the planned contrasts used to test the hypotheses presented in Fig. 1.

Hypothesis	Contrast weights				Contrast *p* value
	Low	Medium	High	Impossible	
1	–3	–1	+1	+3	.48
2	–3	+1	+5	–3	<.001
3	–1	+3	–1	–1	.35
4	+3	–1	–1	–1	.01
5	–1	+1	+3	–3	<.01
6	+1	+5	–3	–3	.19
7	–1	+1	+1	–1	<.001
8	–5	–1	+11	–5	<.001

significant results in both cases. A researcher working with one of the two alternative theories would thus also have found supporting evidence in the Richter et al. (2008) experiment.

The objectives of the two types of conceptual multiverse analyses described in the preceding paragraphs slightly differ. The analysis of variability due to the adoption of different hypotheses derived from one and the same theory mainly examines variability that is caused by differences in how researchers interpret the study manipulations—for instance, whether task success is considered to be impossible in the highest difficulty condition. The analysis of variability due to the adoption of different theories has a slightly different objective. It aims to illustrate how well different theories explain the data and to examine how large the outcome variability caused by the adoption of different theories is.

1.4. Results aggregation using bayes factors

Many multiverse analyses only exploratively describe the variability in the research outcome as a function of the potential decisions that researchers can make, frequently not controlling for an increase in type I error. There are notable exceptions like specification curve analysis (Simonsohn et al., 2020), but most analyses do not offer any integration of the results. In the context of a multiverse analysis of the effects of decisions on theories and hypotheses, an integration can be provided using likelihood ratios or Bayes Factors to directly compare the relative fit of the various hypotheses with the data. Likelihood ratios and Bayes Factors compare the likelihood of the data under one hypothesis with the likelihood of the data under an alternative hypothesis and provide thus a relative measure of how well the data fit the individual hypotheses. Bayes Factors range from near zero to infinity and indicate how much support there is for one hypothesis—or model—over the other. A Bayes Factor of 1 indicates that both hypotheses are equally strongly supported by the data, whereas smaller and larger Bayes Factors favor one of the two hypotheses. Even if Bayes Factors constitute a continuous measure of support, some researchers provided categorical labels to help interpretation. For instance, Lee and Wagenmakers (2014) suggested that Bayes Factors ranging from 1 to 3 and from 1/3 to 1 should be considered anecdotal evidence, from 3 to 10 and from 1/10 to 1/3 moderate evidence, and from 10 to 30 and from 1/30 to 1/10 strong evidence. It is straightforward to calculate the required likelihood ratios or Bayes Factors using the information that is available from the planned contrast analyses (e.g., Glover and Dixon, 2004; Masson, 2011; Richter, 2016)—mainly the sum of squares associated with the various sources of variance. Table 2 presents the Bayes Factors for the comparison of the eight hypotheses discussed in the preceding paragraphs.[3]

Table 2 illustrates that some hypotheses—Hypotheses 2, 5, and 8—consistently outperform the other hypotheses. This additional aggregation of the data beyond the presentation of whether the planned

[3] Given that likelihood ratios provide in this context the same information as the Bayes Factors, we refrained from presenting likelihood ratios.

Table 2
Bayes Factors for the comparison of the hypotheses presented in Fig. 1.

Hypothesis	1	2	3	4	5	6	7	8
1	-	4.16×10^{-32}	0.92	0.03	4.48×10^{-30}	0.66	3.27×10^{-15}	1.70×10^{-31}
2	2.40×10^{31}	-	2.22×10^{31}	7.53×10^{28}	107.68	1.58×10^{31}	7.85×10^{16}	4.09
3	1.08	4.51×10^{-32}	-	0.03	4.85×10^{-30}	0.71	3.54×10^{-15}	1.85×10^{-31}
4	31.92	1.33×10^{-30}	29.47	-	1.43×10^{-28}	20.93	1.04×10^{-13}	5.44×10^{-30}
5	2.23×10^{29}	0.01	2.06×10^{29}	7.00×10^{27}	-	1.46×10^{29}	7.29×10^{14}	0.04
6	1.53	6.35×10^{-32}	1.41	0.05	6.83×10^{-30}	-	4.98×10^{-15}	2.60×10^{-31}
7	3.06×10^{14}	1.27×10^{-17}	2.83×10^{14}	9.59×10^{12}	1.37×10^{-15}	2.01×10^{14}	-	5.22×10^{-17}
8	5.87×10^{30}	0.24	5.42×10^{30}	1.84×10^{29}	26.30	3.85×10^{30}	1.92×10^{16}	-

Notes. Cells with strong relative evidence in favor of the hypothesis listed in the first column are highlighted in light grey. Cells with strong relative evidence against the hypotheses listed in the first column are highlighted in dark grey.

contrast tests were significant for the individual hypotheses shifts the multiverse analysis from exploratory to explanatory. Instead of merely illustrating that the research outcome varies as a function of decisions made when selecting theories and hypotheses, it allows researchers to illustrate which hypotheses are more supported by the data than others and thus prepares future focused work that advances the field. For instance, the suggested multiverse analysis can help to select the specific hypotheses that are than tested against one another in decisive studies enabling a strong inference (Platt, 1964) or multiple working hypotheses (Chamberlin, 1897) approach. It is noteworthy that this benefit of the suggested aggregation constitutes a qualitative difference to most other multiverse analysis approaches: Instead of merely constating the amount of variability that researcher decisions cause, aggregating the data with Bayes Factors—or likelihood ratios—provides information about which decisions are more supported by the data.

1.5. Recommendations

We encourage researchers interested in a multiverse analysis to not only consider the impact that methods and data analysis-related decisions have on research outcome, but also to consider the impact that decisions about guiding theoretical frameworks and hypotheses have. These decisions are the first that are normally made when designing a study and can have a considerable impact on the outcome—as illustrated in the discussed example. Exclusively focussing multiverse analyses on methodological and analytical decisions would miss out on an important source of variation. Integrating the conceptual plane of decisions about theories and hypotheses in multiverse analysis will provide a more comprehensive approach to studying the sources of research outcome variability. However, the suggested analysis is only possible if sufficient information for a re-analysis of the data is provided. Cell means, standard deviations, and number of participants per condition are sufficient in the case of between-person designs as illustrated in the preceding sections, but the creation of a surrogate dataset is not possible with these pieces of information in the case of repeated-measures designs, mixed-model designs, or correlational designs. We thus strongly recommend that researchers make their full data sets available so that other researchers have access to all information required for a multiverse analysis of the impact of decisions about theories and hypotheses on research outcome. We also recommend that researchers do not conduct multiverse analysis with the aim to merely describe the amount of variability but with the aim to provide some kind of integration that goes beyond the descriptive level. In the case of variability due to theories and hypotheses, the suggested Bayes Factor-based analysis provides such an integration.

CRediT authorship contribution statement

Michael Richter: Writing – original draft, Conceptualization. **Guido H.E. Gendolla:** Writing – review & editing.

Data availability

No data was used for the research described in the article.

References

Bloom, P.A., VanTieghem, M., Gabard-Durnam, L., Gee, D.G., Flannery, J., Caldera, C., Goff, B., Telzer, E.H., Humphreys, K.L., Fareri, D.S., Shapiro, M., Algharazi, S., Bolger, N., Aly, M., Tottenham, N., 2022. Age-related change in task-evoked amygdala-prefrontal circuitry: a multiverse approach with an accelerated longitudinal cohort aged 4-22 years. Hum. Brain Mapp. 43 (10), 3221–3244. https://doi.org/10.1002/hbm.25847.

Bouzidi, Y.S., Gendolla, G.H.E., 2024. Cognitive conflict does not always mean high effort: task difficulty's moderating effect on cardiac response. Psychophysiology. https://doi.org/10.1111/psyp.14580.

Brehm, J.W., Self, E.A., 1989. The intensity of motivation. Annu. Rev. Psychol. 40 (1), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545.

Breznau, N., Rinke, E.M., Wuttke, A., Nguyen, H.H.V., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H.K., Auer, D., Azevedo, F., Bahnsen, O., Balzer, D., Bauer, G., Bauer, P.C., Baumann, M., Baute, S., Benoit, V., Bernauer, J., Berning, C., Zoltak, T., 2022. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. Proc. Natl. Acad. Sci. USA 119 (44). https://doi.org/10.1073/pnas.2203150119 e2203 150119.

Brinkmann, K., Richter, M., Gendolla, G.H.E., 2021. The intensity side of volition. Z. Sportpsychol. 28 (3), 97–108. https://doi.org/10.1026/1612-5010/a000323.

Chamberlin, T.C., 1897. Studies for students: the method of multiple working hypotheses. J. Geol. 5 (8), 837–848. https://doi.org/10.1086/607980.

Clayson, P.E., 2024. Beyond single paradigms, pipelines, and outcomes: embracing multiverse analyses in psychophysiology. Int. J. Psychophysiol. 197, 112311. https://doi.org/10.1016/j.ijpsycho. 2024.112311.

Correll, J., Park, B., Judd, C.M., Wittenbrink, B., 2002. The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. J. Pers. Soc. Psychol. 83, 1314–1329. https://doi.org/10.1037//0022-3514.83.6.1314.

El Bahri, M., Wang, X., Biaggi, T., Falissard, B., Naudet, F., Barry, C., 2022. A multiverse analysis of meta-analyses assessing acupuncture efficacy for smoking cessation evidenced vibration of effects. J. Clin. Epidemiol. 152, 140–150. https://doi.org/10.1016/j.jclinepi. 2022.09.001.

Engzell, P., Mood, C., 2023. Understanding patterns and trends in income mobility through multiverse analysis. Am. Sociol. Rev. 88 (4), 600–626. https://doi.org/10.1177/00031 224231180607.

Fairclough, S.H., Ewing, K., 2017. The effect of task demand and incentive on neurophysio- logical and cardiovascular markers of effort. Int. J. Psychophysiol. 119, 58–66. https://doi.org/10.1016/j.ijpsycho.2017.01.007.

Falk, J.R., Gollwitzer, P.M., Oettingen, G., Gendolla, G.H.E., 2024. Noise annoys-but personal choice can attenuate noise effects on cardiac response reflecting effort. Psychophysiology 61, e14502. https://doi.org/10.1111/PSYP.14502.

Framorando, D., Falk, J.R., Gollwitzer, P.M., Oettingen, G., Gendolla, G.H.E., 2023. Can personal task choice shield against fear and anger prime effects on effort? A study on cardiac response. Biol. Psychol. 181, 108616. https://doi.org/10.1016/j.biopsycho.2023.108616.

Fried, E.I., 2021. Lack of theory building and testing impedes progress in the factor and network literature. Psychol. Inq. 31 (4), 271–288. https://doi.org/10.1080/1047840X. 2020.1853461.

Gendolla, G.H.E., Wright, R.A., Richter, M., 2019. Advancing issues in motivation intensity research : Updated insights from the cardiovascular system. In: Ryan, R.M. (Ed.), The Oxford Handbook of Human Motivation, 2nd ed. Oxford University Press, pp. 373–392.

Glover, S., Dixon, P., 2004. Likelihood ratios: a simple and flexible statistic for empirical psychologists. Psychon. Bull. Rev. 11 (5), 791–806. https://doi.org/10.3758/bf03196706.

Harder, J.A., 2020. The multiverse of methods: extending the multiverse analysis to address data-collection decisions. Perspect. Psychol. Sci. 15 (5), 1158–1177. https://doi.org/10.1177/1745691620917678.

Kelsey, R.M., 2012. Beta-adrenergic cardiovascular reactivity and adaptation to stress: The cardiac pre-ejection period as an index of effort. In: Wright, R.A., Gendolla, G.H. E. (Eds.), How Motivation Affects Cardiovascular Response: Mechanisms and Applications. American Psychological Association, pp. 43–60.

Klawohn, J., Meyer, A., Weinberg, A., Hajcak, G., 2020. Methodological choices in event-related potential (ERP) research and their impact on internal consistency reliability and individual differences: an examination of the error-related negativity (ERN) and anxiety. J. Abnorm. Psychol. 129 (1), 29–37. https://doi.org/10.1037/abn0000458.

Larson, D.A., 1992. Analysis of variance with just summary statistics as input. Am. Stat. 46 (2). https://doi.org/10.2307/2684186.

Lee, M.D., Wagenmakers, E.-J., 2014. Bayesian Cognitive Modelling. Cambridge University Press, A practical course. https://doi.org/10.1017/CBO9781139087759.

Lewis, M.W., Bradford, D.E., Pace-Schott, E.F., Rauch, S.L., Rosso, I.M., 2023. Multiverse analyses of fear acquisition and extinction retention in posttraumatic stress disorder. Psychophysiology 60 (7), e14265. https://doi.org/10.1111/psyp.14265.

Loken, E., Gelman, A., 2014. The statistical crisis in science. Am. Sci. 102 (6). https://doi.org/10.1511/2014.111.460.

Mallat, C., Cegarra, J., Calmettes, C., Capa, R.L., 2020. A curvilinear effect of mental workload on mental effort and behavioral adaptability: an approach with the pre-ejection period. Hum. Factors 62 (6), 928–939. https://doi.org/10.1177/0018720819855919.

Masson, M.E., 2011. A tutorial on a practical Bayesian alternative to null-hypothesis signifi- cance testing. Behav. Res. Methods 43 (3), 679–690. https://doi.org/10.3758/s13428-010-0049-5.

McBee, M.T., Brand, R.J., Dixon Jr., W.E., 2021. Challenging the link between early childhood television exposure and later attention problems: a multiverse approach. Psychol. Sci. 32 (4), 496–518. https://doi.org/10.1177/0956797620971650.

Patel, C.J., Burford, B., Ioannidis, J.P., 2015. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J. Clin. Epidemiol. 68 (9), 1046–1058. https://doi.org/10.1016/j.jclinepi.2015.05.029.

Platt, J.R., 1964. Strong inference: certain systematic methods of scientific thinking may produce much more rapid progress than others. Science 146 (3642), 347–353. https://doi.org/10.1126/science.146.3642.347.

Richter, M., 2016. Residual tests in the analysis of planned contrasts: problems and solutions. Psychol. Methods 21 (1), 112–120.

Richter, M., Friedrich, A., Gendolla, G.H.E., 2008. Task difficulty effects on cardiac activity. Psychophysiology 45 (5), 869–875. https://doi.org/10.1111/j.1469-8986.2008.00688.x.

Richter, M., Gendolla, G.H.E., Wright, R.A., 2016. Three decades of research on motivational intensity theory: what we have learned about effort and what we still don't know. Adv. Motiv. Sci. 3, 149–186. https://doi.org/10.1016/bs.adms.2016.02.001.

Rosenthal, R., Rosnow, R.L., 1985. Contrast Analysis. Focused Comparisons in the Analysis of Variance. Cambridge University Press.

Sadus, K., Schubert, A.L., Loffler, C., Hagemann, D., 2024. An explorative multiverse study for extracting differences in P3 latencies between young and old adults. Psychophysio- logy 61 (2), e14459. https://doi.org/10.1111/psyp.14459.

Simonsohn, U., Simmons, J.P., Nelson, L.D., 2020. Specification curve analysis. *Nat.* Hum Behav 4 (11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z.

Sjouwerman, R., Illius, S., Kuhn, M., Lonsdorf, T.B., 2022. A data multiverse analysis investigating non-model based SCR quantification approaches. Psychophysiology 59 (12), e14130. https://doi.org/10.1111/psyp.14130.

Slade, K., Kramer, S.E., Fairclough, S., Richter, M., 2021. Effortful listening: sympathetic activity varies as a function of listening demand but parasympathetic activity does not. Hear. Res. 410, 108348. https://doi.org/10.1016/j.heares.2021.108348.

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing transparency through a multiverse analysis. Perspect. Psychol. Sci. 11 (5), 702–712. https://doi.org/10.1177/1745691616658637.

van der Wel, P., van Steenbergen, H., 2018. Pupil dilation as an index of effort in cognitive control tasks: a review. Psychon. Bull. Rev. 25 (6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y.

Voracek, M., Kossmeier, M., Tran, U.S., 2019. Which data to meta-analyze, and how? Z. Psychol. 227 (1), 64–82. https://doi.org/10.1027/2151-2604/a000357.

Wright, R.A., 1996. Brehm's theory of motivation as a model of effort and cardiovascular response. In: Gollwitzer, P.M., Bargh, J.A. (Eds.), The Psychology of Action: Linking Cognition and Motivation to Behavior. Guilford Press, pp. 424–453.