



Vazul: An R Package for Analysis Blinding

Tamás Nagy 

ELTE Eotvos Lorend University

Alexandra Sarafoglou 

University of Amsterdam

Abstract

The abstract of the article.

Keywords: keywords, not capitalized, Java.

1. Introduction

In data analysis, researchers face the challenge of maintaining objectivity and minimizing bias. A central concern is the considerable analytic flexibility present at various stages of the workflow, including data preprocessing, variable selection, model specification, and statistical testing. This flexibility has been described as the “garden of forking paths” (Gelman and Loken 2013), meaning that possibly hundreds or thousands of plausible, non-redundant analytic strategies can be applied to the same dataset (Steegen, Tuerlinckx, Gelman, and Vanpaemel 2016; Simonsohn, Nelson, and Simmons 2020). Crucially, different analytic paths can yield different results, and analysts may, intentionally or not, gravitate toward those that align with their expectations.

The result can be biased or even spurious findings (Ioannidis 2005b; Simmons, Nelson, and Simonsohn 2011; Gelman and Loken 2016), which have eroded trust in science and contributed to replication failures across medical sciences, psychology, economics, and computer science (Ioannidis 2005a; Open Science Collaboration 2015; Baker 2016; Nosek, Hardwicke, Moshontz, Allard, Corker, Dreber, Fidler, Hilgard, Kline Struhl, Nuijten *et al.* 2022; Pashler and Wagenmakers 2012; Camerer, Dreber, Forsell, Ho, Huber, Johannesson, Kirchler, Almenberg, Altmejd, Chan *et al.* 2016; Cockburn, Dragicevic, Besançon, and Gutwin 2020).¹ To ensure the reliability of empirical findings, it is therefore essential to prevent the idiosyncrasies of the

¹But see Gilbert, King, Pettigrew, and Wilson (2016), Muradchanian, Hoekstra, Kiers, and van Ravenzwaaij (2021), Savitz, Wise, Bond, Hatch, Neube, Wesselink, Willis, Yland, and Rothman (2024), and McShane, Bradlow, Lynch Jr, and Meyer (2024) for critiques of how replication failure is measured and of the design and validity of some replication attempts.

data from feeding back into the formulation of the hypotheses being tested (HARKing, Kerr 1998) or from prompting researchers to exploit their analytic freedom to accentuate desired outcomes (Simmons *et al.* 2011). In response, researchers have proposed methodologies to limit the influence of bias in data analysis. One such method is analysis blinding, which supplies analysts with an altered version of the real data that conceals biasing features until the analysis plan is finalized.

Analysis blinding, or blind analysis (MacCoun and Perlmutter 2015), is a methodological approach designed to protect the confirmatory status of an analysis by concealing crucial test-relevant aspects of the data from the analysts—for example, by scrambling dependent variables or masking key labels. The procedure typically involves two independent parties: a data manager, who has full access to the raw data and is responsible for implementing the blinding steps, and an analyst, who is tasked with developing the analysis plan. After the data manager applies the blinding procedure, the analyst receives the resulting modified dataset: the blinded data.

Working with the blinded data, the analyst can explore and preprocess the dataset and develop an analysis plan without being influenced by whether a particular analytic choice produces the desired result, as any patterns in the blinded data are, by design, only a product of chance. Importantly, although the test-relevant elements are concealed, other meaningful characteristics remain intact, including demographic information, secondary variables, and the distributional properties of both dependent and independent variables. As a result, the analyst retains all information necessary to tailor the analytic approach to the specific, and potentially unexpected, features of the data. Once the analysis plan has been finalized using the blinded data, the analyst is granted access to the raw, unblinded dataset, and the confirmatory analysis can then be carried out strictly according to the pre-established plan.

Analysis blinding ties in with other methodologies to minimize bias in the research cycle, such as single-blind or double-blind experimental designs, where participants and/or experimenters are unaware of certain aspects of the study to prevent bias in data collection (Schulz and Grimes 2002). Here, the analysts is intentionally kept unaware of certain aspects of the data to prevent bias in data analysis, which is why the methodology is also referred to as triple-blind (Schulz and Grimes 2002).

Although many blinding techniques are conceptually straightforward, such as scrambling outcome variables or masking variable names and factor labels, their practical implementation can become technically challenging once real research designs and data structures are considered. For instance, in hierarchical models, researchers may need to scramble the outcome variable to destroy the effect of interest while still preserving information within grouping levels, such as country-level means in cross-cultural datasets (e.g., Sarafoglou, Hoogeveen, and Wagenmakers 2023). Similarly, in repeated-measures designs, condition labels may need to be masked so that each participant’s condition structure is preserved while permuting the mapping of conditions across participants to obscure the true experimental effect. These complexities make manual analysis blinding difficult to carry out consistently and correctly. To lower barriers for data managers and facilitate the broader adoption of analysis blinding in research labs, dedicated software tools, such as R packages, are needed to implement these procedures reliably.

To meet this need, the **vazul** package in R (<https://CRAN.R-project.org/package=vazul>) offers flexible and reproducible tools for implementing analysis blinding. It provides func-

tionality for data scrambling as well as data and variable masking. In the remainder of this article, we introduce the methodology in more detail, describe the functionalities of the **vazul** package, and illustrate its application through practical examples. We conclude with a brief discussion and outline future directions.

1.1. Analysis Blinding: Safeguarding Against Bias Without Preregistration

Originally, analysis blinding gained traction in astrophysics in the early 2000s ([MacCoun and Perlmutter 2018](#)) as a means of safeguarding analysts against bias. Since then, the methodology has also been advocated in the social and behavioral sciences as an alternative or complement to preregistration ([Dutilh, Sarafoglou, and Wagenmakers 2019](#); [Nagy, Hergert, Elsherif, Wallrich, Schmidt, Waltzer, Payne, Gjoneska, Seetahul, Wang, Scharfenberg, Tyson, Yang, Skvortsova, Alarie, Graves, Sotola, Moreau, and Rubínová 2025](#); [MacCoun 2021](#); [MacCoun and Perlmutter 2015](#); [MacCoun and Perlmutter 2018](#); [Aczel, Szaszi, Sarafoglou, Kekcs, Kucharský, Benjamin, Chambers, Fisher, Gelman, Gernsbacher, Ioannidis, Johnson, Jonas, Kousta, Lilienfeld, Lindsay, Morey, Munafò, Newell, Pashler, Shanks, Simons, Wicherts, Albarracin, Anderson, Antonakis, Arkes, Back, Banks, Beevers, Bennett, Bleidorn, Boyer, Cacciari, Carter, Cesario, Clifton, Conroy, Cortese, Cosci, Cowan, Crawford, Crone, Curtin, Engle, Farrell, Fearon, Fichman, Frankenhuus, Freund, Gaskell, Giner-Sorolla, Green, Greene, Harlow, Hoces de la Guardia, Isaacowitz, Kolodner, Lieberman, Logan, Mendes, Moersdorf, Nyhan, Pollack, Sullivan, Vazire, and Wagenmakers 2020](#)). This interest partly reflects the fact that analysis blinding addresses several of preregistration’s major shortcomings. Preregistration can restrict analysts so rigidly that they are unable to adapt their analyses to unexpected peculiarities in the data. Although deviations from preregistered plans are both possible and accepted when transparently disclosed in the final manuscript, such deviations run counter to the spirit of preregistration, as they reintroduce data-dependent decisions that may bias the results. At the same time, text-based preregistrations, that is, the description of analysis plans in a preregistration template, often fail to be “specific, precise, and exhaustive” ([Wicherts, Veldkamp, Augusteijn, Bakker, van Aert, and van Assen 2016](#), 2), which leaves substantial degrees of freedom unaccounted for.

In contrast to preregistration, analysis blinding allows researchers to maintain flexibility in their analysis plans, as they can explore the data without being constrained by a rigid pre-registered protocol. This flexibility is especially important in analyses involving advanced statistical modeling or preprocessing, where not all decisions can be anticipated in advance and researchers must make data-dependent choices. For instance, many psychological constructs, such as well-being, are measured with multi-item scales whose factor structure must be examined before further analysis; whether the items load onto a single factor or multiple factors determines subsequent analytic steps, including whether composite scores can be computed or whether a more complex measurement model is required (see e.g., [Hoogeveen, Sarafoglou, van Elk, and Wagenmakers 2023](#)). The flexibility gained by analysis blinding enables researchers to develop analysis strategies that are appropriately tailored to the specific characteristics of the dataset without the need to anticipate and specify all eventualities in advance. Consequently, researchers who analyze their data with analysis blinding are less likely to deviate from their developed analysis plan compared to researchers who preregister their analytic strategy in a text-based format ([Sarafoglou *et al.* 2023](#)).

Moreover, similar in spirit to approaches proposed in code-based preregistration (Peikert, Van Lissa, and Brandmaier 2021; Van Lissa, Brandmaier, Brinkman, Lamprecht, Peikert, Struiksma, and Vreede 2021), analysis blinding naturally produces analysis plans that are specific, precise, and exhaustive, because they are written directly in code rather than described verbally in preregistration templates. As with preregistration, analysis scripts based on the blinded data can be uploaded or attached to preregistration platforms (e.g., the Open Science Framework) to provide a transparent record of the planned analysis before the data are unblinded. In short, analysis blinding can serve as a valuable alternative or additional safeguard to preregistered analysis plans: it effectively protects against bias while providing analysts with the flexibility needed to develop an analysis strategy that is optimally suited to their data.

References

- Aczel B, Szaszi B, Sarafoglou A, Kekecs Z, Kucharský v, Benjamin D, Chambers CD, Fisher A, Gelman A, Gernsbacher MA, Ioannidis JP, Johnson E, Jonas K, Kousta S, Lilienfeld SO, Lindsay DS, Morey CC, Munafò M, Newell BR, Pashler H, Shanks DR, Simons DJ, Wicherts JM, Albarracín D, Anderson ND, Antonakis J, Arkes H, Back MD, Banks GC, Beevers C, Bennett AA, Bleidorn W, Boyer TW, Cacciari C, Carter AS, Cesario J, Clifton C, Conroy RM, Cortese M, Cosci F, Cowan N, Crawford J, Crone EA, Curtin J, Engle R, Farrell S, Fearon P, Fichman M, Frankenhuys W, Freund AM, Gaskell MG, Giner-Sorolla R, Green DP, Greene RL, Harlow LL, Hoces de la Guardia F, Isaacowitz D, Kolodner J, Lieberman D, Logan GD, Mendes WB, Moersdorf L, Nyhan B, Pollack J, Sullivan C, Vazire S, Wagenmakers EJ (2020). “A Consensus-Based Transparency Checklist.” *Nature Human Behaviour*, **4**, 4–6. doi:[10.1038/s41562-019-0772-6](https://doi.org/10.1038/s41562-019-0772-6).
- Baker M (2016). “1,500 scientists lift the lid on reproducibility.” *Nature News*, **533**, 452. doi:[10.1038/533452a](https://doi.org/10.1038/533452a).
- Camerer CF, Dreber A, Forsell E, Ho TH, Huber J, Johannesson M, Kirchler M, Almenberg J, Altmejd A, Chan T, et al. (2016). “Evaluating replicability of laboratory experiments in economics.” *Science*, **351**(6280), 1433–1436. doi:[10.1126/science.aaf0918](https://doi.org/10.1126/science.aaf0918).
- Cockburn A, Dragicevic P, Besançon L, Gutwin C (2020). “Threats of a replication crisis in empirical computer science.” *Communications of the ACM*, **63**(8), 70–79. doi:[10.1145/3360311](https://doi.org/10.1145/3360311).
- Dutilh G, Sarafoglou A, Wagenmakers EJ (2019). “Flexible Yet Fair: Blinding Analyses in Experimental Psychology.” *Synthese*, **198**, S5745–S5772. doi:[10.1007/s11229-019-02456-7](https://doi.org/10.1007/s11229-019-02456-7).
- Gelman A, Loken E (2013). “The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time.” *Department of Statistics, Columbia University*, **348**. URL http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.
- Gelman A, Loken E (2016). “The statistical crisis in science.” In M Pitici (ed.), *The best writing on mathematics*, volume 102, pp. 305–318. Princeton University Press.

- Gilbert DT, King G, Pettigrew S, Wilson TD (2016). “Comment on “Estimating the reproducibility of psychological science”.” *Science*, **351**(6277), 1037–1037. [doi:10.1126/science.aad7243](https://doi.org/10.1126/science.aad7243).
- Hoogeveen S, Sarafoglou A, van Elk M, Wagenmakers EJ (2023). “A Many-Analysts Approach to the Relation between Religiosity and Well-Being: Reflection and Conclusion.” *Religion, Brain & Behavior*, **13**(3), 356–363. [doi:10.1080/2153599X.2022.2070263](https://doi.org/10.1080/2153599X.2022.2070263).
- Ioannidis JP (2005a). “Contradicted and initially stronger effects in highly cited clinical research.” *JAMA*, **294**(2), 218–228. [doi:10.1001/jama.294.2.218](https://doi.org/10.1001/jama.294.2.218).
- Ioannidis JP (2005b). “Why most published research findings are false.” *PLoS medicine*, **2**(8), e124. [doi:10.1371/journal.pmed.1004085](https://doi.org/10.1371/journal.pmed.1004085).
- Kerr N (1998). “HARKing: Hypothesizing After the Results are Known.” *Personality and Social Psychology Review*, **2**, 196–217. [doi:10.1207/s15327957pspr0203_4](https://doi.org/10.1207/s15327957pspr0203_4).
- MacCoun R (2021). “Blinding to remove biases in science and society.” In R Hertwig, C Engel (eds.), *Deliberate ignorance: Choosing not to know*, pp. 51–64. MIT Press, Cambridge.
- MacCoun R, Perlmutter S (2015). “Blind analysis: Hide results to seek the truth.” *Nature*, **526**, 187–190. [doi:10.1038/526187a](https://doi.org/10.1038/526187a).
- MacCoun R, Perlmutter S (2018). “Blind Analysis as a Correction for Confirmatory Bias in Physics and in Psychology.” In SO Lilienfeld, I Waldman (eds.), *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions*, pp. 297–322. John Wiley and Sons. [doi:10.1002/9781119095910.ch15](https://doi.org/10.1002/9781119095910.ch15).
- McShane BB, Bradlow ET, Lynch Jr JG, Meyer RJ (2024). ““Statistical Significance” and Statistical Reporting: Moving Beyond Binary.” *Journal of Marketing*, p. 00222429231216910. [doi:10.1177/00222429231216910](https://doi.org/10.1177/00222429231216910).
- Muradchanian J, Hoekstra R, Kiers H, van Ravenzwaaij D (2021). “How best to quantify replication success? A simulation study on the comparison of replication success metrics.” *Royal Society Open Science*, **8**, 201697. [doi:10.1098/rsos.201697](https://doi.org/10.1098/rsos.201697).
- Nagy T, Hergert J, Elsherif MM, Wallrich L, Schmidt K, Waltzer T, Payne JW, Gjoneska B, Seetahul Y, Wang YA, Scharfenberg D, Tyson G, Yang YF, Skvortsova A, Alarie S, Graves K, Sotola LK, Moreau D, Rubínová E (2025). “Bestiary of Questionable Research Practices in Psychology.” *Advances in Methods and Practices in Psychological Science*, **8**(3), 25152459251348431. ISSN 2515-2459, 2515-2467. [doi:10.1177/25152459251348431](https://doi.org/10.1177/25152459251348431).
- Nosek BA, Hardwicke TE, Moshontz H, Allard A, Corker KS, Dreber A, Fidler F, Hilgard J, Kline Struhl M, Nuijten MB, et al. (2022). “Replicability, robustness, and reproducibility in psychological science.” *Annual review of psychology*, **73**, 719–748. [doi:10.1146/annurev-psych-020821-114157](https://doi.org/10.1146/annurev-psych-020821-114157).
- Open Science Collaboration (2015). “Estimating the Reproducibility of Psychological Science.” *Science*, **349**, aac4716. [doi:10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716).

- Pashler H, Wagenmakers EJ (2012). “Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence?” *Perspectives on Psychological Science*, **7**, 528–530. [doi:10.1177/1745691612465253](https://doi.org/10.1177/1745691612465253).
- Peikert A, Van Lissa CJ, Brandmaier AM (2021). “Reproducible research in R: A tutorial on how to do the same thing more than once.” *Psych*, **3**(4), 836–867. [doi:10.3390/psych3040053](https://doi.org/10.3390/psych3040053).
- Sarafoglou A, Hoogeveen S, Wagenmakers EJ (2023). “Comparing analysis blinding with preregistration in the Many-Analysts Religion Project.” *Advances in Methods and Practices in Psychological Science*, **6**(1), 25152459221128319. [doi:10.1177/25152459221128319](https://doi.org/10.1177/25152459221128319).
- Savitz DA, Wise LA, Bond JC, Hatch EE, Ncube CN, Wesselink AK, Willis MD, Yland JJ, Rothman KJ (2024). “Responding to Reviewers and Editors About Statistical Significance Testing.” *Annals of Internal Medicine*, **177**, 385–386. [doi:10.7326/M23-2430](https://doi.org/10.7326/M23-2430).
- Schulz KF, Grimes DA (2002). “Blinding in randomised trials: Hiding who got what.” *The Lancet*, **359**(9307), 696–700.
- Simmons J, Nelson L, Simonsohn U (2011). “False-positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant.” *Psychological Science*, **22**, 1359–1366. [doi:10.1177/0956797611417632](https://doi.org/10.1177/0956797611417632).
- Simonsohn U, Nelson LD, Simmons JP (2020). “Specification curve analysis.” *Nature Human Behaviour*, **4**, 1208–1214. [doi:10.1038/s41562-020-0912-z](https://doi.org/10.1038/s41562-020-0912-z).
- Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W (2016). “Increasing transparency through a multiverse analysis.” *Perspectives on Psychological Science*, **11**, 702–712. [doi:10.1177/1745691616658637](https://doi.org/10.1177/1745691616658637).
- Van Lissa CJ, Brandmaier AM, Brinkman L, Lamprecht AL, Peikert A, Struiksma ME, Vreede BM (2021). “WORCS: A workflow for open reproducible code in science.” *Data Science*, **4**(1), 29–49. [doi:10.3233/DS-210031](https://doi.org/10.3233/DS-210031).
- Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM (2016). “Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking.” *Frontiers in Psychology*, **7**, 1832. [doi:10.3389/fpsyg.2016.01832](https://doi.org/10.3389/fpsyg.2016.01832).

Affiliation:

Tamás Nagy
 ELTE Eotvos Lorand University
 Institute of Psychology,
 ELTE Eotvos Lorand University,
 Budapest, Hungary
 E-mail: nagy.tamas@ppk.elte.hu

Alexandra Sarafoglou
University of Amsterdam
Department of Psychology,
University of Amsterdam,
Amsterdam, The Netherlands
E-mail: a.s.g.sarafoglou@uva.nl