**Factors impacting effective altruism: Revisiting heuristics and biases in charity in a replication and extensions Registered Report of Baron and Szymanska (2011)**

Mannix Chan  
ORCID: 0000-0001-9699-9522  
Department of Psychology, University of Hong Kong  
[mannix42@connect.hku.hk](mailto:mannix42@connect.hku.hk) / [mannixchan@proton.me](mailto:mannixchan@proton.me)

^Gilad Feldman  
ORCID: 0000-0003-2812-6599  
University of Hong Kong  
gfeldman@hku.hk / giladfel@gmail.com

^Corresponding author

## Author bios:

Mannix Chan was a thesis student at the University of Hong Kong during the academic year 2022-3.

Gilad Feldman is an assistant professor with the University of Hong Kong Psychology department. His research focuses on judgement and decision-making.

**Declaration of Conflict of Interest:**

The author(s) declare no potential conflicts of interests with respect to the authorship and/or publication of this article.

**Financial disclosure/funding:**

This project is supported by the University of Hong Kong Teaching Development Grant.

**Authorship declaration:**

Mannix Chan conducted the replication as part of his thesis in Psychology.

Gilad Feldman guided the replication efforts, supervised each step in the project, ran data collection, conducted the pre-registration, and edited the manuscript for submission.

[Open to revision following in-principle acceptance.]

**Corresponding author**

Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; [gfeldman@hku.hk](mailto:gfeldman@hku.hk) ; 0000-0003-2812-6599

**Rights:**

CC-BY or an equivalent license is applied to the AAM arising from this submission. (clarification)

**Citation of the target research article:**

Baron, J., & Szymanska, E. (2011). Heuristics and Biases in Charity. In D. M. Oppenheimer & C. Y. Olivola (Eds.), *The Science of Giving: Experimental Approaches to the Study of Charity* (pp. 215–235). Psychology Press. <https://doi.org/10.4324/9780203865972-24>

## Contributor Roles Taxonomy

|  |  |  |
| --- | --- | --- |
| **Role** | **Mannix Chan** | **Gilad Feldman** |
| Conceptualization | X | X |
| Pre-registration | X | X |
| Data curation |  | X |
| Formal analysis | X |  |
| Funding acquisition |  | X |
| Investigation | X |  |
| Pre-registration peer review / verification |  | X |
| Data analysis peer review / verification |  | X |
| Methodology | X |  |
| Project administration |  | X |
| Resources |  | X |
| Software | X |  |
| Supervision |  | X |
| Validation |  | X |
| Visualization | X |  |
| Writing-original draft | X |  |
| Writing-review and editing |  | X |

# Abstract

[IMPORTANT: The method and results sections were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. No actual pre-registration or data collection has taken place yet; all statistics and sentences written in the past tense are placeholders.]

Individuals who donate to charity may be affected by various biases and donate inefficiently. In a replication and extension Registered Report with a US Amazon Mechanical Turk sample using CloudResearch (*N* = 1400), we replicated Studies 1 to 4 in Baron and Szymanska (2011) with extensions on reputation and overhead funding.

[The following findings are based on simulated random noise and will be updated after data collection:]

We found support for the effect of diversification on donations, no support for the effects of waste/overhead, past costs, and forced charity on donations, and mixed support for the ingroup effect on donations ([effects and CIs]). Extending the replication, we found no support for the impact of reputation on donation allocation ([effects and CIs]), and no support for impact of paid-for overhead costs on donation allocation ([effects and CIs]). We discuss the implications and validity of these findings. All materials, data, and code were made available on: <https://osf.io/bep78/>.

*Keywords:* effective altruism, heuristics, utilitarianism, donations, efficacy, charity, cognitive biases, registered report, replication

# PCIRR-Study Design Table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Question | Hypothesis | Sampling plan | Analysis plan  (See results section) | Rationale | Interpretation given different outcomes | Theory impact |
| How does charity perceived waste (overhead) impact donations? | People prefer to donate to charities with lower perceived waste, even when efficiency is held constant (H1) | This  study aims to  recruit 1400 participants online on Amazon MTurk via CloudResearch. | One sample t-tests against mid-point. | Power analyses indicate that this planned sample size should be well-powered enough  to detect effects  much weaker  than the smallest  effects in the  target article. See the power  analysis section of this manuscript. | We interpret replication results based on criteria in LeBel et al. (2019) by comparing our replication effect sizes and confidence intervals to the original effect sizes in the target article.  We will conclude a successful replication if at least 80% of the studies (i.e., 4 or 5, out of 5) showed a signal in the same direction as the original study by Baron and Szymanska (2011), a failed replication if only one or no studies (out of 5) showed a signal in the same direction as the original, and any mixed findings with lower than 80% and above 20% (i.e., 2 or 3, out of 5) to be a mixed results replication. | The theory that people donate inefficiently due to various heuristics and biases. |
| How do charity past costs (average benefit per dollar) impact donations? | People prefer to donate to charities with lower the past costs (average benefit per dollar), even when past costs are irrelevant in the context (H2) |
| In a choice between two charities, would people prefer diversification or efficiency? | People tend to diversify their donations (donate to a larger number of charities), even when it means that their donations are less efficient overall (H3) | Series of one sample t-tests against mid-point. paired t-test, one-way repeated measures ANOVA |
| In a choice between local and foreign charities, would people prefer to donate to causes in their own country over foreign countries? | People prefer to donate to causes within their own country than to causes in other countries (H4) | One sample t-test, one-way repeated measures ANOVA |
| Do people prefer to do good through government tax or voluntary charity? | People prefer to help through voluntary donations over forced charity (government taxes) (H5) |
| Would people prefer to donate publicly or anonymously? | People prefer to donate publicly than to donate anonymously (H6) | One sample t-test |
| Would people donate more to charities where the overhead is paid for by another donor? | People prefer to donate to charities with overhead paid for by other donors (H7) |

# Factors impacting effective altruism: Revisiting heuristics and biases in charity in a replication and extensions Registered Report of Baron and Szymanska (2011)

[IMPORTANT:   
This section is written in the past tense to simulate what the manuscript will look like after the data collection; no pre-registration or data collection has taken place yet.]

## Background

There are many factors that influence a person’s decision to donate money to a charitable cause. Bekkers and Wiepking (2011) identified eight separate mechanisms which drive charitable giving: awareness of need, solicitation, costs and benefits, altruism, reputation, psychological benefits, values, and efficacy. Some or all of these mechanisms may drive a person’s choice to donate.

Baron and Szymanska (2011) proposed that utilitarianism, which they defined as “the totality of good that comes about from a choice”, should be the objective standard from which to evaluate the efficiency of any given donation. They argued that charitable donations should be made aiming to maximize the most good possible using the same amount of money. However, people might not donate according to these standards due to mental shortcuts driven by cognitive constraints aiming to minimize use of cognitive resources, which end up going counter to the intended goal. Baron and Szymanska (2011) coined those as “non-utilitarian heuristics”, and their research demonstrated five heuristics or biases that cause deviations from the utilitarian model: 1) waste, 2) average cost, 3) diversification, 4) nationalism, and 5) forced charity.

We report a replication and extension Registered Report of Baron and Szymanska (2011) with the following goals. Our first goal was to conduct a close, independent, and well-powered replication of the target article and the effects of various non-utilitarian heuristics that drive charitable donations. Our second goal was to extend the target article’s design by addressing several unanswered further manipulating these heuristics. We hope to gain a better understanding of the different effects that result in suboptimal donations not aligned with maximizing overall good.

We begin by introducing the various heuristics and biases covered in Baron and Szymanska (2011), then discuss our motivation for the current replication study and the target’s hypotheses and study design, and conclude with our replication and extension design, needed adjustments, and added extensions.

### 1. Waste/overhead effect

Efficacy is one of the eight mechanisms that drive people to donate to charity. However, it can be difficult to evaluate the amount of good a donation of a certain amount will do, whereas it is often easier to evaluate other simpler factors (evaluability bias; Hsee, 1996). One such factor that people seem to pay more attention to is the relative amount of money a charity spends on overhead. Baron and Szymanska (2011) demonstrated that people appear biased against charities that have a higher overhead, even if those charities are actually more efficient in the goods they do, taking overhead costs into account.

Recent followup research by Caviola et al. (2014) showed further support for the idea that people seem more willing to donate more to charities with low overhead ratio, regardless of cost effectiveness.

### 2. Past costs effect

Baron and Szymanska (2011) demonstrated that people are less willing to donate when presented with a charity’s past costs. To test the past

They argued that the utilitarian approach to maximizing efficiency in charitable donations should be to evaluate the “marginal benefit per marginal dollar”; i.e. the extra benefit gained for a new contribution of a certain size. This means that people tend to be biased, for example, towards a charity which was cheaper to set up compared to one that took more money to set up, even if they could do the same or more good with a new donation of the same size, since they inaccurately take into account the previous efficiency of the charity in their evaluation, not just their current efficiency.

### 3. Diversification effect

The diversification effect is the phenomenon that people seek variety even when there is no reason to diversify (Read & Loewenstein, 1995). In the context of charitable giving, this effect can manifest in the tendency towards giving to many charities over a single charity in pursuit of a fair distribution (Fox et al., 2015). Baron and Szymanska (2011) argued that this effect extends even to cases where the many charities are, overall, less efficient than the single charity.

As a control condition, they also asked participants of Studies 1 and 2 how much they would allocate towards a charity running one project and a charity running several if they were equally effective; in this case, participants’ allocations were not found to differ from equal distribution.

### 4. Nationalism/Ingroup effect

Another possible non-utilitarian heuristic that Baron and Szymanska (2011) investigated is the effect of parochialism, which they defined as “a type of ingroup bias in which people weigh the welfare of their own group more heavily than those of outsiders”, commonly causing them to act in ways that benefit themselves over others.

Baron and Szymanska (2011) framed this effect through the lens of nationalism; however, we can also look at the effect of parochialism as being a manifestation of the ingroup effect, where people tend to more positively evaluate a group which they belong to compared to an analogous group to which they do not (Mullen et al., 1992).

Other studies have also shown that the ingroup effect appears in the context of charitable donations. For example, James and Zagefka (2017) found that people were willing to donate approximately 30% more money when told that the victims of a flood were from their own country over that of a fictional one; in a study of real-life online crowdfunding, Burtch et al. (2014) found that lenders preferred to lend money to those who were closer to them both in terms of culture and in physical distance.

### 5. Forced-charity/Government-taxes effect

Baron and Szymanska (2011) argued that people tend to prefer third-party organizations that collect voluntary donations for given beneficiaries over tax-supported government aid programs. They labeled these systems of taxation as “forced charity”, and argued that if voluntary donations and taxation result in the same amount of benefit to the same group of people, a utilitarianism standard would mean that these two programmes should be evaluated equally, and that preferring one over the other would be a form of bias.

## Choice of target for replication: Baron and Szymanska (2011)

We chose to replicate and extend Baron and Szymanska (2011) based on several factors.

First, the article has had an impact on scholarly research and practice, especially in the field of effective altruism. At the time of writing (March 2023), this book chapter has received 109 citations according to Google Scholar, and has set the foundations for a new field on the psychology of (in)effective altruism. This new domain has led to some groundbreaking work by scholars like Caviola et al. (2021) and Butts et al. (2019), as well as other impactful studies in the field such as Burtch et al. (2014) and Berman et al. (2018), investigating charitable behavior as affected by cultural differences and subjective preferences respectively.

Second, the design of the studies in this paper allowed for the straightforward inclusion of extensions to allow for additional tests and insights on impediments to effective altruism. The formatting of the studies’ questionnaires lent itself well to the inclusion of selected extensions examining other factors that may preclude the effectiveness of participant choice in donation allocations, which we achieved by inserting items of a similar format to the original studies without heavily increasing the complexity of the replication, thus striking a balance between coverage of different impediments to effective altruism and the complexity of the study and the associated data analyses.

Finally, to the best of our knowledge, there are currently no published independent direct replications of this chapter, and there is much potential in revisiting and expanding on many of its insights. The target chapter was very brief on both the description of what was done, the analyses conducted, and the results, and we hope that our reproduction of all the materials, the procedure, the analyses, and the results, would make it easier for others to follow and expand on this important work.

For these reasons, following the recent growing recognition of the importance of reproducibility and replicability in psychological science (Nosek et al., 2022), we aimed to revisit Baron and Szymanska (2011) by conducting a close, independent, and well-powered replication and extension Registered Report.

## Baron and Szymanska (2011): Findings and hypotheses

Four studies were conducted in Baron and Szymanska (2011) and we aimed at replicating all four studies. The studies were originally conducted as online questionnaires; the original authors directed us to the original questionnaires’ availability online. We therefore only needed to make adjustments of the study design using the old platform to Qualtrics, and we added extensions by inserting additional items into two of the studies (to be discussed in a subsequent section).

We summarized the findings in the target article in Table 1. Please note that for ease of reading, this manuscript (as well as the tables included within) follows the target article in that findings are sorted by hypothesis and not by study, as most hypotheses were tested over several different studies. Additionally, labels for each hypothesis follow that of the original study for ease of comparison.

**Table 1**

*Baron and Szymanska (2011): Summary of findings*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | 95% Confidence Interval | |
| Hypothesis | Study | *p* | *t* | *df* | Cohen’s *d* | Lower | Upper |
| 1 (Waste) | 1 | = .0000 | N/A | N/A | N/A | N/A | N/A |
|  | 2 | N/A | N/A | N/A | N/A | N/A | N/A |
|  | 3 | = .001 | 3.39 | 83 | 0.37 | 0.15 | 0.56 |
| 2 (Average cost) | 1 | N/A\* | N/A | N/A | N/A | N/A | N/A |
| 2 | = .0005 | 3.66 | 76 | 0.42 | 0.19 | 0.64 |
| 3 (Diversification) | 1 | N/A | N/A | N/A | N/A | N/A | N/A |
| 2 | N/A | N/A | N/A | N/A | N/A | N/A |
| 3 | N/A | N/A | N/A | N/A | N/A | N/A |
| 4 (means of responses) | = .0019 | 3.22 | 77 | 0.36 | 0.13 | 0.59 |
| 4 (proportion of responses) | = .0006 | 3.61 | 77 | 0.41 | 0.18 | 0.64 |
| 4 (Nationalism) | 1 | N/A | N/A | N/A | N/A | N/A | N/A |
| 2 | N/A | N/A | N/A | N/A | N/A | N/A |
|  | 3 | = .0000 | N/A | N/A | N/A | N/A | N/A |
| 5 (Forced charity) | 4 (version 1) | = .09 | N/A | N/A | N/A | N/A | N/A |
|  | 4 (version 3) | = .0057 | 2.84 | 77 | 0.32 | 0.09 | 0.55 |
|  | 4 (version 4) | = .0000 | 5.96 | 77 | 0.70 | 0.43 | 0.93 |

*Note*. Versions of studies without a reported p value are omitted for brevity. Items marked with N/A were not reported by the original authors or cannot be calculated. p values marked with an asterisk (\*) means that the result was reported as significant by the original authors but the p value was not mentioned. We calculated the Cohen’s *d* and confidence interval values wherever possible; see the section on effect size calculations in the supplementary materials.

## Extensions

### 6. Reputation / Publicity

In the context of charitable giving, reputation refers to “the social consequences of donations for the donor” (Bekkers & Wiepking, 2011). As the act of donating money is usually seen as a positive thing to do, the act of being observed donating money to a charitable cause could lead to the positive consequence of one’s reputation increasing. A meta-analysis conducted by Bradley et al. (2018) found that the feeling of being observed by others, whether actual or perceived, has a small but positive effect on prosocial behavior. For example, in their study, Alpizar et al. (2008) found that monetary donations made in public were 25% larger than ones made in private.

Therefore, we hypothesized that due to the additional benefits of having their reputation increase, people would show a preference towards donating to causes that could improve their reputation over ones that do not, even when efficiency is held constant.

### 7. Overhead costs external funding (Hypothesis 7)

Hypothesis 1 of the replication states that participants are less willing to donate to charities with a higher perceived waste or overhead, even when efficiency is held constant. However, a study by Gneezy et al. (2014) found that donors were more willing to donate to a charity with overhead costs if the costs were covered by another donor. Both the donation rate and the total amount of donations were found to increase when donors were told that the overhead costs are covered by another donor when compared to control groups and with other manipulations. Moreover, the findings in this paper were successfully replicated with an effect in the same direction in a mass replication effort by Camerer et al. (2018).

Extending these findings, we hypothesized that when people are presented with both options (a charity where overhead is covered by another donor and a charity where overhead is paid for by the donor), people have a preference towards the charity in which overhead is covered by another donor, even if both charities are equally effective and spend the same amount on overhead.

We constructed the hypotheses of our replication from the results of the target article, and we summarized the hypotheses of our replication as well as our extensions in Table 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | |  |  |
|  | | |  |  |
|  |  | |  |  |

**Pre-registration and open-science**

We provided all materials, data, and code at: <https://osf.io/bep78/>. This project has received a Peer Community in Registered Report Stage 1 in-principle acceptance (ENTER LINK AFTER IPA); (ENTER LINK AFTER IPA) after which we created a frozen pre-registration version of the entire Stage 1 packet at (ENTER LINK AFTER IPA) and proceeded to data collection. All measures, manipulations, and exclusions conducted for this investigation are reported, and data collection was completed before conducting the data analyses.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | |  |  |
| 1 | Waste/overhead | 1, 2, 3 | People prefer to donate to charities with lower perceived waste, even when efficiency is held constant |
| 2 | Past costs | 1, 2 | People prefer to donate to charities with lower the past costs (average benefit per dollar), even when past costs are irrelevant in the context |
| 3 | Diversification | 1, 2, 3, 4 | People tend to diversify their donations (donate to a larger number of charities), even when it means that their donations are less efficient overall. |
| 4 | Nationalism/Ingroup | 1, 2, 3 | People prefer to donate to causes within their own country than to causes in other countries |
| 5 | Forced-charity/Government-taxes | 4 | People prefer to help through voluntary donations over forced charity (government taxes) |
|  |  |  |  |
| 6 | Reputation / Publicity | 1, 2 | People prefer to donate publicly than to donate anonymously |
| 7 | Overhead costs external funding | 1, 2 | People prefer to donate to charities with overhead paid for by other donors |

# Method

[IMPORTANT:   
Method and results were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. For the purpose of the simulation, we wrote these sections in past tense, but no pre-registration or data collection took place yet.]

## Power and SESOI analyses

The original article recruited about 80 participants per study, for a total of approximately 320 participants overall. To prevent the replication from becoming underpowered, we conducted effect size calculations and power analyses based on the information and statistics reported by the target article. Based on the original article’s effect sizes, we found that the largest minimum sample size required in one study was 178 (for Hypothesis 5: Study 4, version 3). We ran the four studies separately, with participants evenly distributed, and therefore multiplied the minimum number of participants by four (178 × 4) resulting in 712.

However, to account for the possibility that the target’s effects were an overestimation, for possible exclusion of participants, and to allow for additional analyses, we conducted a SESOI analysis aiming for the ability to detect a Cohen’s *d* of 0.2 (power = 95%, alpha = 0.05) with one-sample and paired samples t-tests, commonly considered weak effects (Xiao et al., 2023). This required a sample size of 327 (and multiplied by 4 = 1308), for a larger total sample size of 1400 participants, accounting for possible exclusions due to incomplete data. Simonsohn (2015) suggests an approach of multiplying the initial sample size by 2.5, which would call for a total sample size of 320 × 4 = 800 participants; the total sample size of this replication is larger than this figure.

We provided more information regarding these calculations in the section on “Analysis of the original article” in the supplementary materials.

## Participants

[To demonstrate our results after data collection, we simulated a dataset of 800 participants using Qualtrics and reported our analyses below based on that dataset. Results will later be updated with the real data.]

Based on the SESOI analysis, we recruited a total of 1400 American Amazon Mechanical Turk (MTurk) participants through CloudResearch (Litman et al., 2017) (*Mage* = 58.59, *SD* = 24.17; 353 males, 356 females, 691 other/did not disclose). We summarized a comparison of the target article sample and the replication sample in Table 3.

Based on our experience of running similar judgment and decision making replications on MTurk, to ensure high quality data collection, the following CloudResearch options were employed: Duplicate IP Block, Duplicate Geocode Block, Suspicious Geocode Block, Verify Worker Country Location, Enhanced Privacy, CloudResearch Approved Participants, and Block Low Quality Participants. We also employed the following Qualtrics fraud and spam prevention measures: reCAPTCHA, prevent multiple submissions, prevent ballotstuffing, bot detection, security scan monitor, and relevantID.

[The assignment pay is based on the federal wage of 7.25USD/hour, per minute, so for example 5-8 minutes survey would be paid 1 USD per participant. We first pretested survey duration with 30 participants to make sure our time run estimate was accurate and adjusted pay as needed, the data of the 30 participants was not analyzed other than to assess survey completion duration and needed pay adjustments. For those pretest participants, if survey duration was longer than expected, they were paid a bonus as pay adjustment. The pretest participants' responses were included in the final analysis.]

**Table 3**

*Comparison between original study and replication participant demographics*

|  |  |  |
| --- | --- | --- |
|  | Baron and Szymanska (2011) | US MTurk workers (2023) |
| Sample size | ~320 | 1400 | |
| Geographic origin | Mostly Americans | US Americans | |
| Gender | About 80% female | 353 males, 356 females, 691 other/did not disclose | |
| Median age (years) | About 42 | 58 | |
| Average age (years) | Not given | 58.59 | |
| Age SD (years) | Not given | 24.17 | |
| Age range (years) | 20-80 | 18-100 | |
| Medium (location) | Online questionnaire | Online questionnaire | |
| Compensation | Nominal payment | Nominal payment | |
| Year | 2011 | 2023 | |

## 

## Design and procedure: Replication

We reconstructed the target’s stimuli and adjusted it to an online Qualtrics survey based on the information provided in the article. Participants indicated their consent, with four questions confirming their eligibility, understanding, and agreement with study terms, which they must answer with a “yes” and required responses in order to proceed to the study. Three of the four questions also served as attention checks, with the options order being rotated (yes, no, not sure).

All participants completed only one of the four studies. At the end of the study, they answered a number of funneling questions and provided their demographic information, two of which asked for the participant’s age and gender (*male/female/other/rather not disclose*), similar to the target article, then were debriefed.

The complete list of items and questions considered in our analysis for both the replication and the extension sections in all four studies can be found in the “Materials used in the replication + extension” section of the supplementary.

[For review: The Qualtrics survey .QSF file and an exported DOCX file are provided on the OSF folder. A preview link of the Qualtrics survey is provided on:

<https://hku.au1.qualtrics.com/jfe/preview/previewId/52089430-f111-4a71-ac07-b0070ee98486/SV_6eQJIy9YYTirmXI?Q_CHL=preview&Q_SurveyVersionID=current> ]

## 

## Manipulations

Participants saw the full set of items and questions in the study they were sorted into. Participants were first told to imagine that they have enough money for it to be easy for them to give some money away for charitable causes without seriously hurting their quality of life, and that they are willing to contribute some of their annual income to such causes. For each item, participants of each study were provided with descriptions of two different conditions, and asked to evaluate the two conditions using various scales.

## Measures

A full list of the items that we used in this section can be found in the supplementary.

### Replication

#### 1. Waste/overhead

Participants of Studies 1 through 3 read a description of two charities which differed in the relative amount of money spent on advertising or overhead (e.g., Study 1: “**A** and **B** help prevent deaths in children. Both of them can prevent 5 deaths for every $1,000 of donations. **A** spends $200 out of every $1,000 of donations on advertising. **B** spends $100.”). They then indicated the money allocation ratio between A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals).

#### 2. Past costs

Participants in Studies 1 and 2 read a description of two charities which differed in the overall cost per life already saved in the past, but that are equally efficient with new donations (e.g. Study 2: “**A** and **B** will each prevent 5 deaths for every $10,000 of new donations. **A** was much more expensive to get started. Thus, the cost per life saved on the average is higher for **A**, because **A** has spent more money in total.”). They then indicated the money allocation ratio between A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals).

#### 3. Diversification

For items in the “unequal efficiency” condition, participants in Studies 2 and 4 read several versions of a description of two charities which differed in their respective efficiency in saving lives (e.g. Study 2, version 1: “**A** can save one life for $10,000. **B** can save one life for $12,500. The people helped are from the same groups, with the same problems”). Participants in Study 2 indicated the money allocation ratio to A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals). Participants in Study 4 were asked the following questions: “What is the right allocation between A and B, ignoring your own feelings?” “What allocation would you feel best about making?” “What allocation between A and B would be the most efficient use of your money?” “What allocation between A and B would do the most good for each $1,000 spent?”. All questions were on a 5-point scale (1 = *All to* ***A***; 3 = *Equally to* ***A*** *and* ***B***; 5 = *All to* ***B***).

Additionally, for items in the “unequal efficiency, several projects versus one” condition, participants in Studies 1, 3, and 4 read a description of two charities which differed in that one charity is less effective but helps more groups of people than the other (e.g., Study 1: “**A** puts all the money into one project, which has a 75% chance of helping many children, and a 25% chance of doing no good at all. **B** puts the money into several different projects, each of which has a 70% chance of helping some children, but a 30% chance of doing no good”). Participants in Study 1 indicated money allocation ratio between A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals). Participants in Studies 3 and 4 indicated what the right allocation of money between A and B was (“What is the right allocation between A and B, ignoring your own feelings?”) on a 5-point scale (1 = *All to* ***A***; 3 = *Equally to* ***A*** *and* ***B***; 5 = *All to* ***B***); participants in Study 4 were also asked the question “What allocation between A and B would be the most efficient use of your money?” (1 = *All to* ***A***; 3 = *Equally to* ***A*** *and* ***B***; 5 = *All to* ***B***).

Finally, for items in the “equal efficiency” condition, participants of Studies 1 and 2 read a description of two charities which differed solely in the number of groups of children they helped (“A puts all the money into one project, which will help 100,000 children. B puts the money into five different projects, each of which will help 20,000 children. (The benefit per child will be the same.)”). They then indicated the money allocation ratio between A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals).

#### 4. Nationalism/Ingroup effect

Participants of Studies 1 through 3 read one or several descriptions of two charities which differed in the groups they help; one helps children in their own country, and the other helps children around the world or in a specific foreign country/region (e.g. Study 1: “**A** helps children who are in your own country. **B** helps children around the world. The children are equally needy.”). Participants of Studies 1 and 2 indicated the money allocation ratio between A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals). Participants in Study 3 indicated perceived right allocation of money between A and B (“What is the right allocation between A and B, ignoring your own feelings?”) on a 5-point scale (1 = *All to* ***A***; 3 = *Equally to* ***A*** *and* ***B***; 5 = *All to* ***B***).

#### 5. Forced-charity/Government-taxes

Participants in Study 4 read several scenarios about raising money. In each scenario, two cases were given: one where the money is raised through taxation (i.e. forced charity), and one where the money is raised by voluntary donations. For example, version 3 had the following:

“Workers in your country who make widgets [imaginary goods] are getting lower wages because of competition from foreign imports. The price of widgets has gone down, and the workers have accepted wage cuts to avoid layoffs.  
**Case A**: The government puts a tax on widgets. The proceeds from the tax are used to help the domestic workers by restoring their wages to their original level.  
**Case B**: A voluntary charity collects funds to help the domestic workers. The funds are sufficient to restore their wages to their original level.”

For each scenario, participants answered the following questions: “Which case would you favor if you had a choice?” “Which case is more fair in distributing the cost and benefits?” “Which case provides more freedom of choice?” (-1 = *Case A*; 0 = *Both cases are equal*; 1 = *Case B*), as well as the question “Which is more important in this scenario?” (-1 = *Fair cost allocation*; 0 = *Both are equal*; 1 = *Freedom of choice*).

### Extensions

#### 6. Reputation / publicity

Participants in Studies 1 and 2 read the description of two charities which differed only in that one publishes donor names and one does not (“**A** and **B** both help thousands of children. A publishes the names of donors and how much they donated on their website. **B** keeps donors anonymous.”). Participants indicated the money allocation ratio between A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals).

#### 7. Overhead costs external funding

Participants in Studies 1 and 2 read the description of two charities which differed only in that one charity used the donations to pay the overhead costs, whereas the other charity had another donor cover the overhead costs of the donation (“**A** and **B** both help thousands of children. Both charities spend 50% of the donations they receive on administrative costs. For each $100 contribution to **A**, $50 will go to helping children and $50 will be used to cover administrative costs. For each $100 contribution to **B**, all $100 will go to helping children; another donor will cover the corresponding $100 administrative cost of this contribution.”). Participants indicated the money allocation ratio between A and B (“How much would you allocate to A/B?”) on an 11-point scale (0 = “*A:100%, B:0%*”; 100 = “*A:0%, B:100%*”*;* options at 10% intervals).

## Evaluation criteria for replication findings

We aimed to compare our replication’s effects with those in the original article, wherever data was available, using the criteria set by LeBel et al. (2019) (see the subsection “Replication evaluation” in the supplementary).

We pre-register our overall strategy to conclude a successful replication if at least 80% of the studies (i.e., 4 or 5, out of 5) showed a signal in the same direction as the original study by Baron and Szymanska (2011), a failed replication if only one or no studies (out of 5) showed a signal in the same direction as the original, and any mixed findings with lower than 80% and above 20% (i.e., 2 or 3, out of 5) to be a mixed results replication.

## 

**Table 4**

*Replication classification based on LeBel et al. (2018)*

|  |  |  |
| --- | --- | --- |
| **Design facet** | **Replication** | **Details of deviation** |
| Effect/hypothesis | Same |  |
| IV construct | Same |  |
| DV construct | Same |  |
| IV operationalization | Same |  |
| DV operationalization | Same |  |
| Population (e.g. age) | Similar | Our replication is restricted to participants from the US; the original study had no such restriction but indicated the sample was “mostly Americans”. |
| IV stimuli | Same |  |
| DV stimuli | Same |  |
| Procedural details | Similar | Minor modifications of the formatting and wording to enhance clarity and comprehension. See the “comparisons and deviations” section of the supplementary materials for more details. |
| Physical settings | Similar | Online |
| Contextual variables | Similar | Participants were recruited online via CloudResearch instead of through a panel. |
| Replication classification | Very close replication |  |

## 

## Data analysis strategy

We conducted data analyses for both the replication and extension sections using RStudio (version 2022.12.0.353, running R version 4.2.2) with the packages "effectsize", "haven", "psych", "reshape", "statsExpressions", and "tidyverse", while graphs were generated using the packages "afex", "dplyr", "ggplot2", "ggstatsplot", "haven", "labelled", "PMCMRplus", "sjlabelled", and "reshape".

### Target alpha .005 and corrections.

The tests for some of the hypotheses involve several analyses on similar dependent variables in the same Study, such as having three analyses in Study 2 to test Hypothesis 3. Following on Editor Dr./Prof. Romain Espinosa suggestion to compensate for multiple analyses, we adjust our target alpha to .005 throughout. We will report raw p-values. For ANOVA analyses we will report Holm corrections for multiple analyses and will report both raw and corrected p-values, yet our criteria for signal will use the corrected p-values against the .005 alpha threshold.

We will complement our NHST reporting with Bayesian analyses reporting although we note that our replication success criteria will follow the NHST signal and directionality per the LeBel et al. (2019) criteria.

### Replication

Data analyses for the replication were conducted according to the information provided in the original article. For most hypotheses, a one-sample t-test was used to compare participant responses with an equal allocation, or to a 100% allocation to the one charity that was objectively more efficient or effective according to the utilitarian standards as set by Baron and Szymanska (2011). Paired t-tests were used to compare participant responses to different items in the same scenario or group of scenarios.

Additionally, for some studies that had multiple versions for the same hypothesis, we conducted one-way repeated measures ANOVAs tests to test for differences.

### Extensions

The data analyses for the extensions will follow the same structure of analogous items in the replication; one-sample t-tests will be used to compare participant responses with an equal allocation. As the items and questions posed to participants of both Study 1 and Study 2 are exactly the same in both extensions, the participant responses in both studies will be coalesced and analyzed as one single data set unlike in the replication.

### Assumption checks

We aimed to follow the target article in their analyses. The data analyzes in the original article were conducted using parametric one-sample and paired samples t-tests; these tests run under assumptions of normality and/or homogeneity of variance. We believe this is justified even if normality is violated given complexities inherent in normality tests and rerunning analyses with non-parametric tests (Knief, & Forstmeier, 2021).

However, if we fail to find support for the hypotheses and the assumptions, we will examine the possibility of normality and/or homogeneity violations for the failed analyses, and make adjustments accordingly. In case we find a violation of normality, we will conduct exploratory complementary non-parametric tests of the same tests to verify the validity of the results without the normality and/or variance homogeneity assumptions (with Wilcoxon tests being used in replacement of one-sample and paired t-tests and the Kruskal–Wallis test in replacement of the one-way ANOVA test), and with a stricter criteria using an alpha criterion to .005 to account for the multiple analyses.

## Supplementary Bayesian reporting

Our main analysis is using Null Hypothesis Significance Testing, which is focused on rejecting the null hypothesis, yet to address the possibility that we may fail to find support for rejecting the null, we complement all our analyses with Bayesian analyses using a prior of BF = 0.707. Bayes factor will be reported in our figures, using the R package ggstatsplot.

# Results

[IMPORTANT:   
Method and results were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. For the purpose of the simulation, we wrote these sections in past tense, but no pre-registration or data collection took place yet.]

## Replication

[Note: All ggstatsplot plots in the Rmarkdown code and outputs now also contain Bayes results based. These will be updated to replace the plots before following data collection in Stage 2.]

### 1. Waste/overhead (Hypothesis 1)

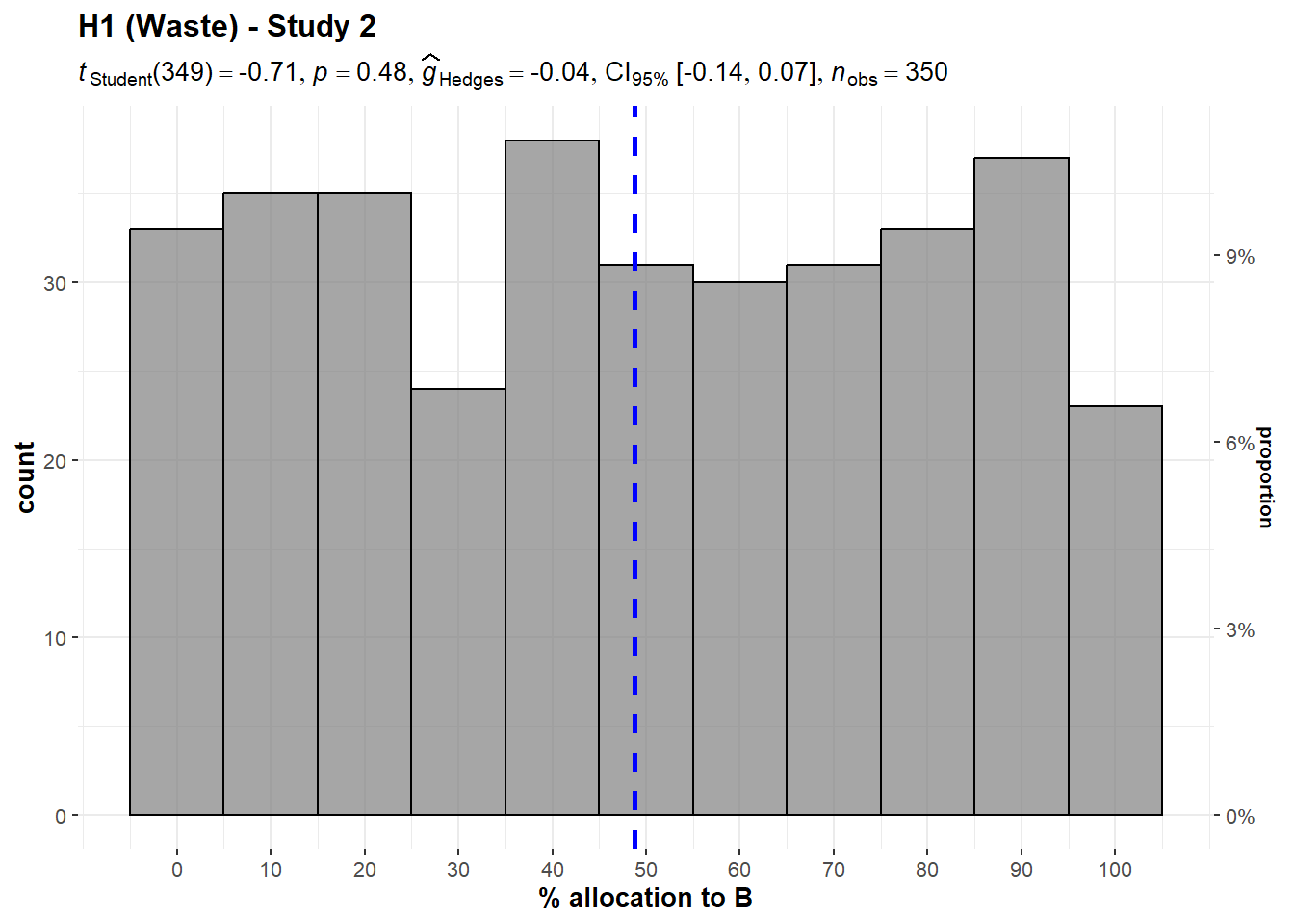
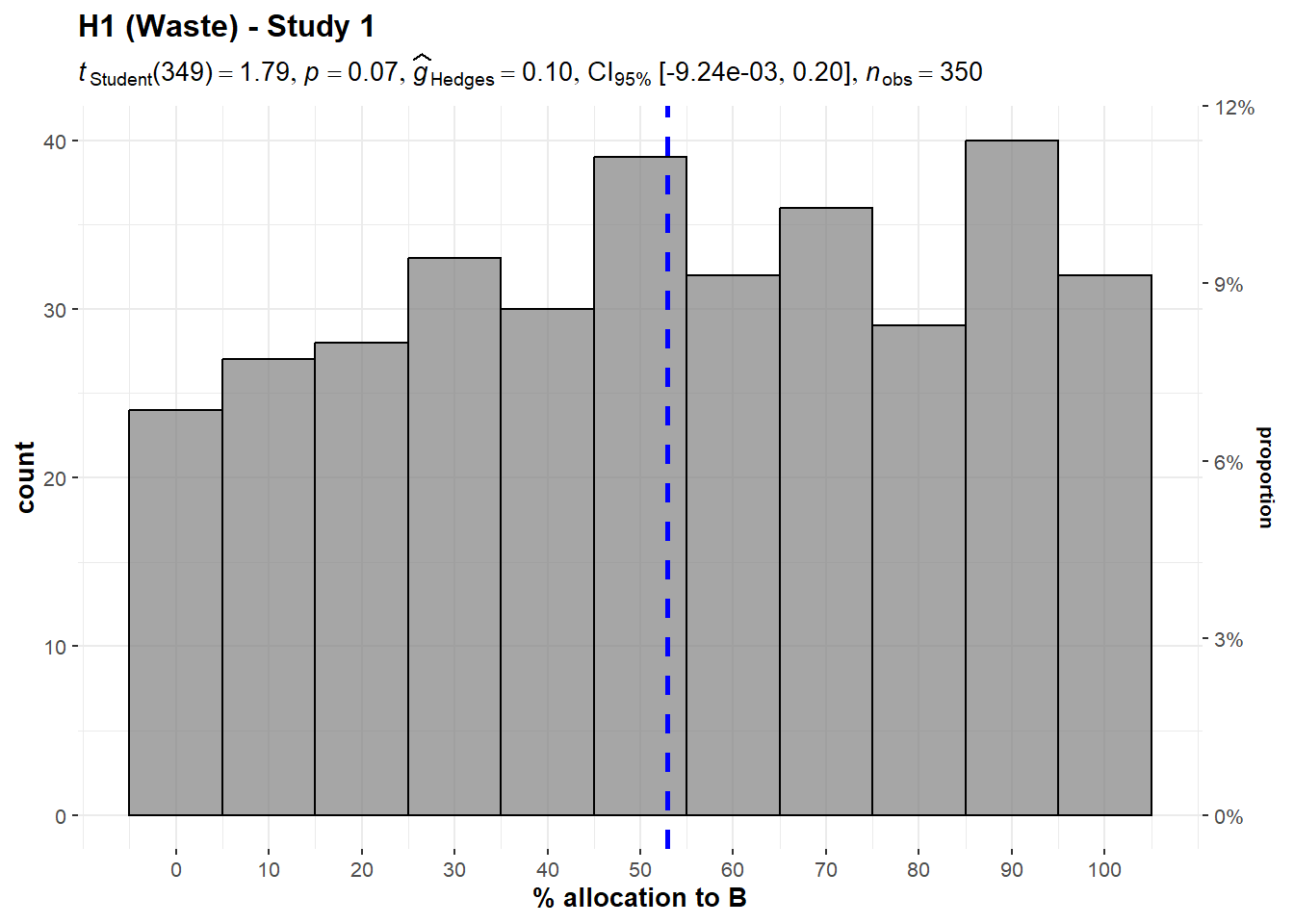
We conducted three one-sample t-tests for Studies 1 to 3, which we summarized and plotted in Figures 1 and 2. The differences were that Studies 1-2 were on a 0-100 scale, whereas Study 3 was on a 1-5 scale and is more explicit about efficiency with remaining funds.

We failed to find support for decreased participant willingness to donate to charities with a higher level of perceived waste in Study 1 (*M* = 52.94, *SD* = 30.72, *t*349 = 1.79, *p* = .074, *d* = 0.10, 95% CI [-0.01, 0.20]) and in Study 2 (*M* = 48.80, *SD* = 31.43, *t*349 = -0.71, *p* = .476, *d* = -0.04, 95% CI [-0.14, 0.07]) (against a midpoint of 50).

In Study 3, we also failed to find support for a decreased participant willingness to donate to charities with a higher level of perceived waste; participants did not allocate significantly more money to the charity with a lower level of perceived waste (against a midpoint of 3; *M* = 2.88, *SD* = 1.42, *t*348 = -1.54, *p* = .124, *d* = -0.08, 95% CI [-0.19, 0.02]).

**Figure 1**

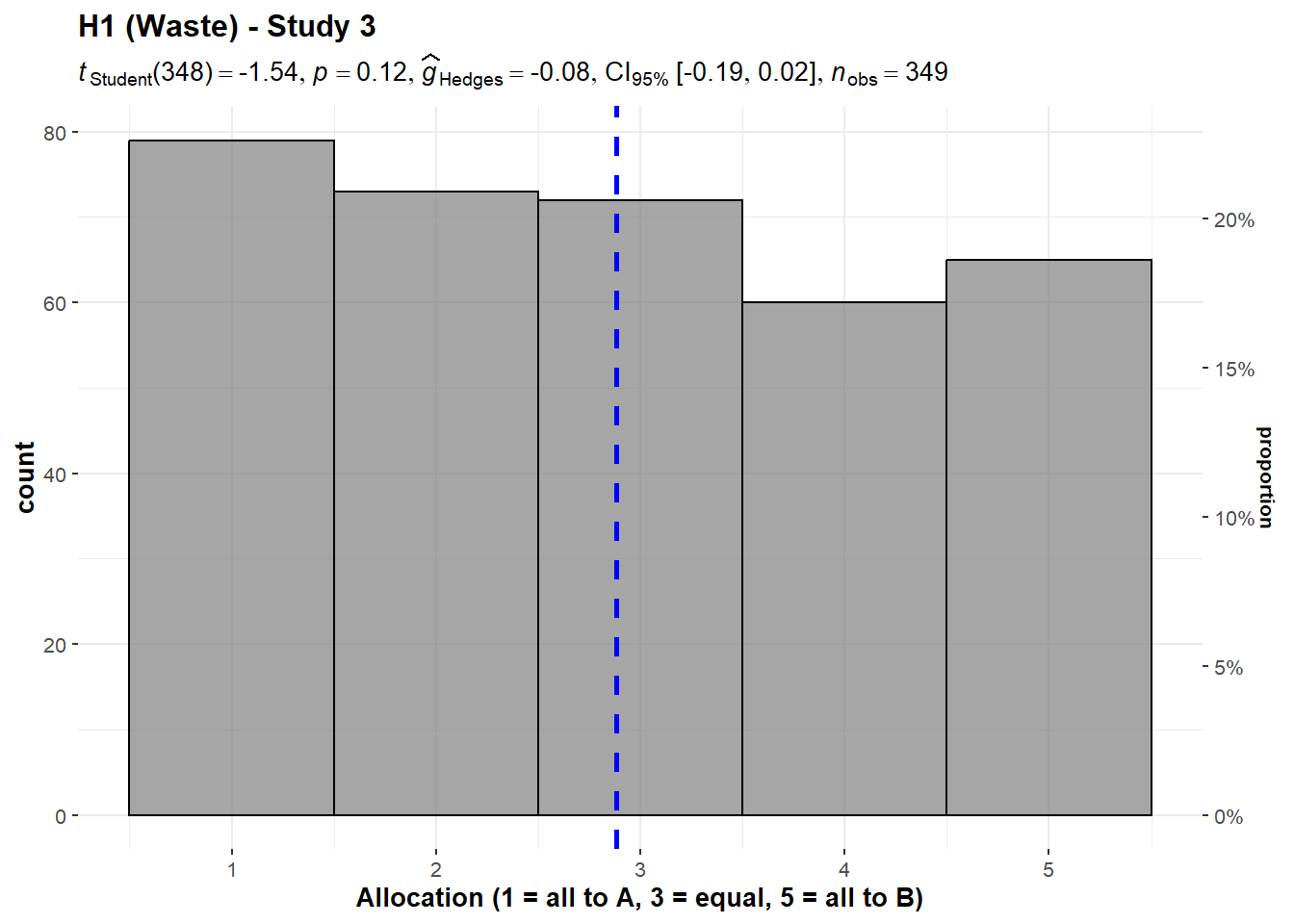
*Waste/overhead, Studies 1 and 2: Histograms for the effect of waste on allocation*



*Note*. The scenario: “**A** and **B** help prevent deaths in children. Both of them can prevent 5 deaths for every $1,000 **of donations** [bolded only in the Study 2 version]. **A** spends $200 out of every $1,000 of donations on advertising. **B** spends $100.”

**Figure 2**

*Waste/overhead, Study 3: Allocation*



*Note.* The scenario: “**A** and **B** help prevent deaths in children. Both of them can prevent 5 deaths for every $1,000 **of donations**. **A** spends $200 out of every $1,000 of donations on overhead expenses, but manages to save 5 lives with the remaining $800. **B** spends $100 out of every $1,000 on overhead, and saves 5 lives with the remaining $900.”

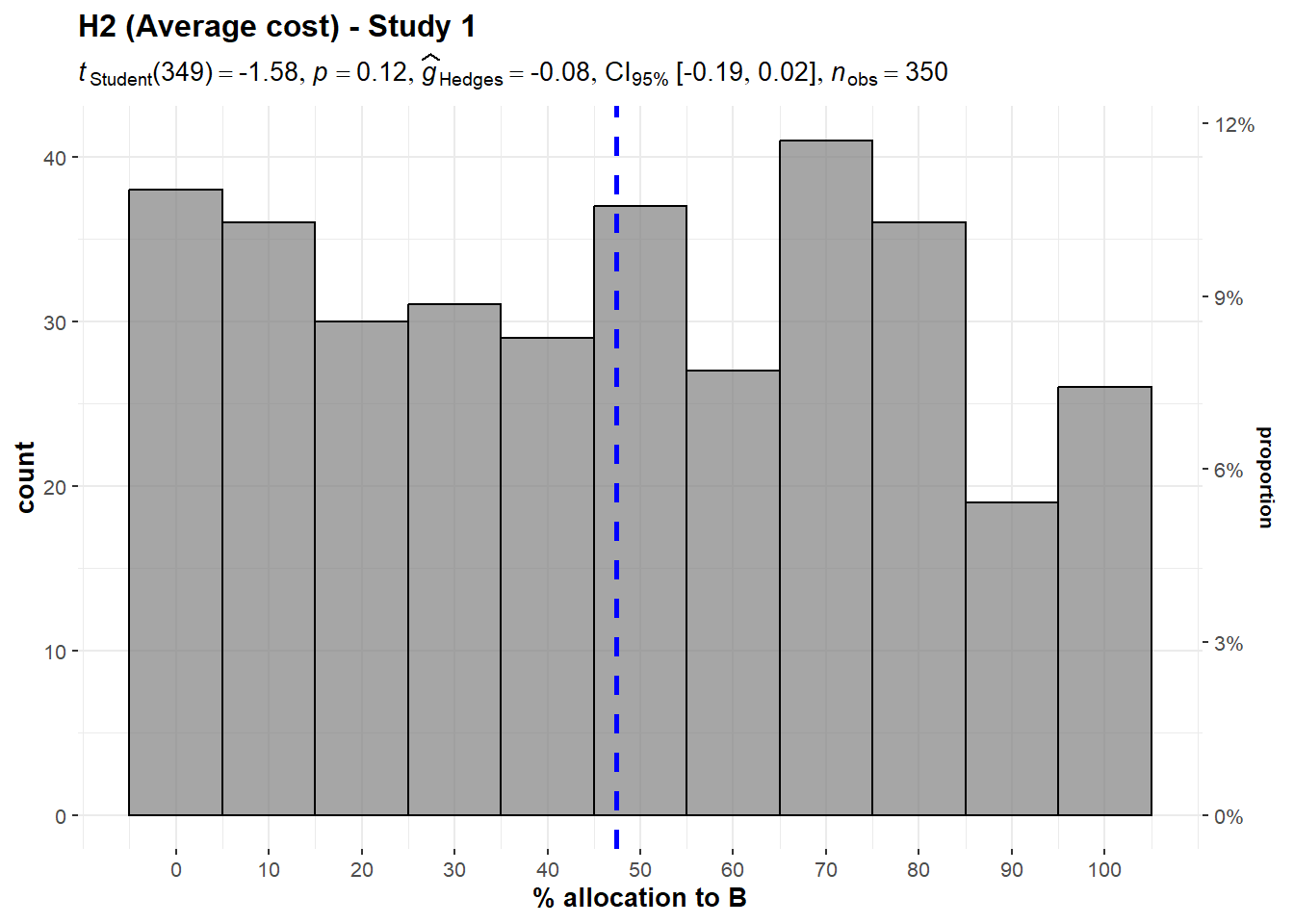
### 2. Past costs (Hypothesis 2)

We conducted two one-sample t-tests for Studies 1 and 2, which we summarized and plotted in Figures 3 and 4.

We failed to find support for the hypothesis that participants attend to average benefit per dollar even when irrelevant in Study 1 (*M* = 47.37, *SD* = 31.18, *t*349 = -1.58, *p* = .116, *d* = -0.08, 95% CI [-0.19, 0.02]) and in Study 2 (*M* = 49.2, *SD* = 31.96, *t*349 = -0.47, *p* = .639, *d* = -0.03, 95% CI [-0.13, 0.08]) (against a midpoint of 50).

**Figure 3**

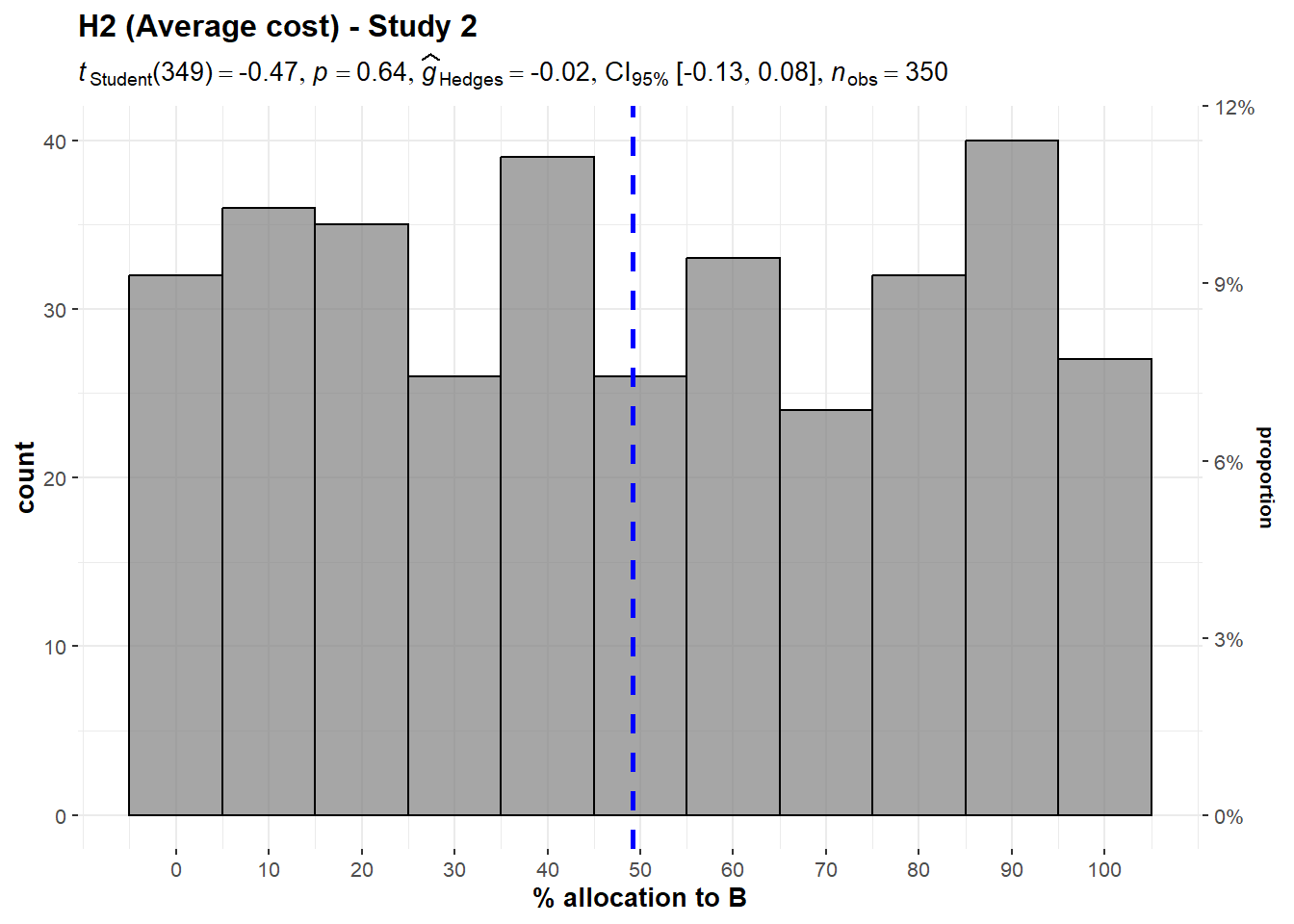
*Past costs, Study 1: Allocation*



*Note*. The scenario: “**A** and **B** help prevent deaths in children. **A** prevents 5 deaths for every $1,000 of donations, on the average, and **B** prevents 6 deaths for every $1,000. Given the donations they have received so far, and the opportunities for expansion, **A** will prevent 5 deaths for each **additional** $1,000 beyond its current level of spending and **B** will also prevent 5 deaths.”

**Figure 4**

*Past costs, Study 2: Allocation*



*Note*. The scenario: “**A** and **B** will each prevent 5 deaths for every $10,000 of new donations. **A** was much more expensive to get started. Thus, the cost per life saved on the average is higher for **A**, because **A** has spent more money in total.”

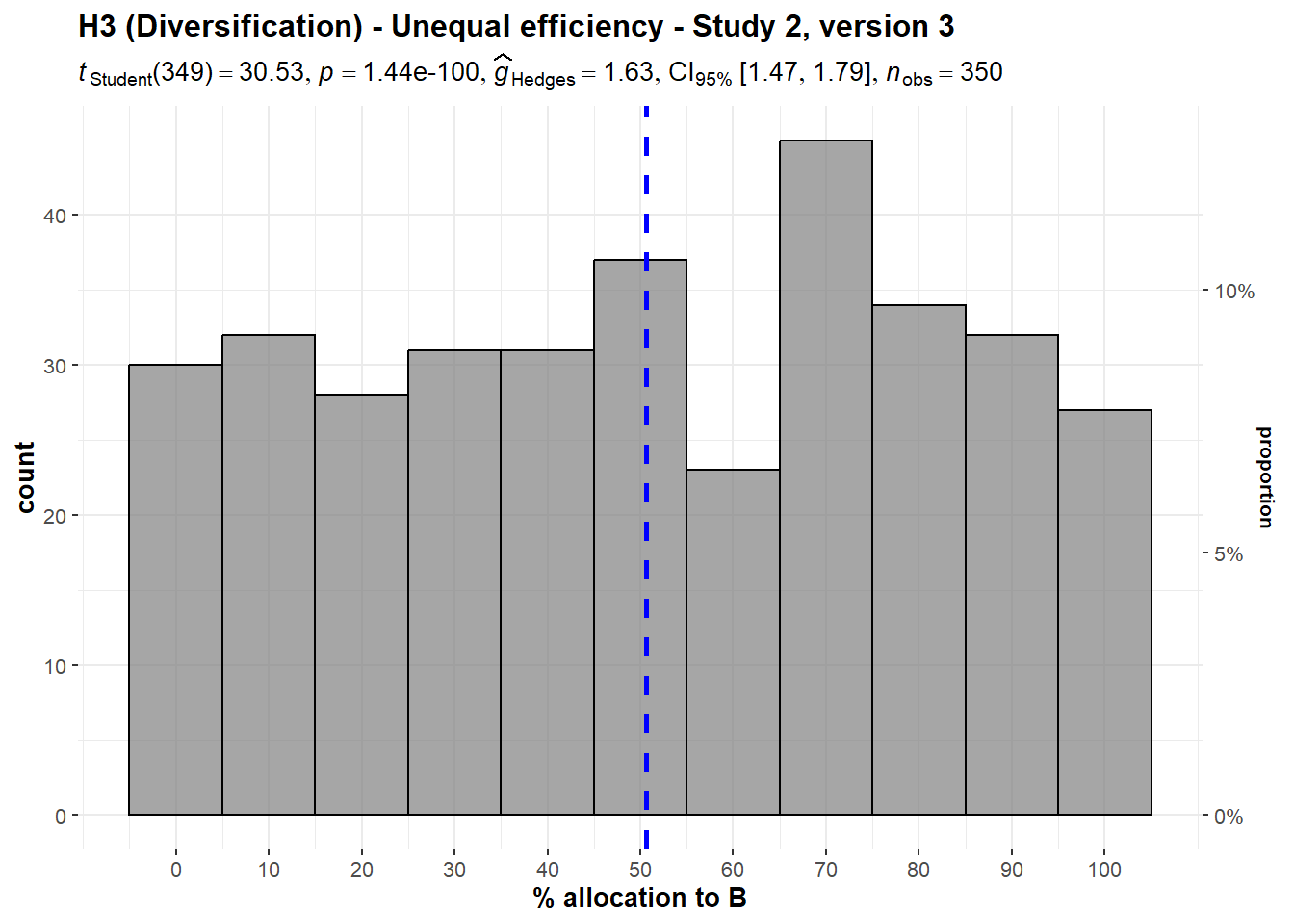
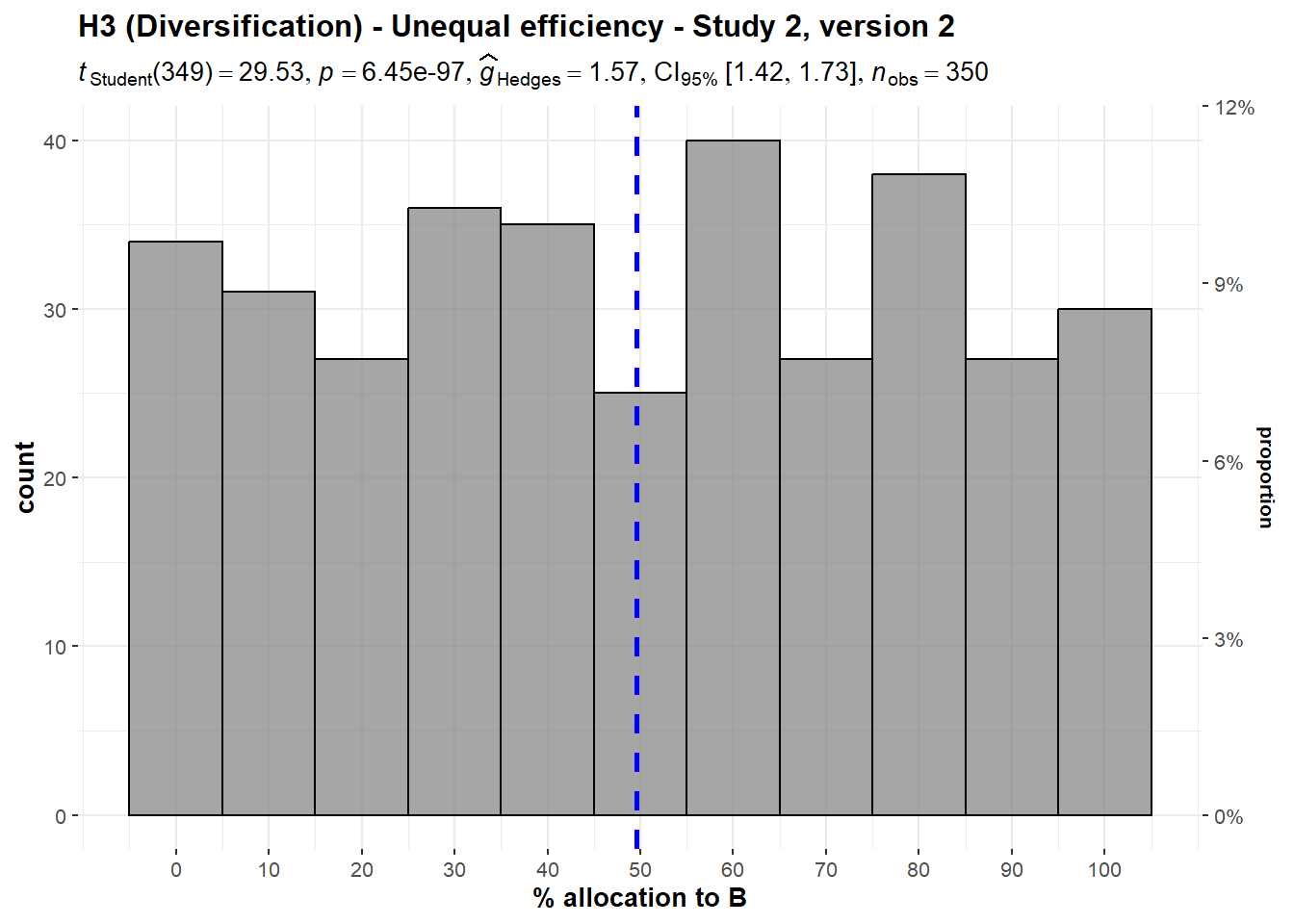
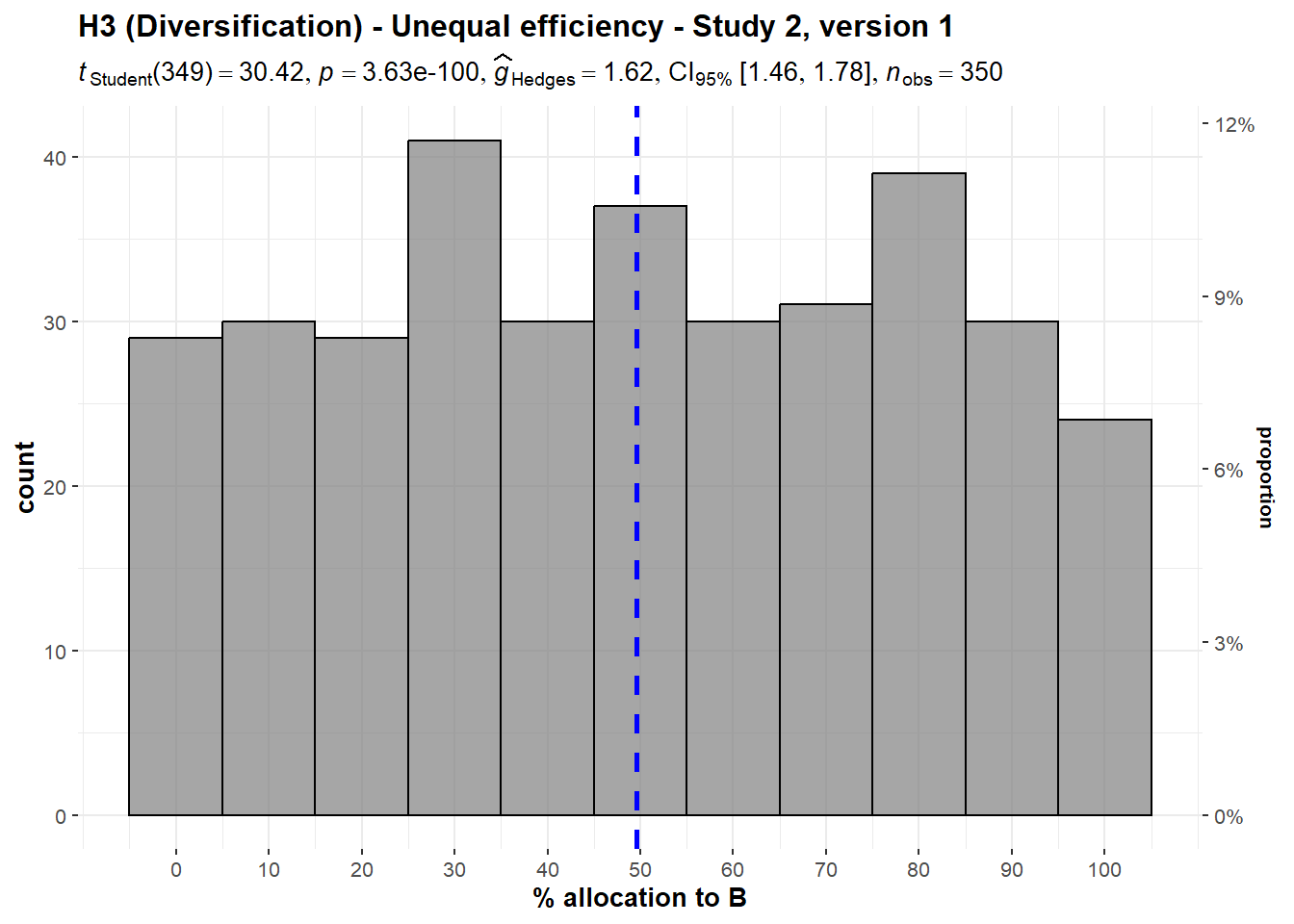
### 3. Diversification effect (Hypothesis 3)

#### Unequal efficiency

We conducted three one-sample t-tests for Study 2, versions 1 to 3, which we summarized and plotted in Figure 5. We found support for the hypothesis that participants would diversify their donations even at the cost of inefficiency, in version 1 (*M* = 49.57, *SD* = 30.49, *t*349 = 30.4, *p* < .001, *d* = 1.63, 95% CI [1.47, 1.79]), version 2 (*M* = 49.54, *SD* = 31.39, *t*349 = 29.5, *p* < .001, *d* = 1.58, 95% CI [1.42, 1.73]), and version 3 (*M* = 50.66, *SD* = 31.04, *t*349 = 30.5, *p* < .001, *d* = 1.63, 95% CI [1.47, 1.79]) (compared to the most efficient allocation, which is allocating 0% to the less efficient charity).

**Figure 5**

*Diversification with unequal efficiency, Study 2: Allocation*



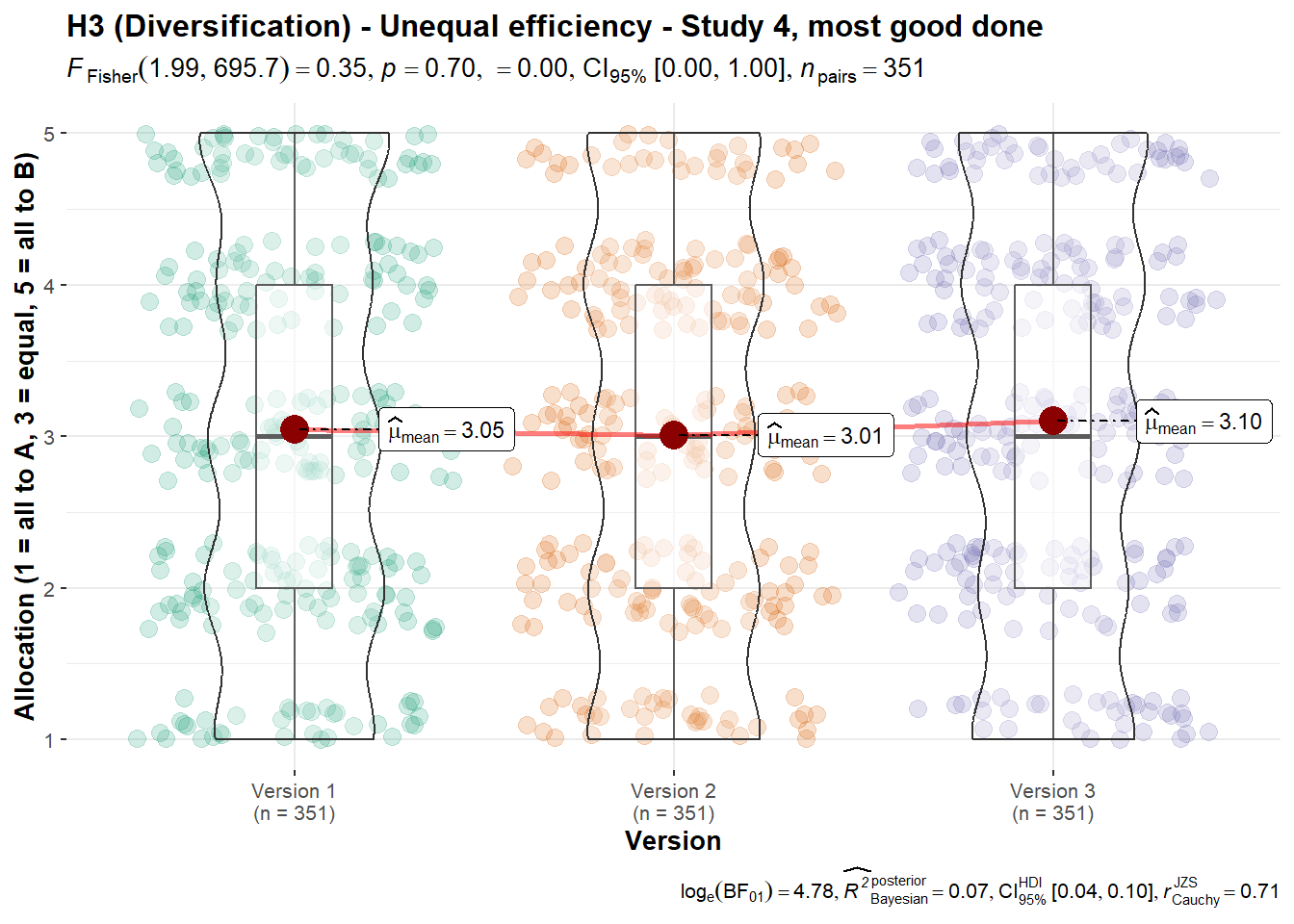
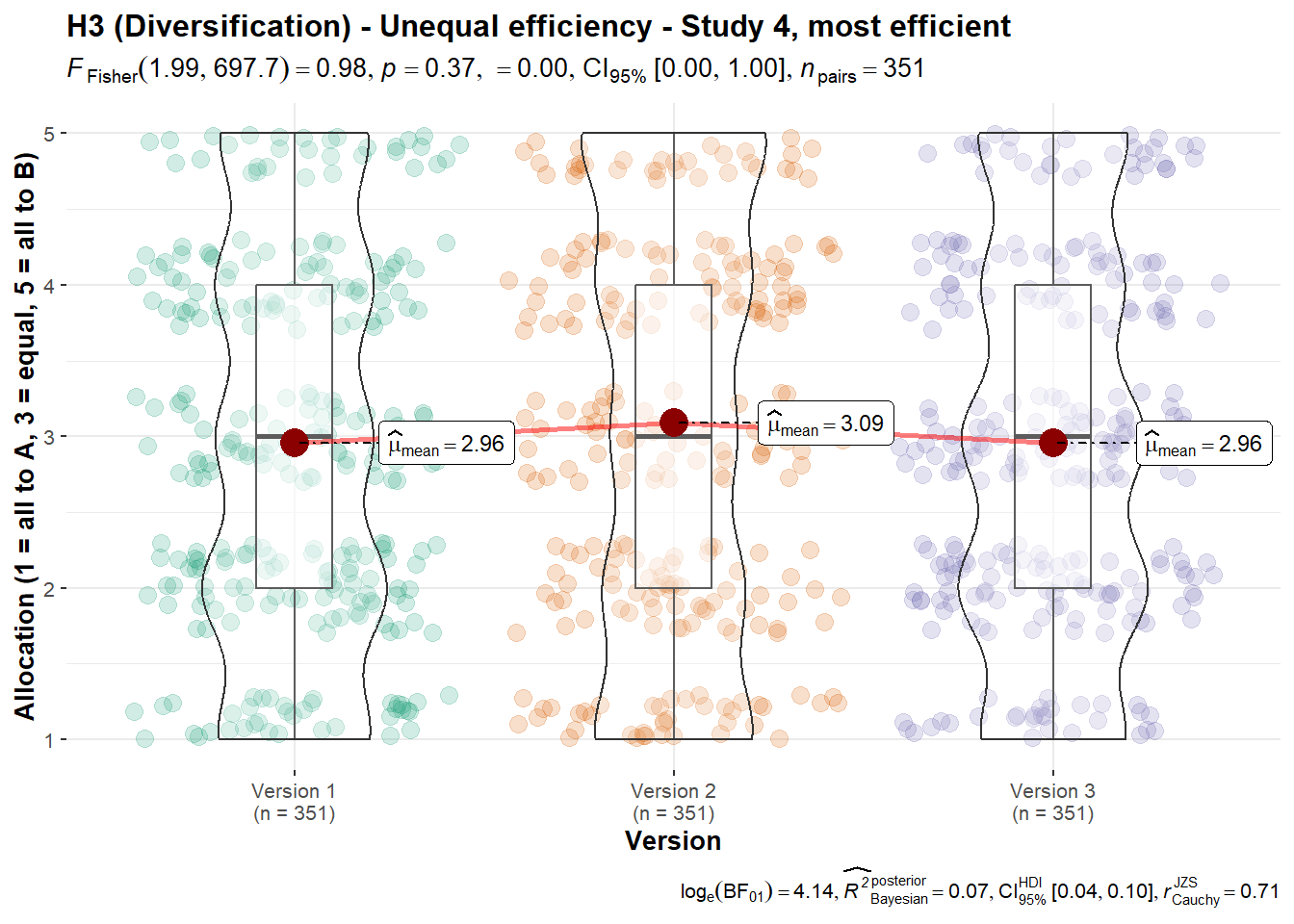
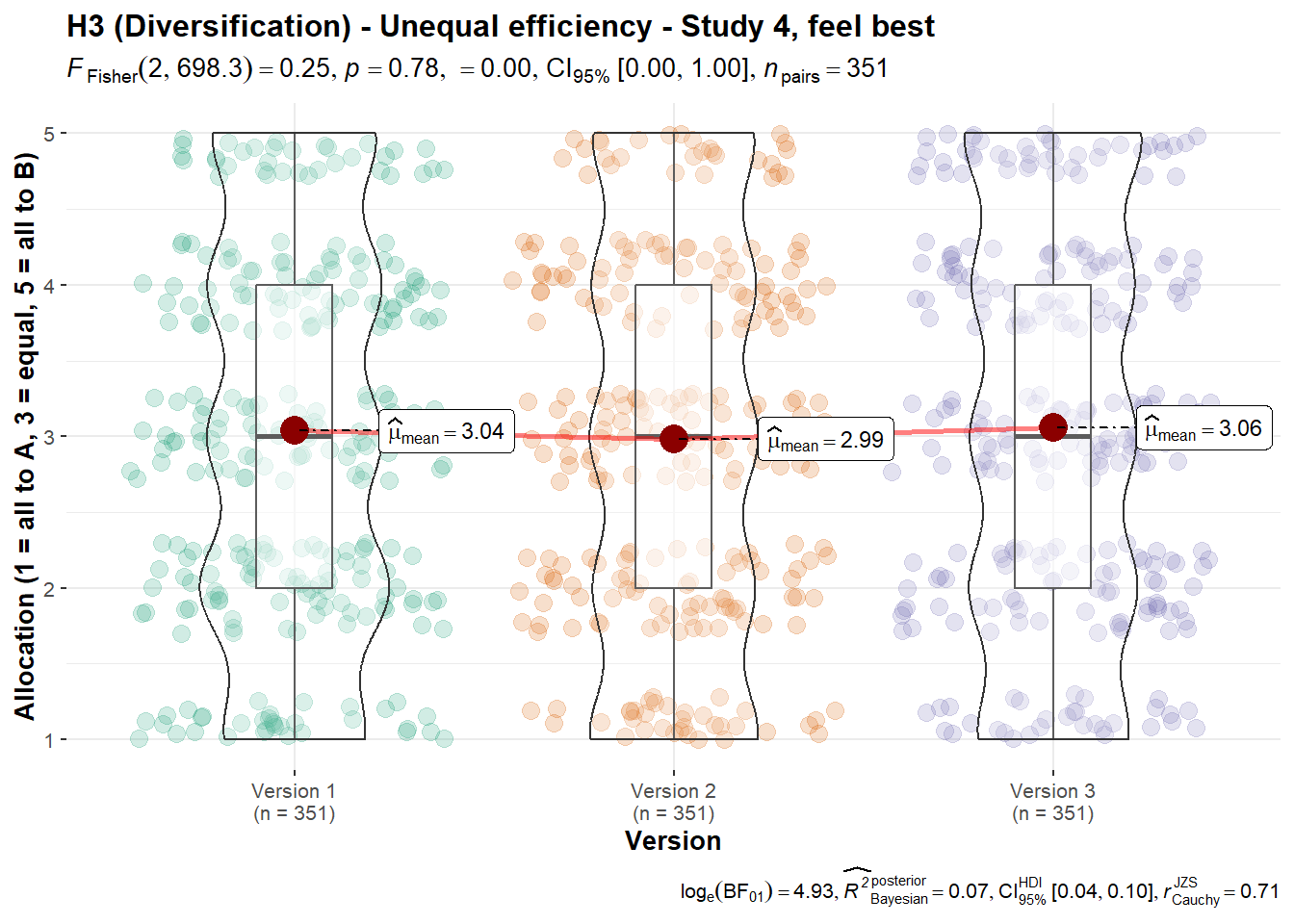
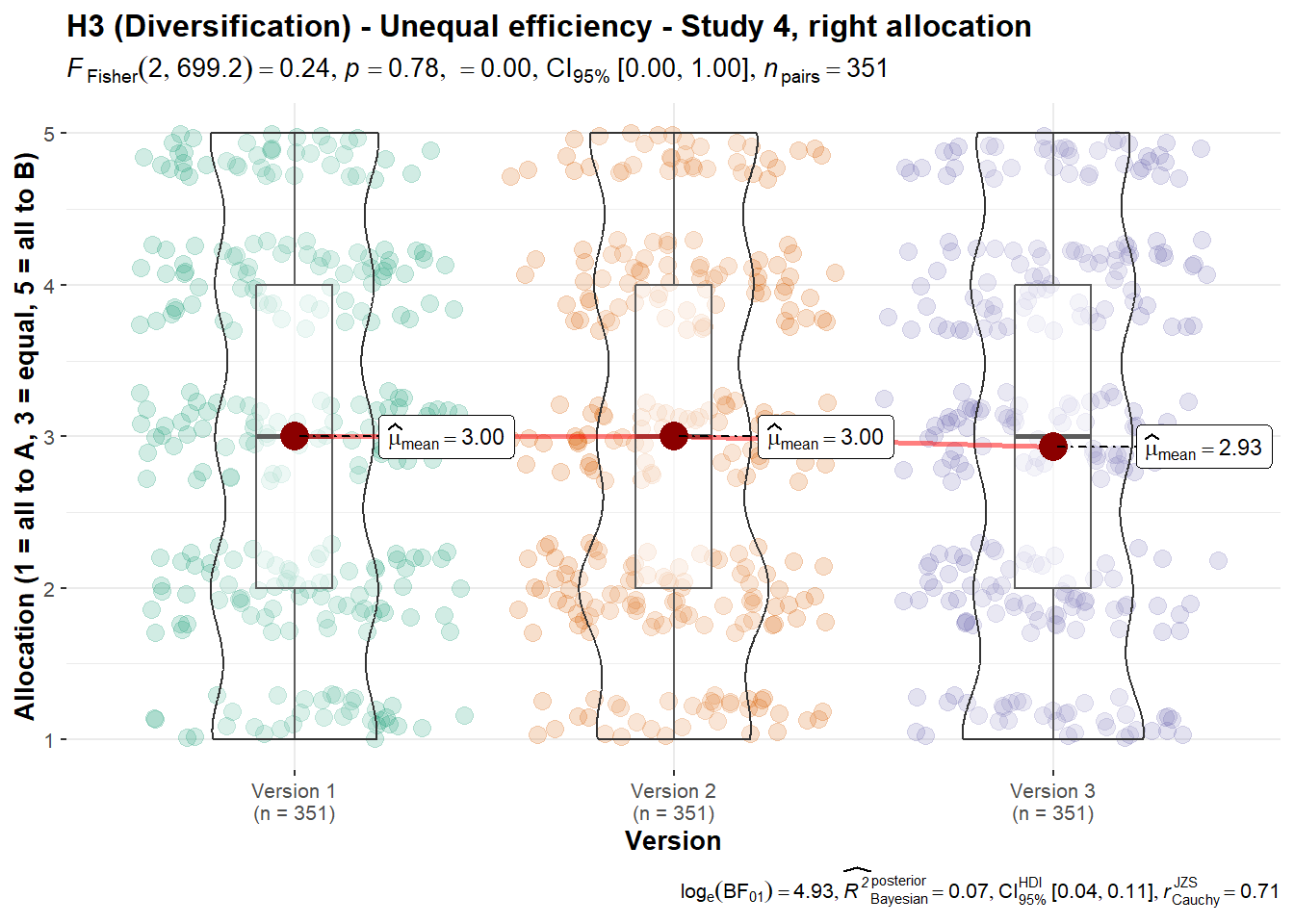
*Note*. Study 2, version 1: **A** can save one life for $10,000. **B** can save one life for $12,500. The people helped are from the same groups, with the same problems.  
Study 2, version 2: **A** can save 5 lives for $50,000. **B** can save 4 lives for $50,000. The people helped are from the same groups, with the same problems.  
Study 2, version 3: **A** and **B** are both involved in preventing death in people with AIDS. **A** uses a method with a 75% chance of success over 5 years. **B** uses a method with a 50% chance of success over 5 years, with the same patients.

We also conducted four one-sample t-tests by collating the responses to versions 1, 2, and 3 in Study 4. We found that participants allocated more than nothing to the less effective charity when asked for the right allocation (*M* = 2.98, *SD* = 1.43, *t*1052 = 44.9, *p* < .001, *d* = 1.38, 95% CI [1.30, 1.47]), the allocation they would feel best about making (*M* = 3.03, *SD* = 1.41, *t*1052 = 46.8, *p* < .001, *d* = 1.44, 95% CI [1.36, 1.53]), the most efficient allocation (*M* = 3.00, *SD* = 1.42, *t*1052 = 45.8, *p* < .001, *d* = 1.41, 95% CI [1.33, 1.50]), and the allocation that would do the most good for each $1,000 spent (*M* = 3.05, *SD* = 1.43, *t*1052 = 46.5, *p* < .001, *d* = 1.43, 95% CI [1.35, 1.52]) (all compared against a lowest point of 1, with 1 being all to the more effective charity and 5 being all to the less effective charity).

We then conducted four one-way repeated measures ANOVA tests and found no support for differences between the versions when asked for the right allocation (*F*(2, 699.2) = 0.242, *p* = .785, partial ω² = 0.00), the allocation they would feel best about making (*F*(2, 698.3) = 0.247, *p* = .781, partial ω² = 0.00), and the most efficient allocation (*F*(1.99, 697.7) = 0.984, *p* = .374, partial ω² = 0.00), as well as the allocation that would do the most good for each $1,000 spent (*F*(1.99, 695.7) = 0.352, *p* = .702, partial ω² = 0.00). We summarized and plotted all analyses in Figure 6.

**Figure 6**

*Diversification with unequal efficiency, Study 4: Allocation*

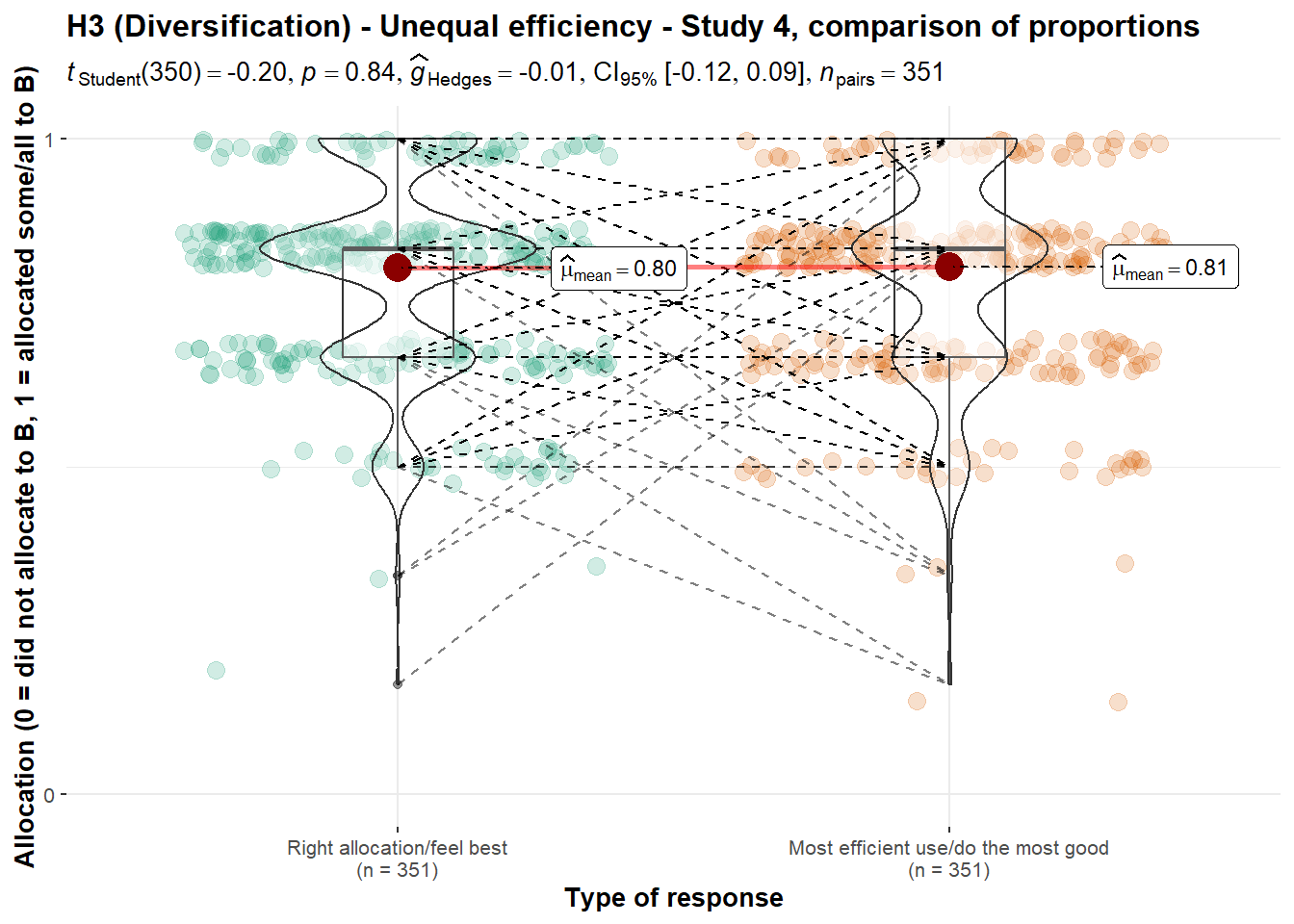
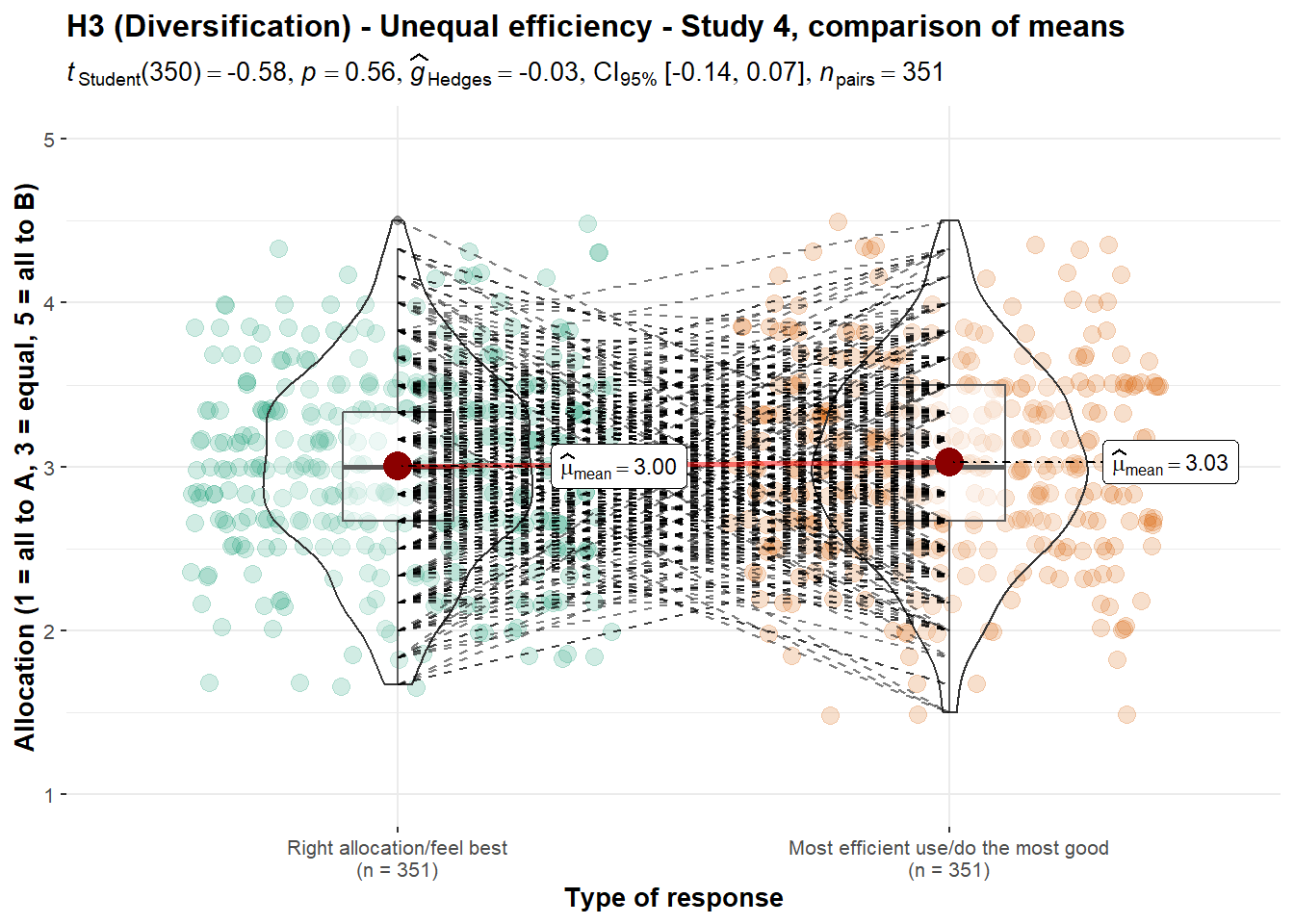


*Note*. Scale is ordinal; jitter was added for visualization.  
Scenarios:  
Study 4, version 1: “**A** and **B** are both involved in preventing death in people with AIDS. **A** uses a method with a 75% chance of success over 5 years. **B** uses a method with a 50% chance of success over 5 years, with the same patients”.  
Study 4, version 2: “**A** can save one life for $10,000. **B** can save one life for $12,500. The people helped are from the same groups, with the same problems”.  
Study 4, version 3: “**A** can save 5 lives for $50,000. **B** can save 4 lives for $50,000. The people helped are from the same groups, with the same problems.”

We then conducted a paired t-test, and found no support for differences between the average response per participant to the “right allocation” and “allocation that feels best” questions combined (*M* = 3.00, *SD* = 0.57) and the average response per participant to the “most efficient allocation” and “allocation that does the most good” questions combined (*M* = 3.03, *SD* = 0.58; *t*350 = -0.58, *p* = .562, *d* = -0.03, 95% CI [-0.14, 0.07]). We also found no support for differences between the proportion of responses that allocated something to the less efficient charity when asked for “right allocation” and the “allocation that feels best” (*M* = 0.8, *SD* = 0.16) to the proportion of responses that allocated something to the less efficient charity when asked for the “most efficient allocation” and the “allocation that does the most good” (*M* = 0.81, *SD* = 0.16; *t*350 = -0.20, *p* = .840, *d* = -0.01, 95% CI [-0.12, 0.09]). Both analyses are summarized and plotted in Figure 7.

**Figure 7**

*Diversification in unequal efficiency, Study 4: Comparison of allocation*



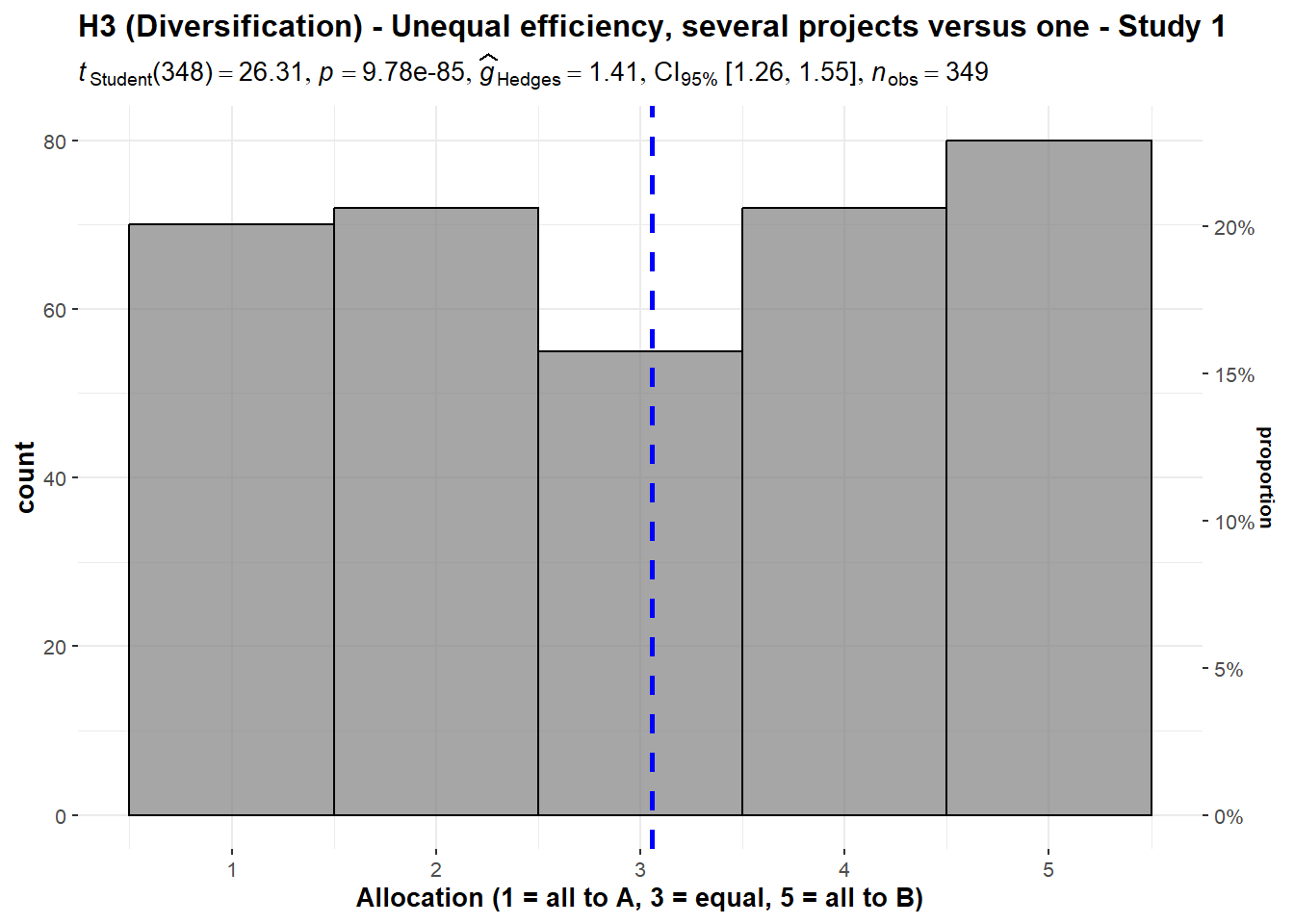
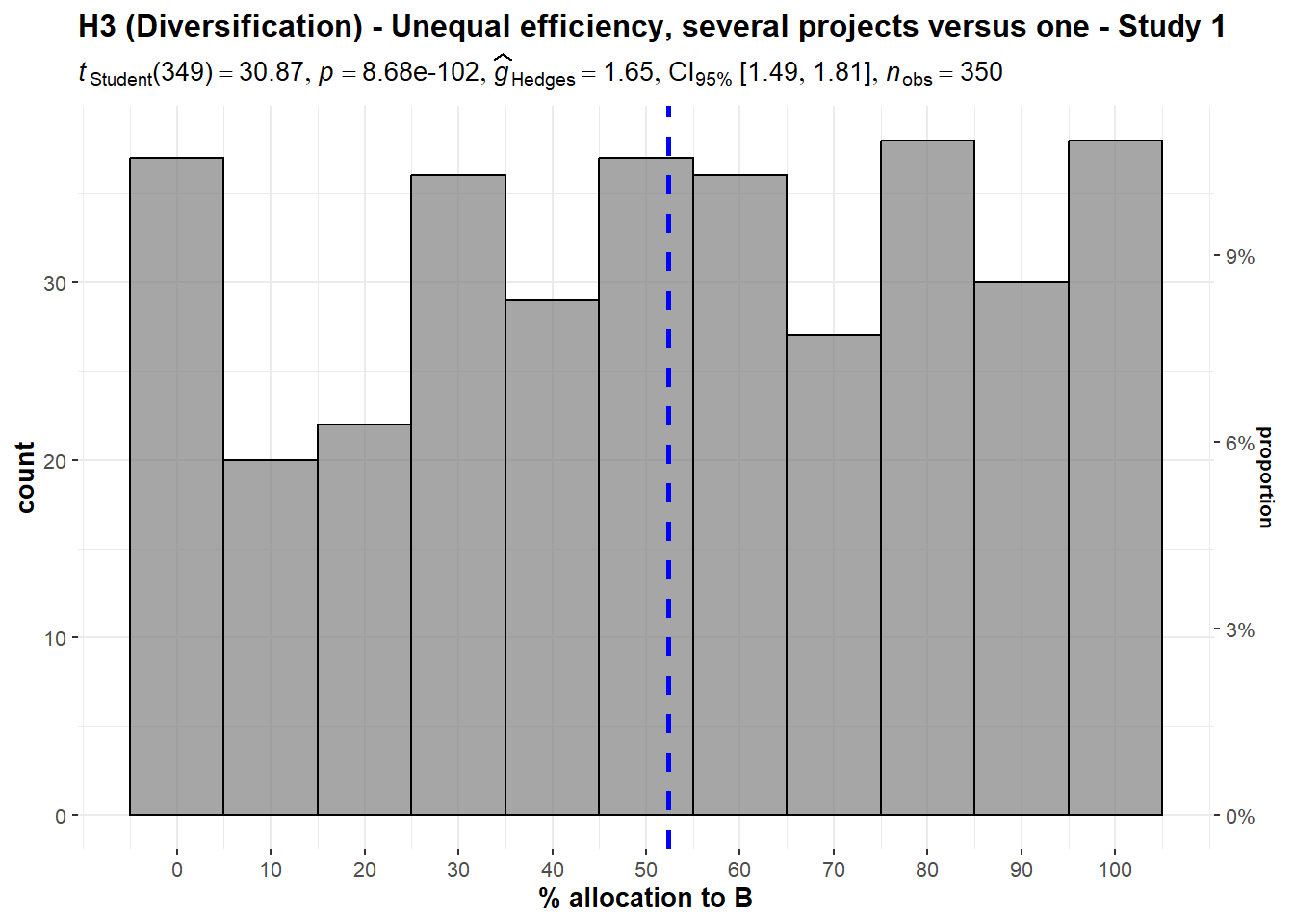
*Note*. Scale is ordinal; jitter was added for visualization purposes.

#### Unequal efficiency, several projects versus one

We conducted two one-sample t-tests for Studies 1 and 3, which we summarized and plotted in Figure 8. We found support for higher than nothing allocation to the less effective charity in Study 1 (against a lowest point of 0; *M* = 52.34, *SD* = 31.73, *t*349 = 30.9, *p* < .001, *d* = 1.65, 95% CI [1.49, 1.81]), and in Study 3 (against a lowest point of 1; *M* = 3.06, *SD* = 1.46, *t*348 = 26.3, *p* < .001, *d* = 1.41, 95% CI [1.26, 1.56]).

**Figure 8**

*Diversification in unequal efficiency, Studies 1 and 3: Allocation (several projects versus one)*



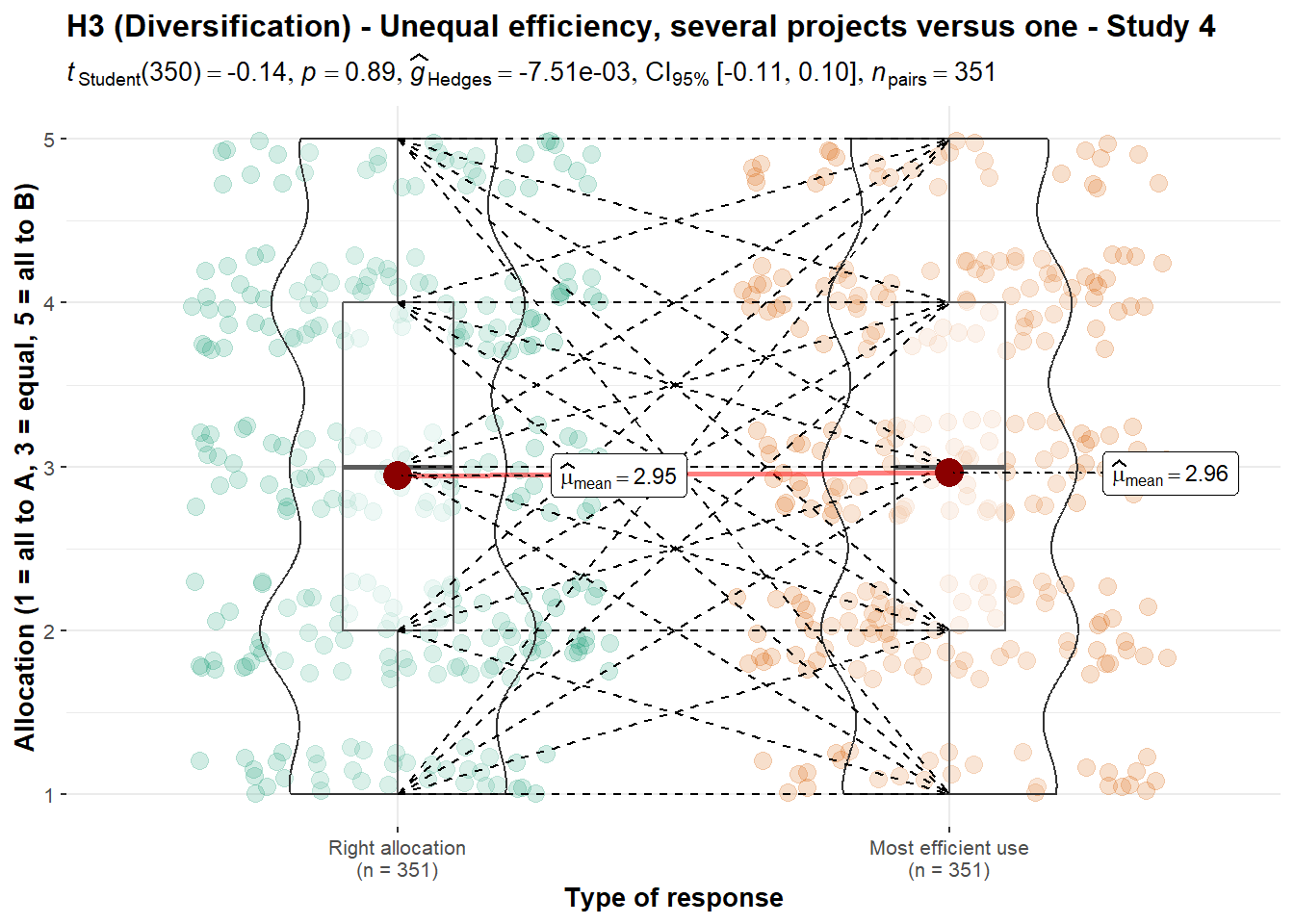
*Note*. Scenarios:  
Study 1: “**A** puts all the money into one project, which has a 75% chance of helping many children, and a 25% chance of doing no good at all. **B** puts the money into several different projects, each of which has a 70% chance of helping some children, but a 30% chance of doing no good”.  
Study 3: “**A** puts $1,000,000 into one project, which has a 75% chance of helping 10,000 children, and a 25% chance of doing no good. **B** puts $200,000 into each of 5 projects ($1,000,000 total). Each of the 5 has a **70%** chance of helping 2,000 children and a 30% chance of doing no good. (If all 5 succeed, then the total benefit is 10,000 children, the same as **A**.)”

We also conducted two one-sample t-tests for Study 4. We found support for higher than nothing allocation to the less effective charity when asked for the right allocation (*M* = 2.95, *SD* = 1.38, *t*350 = 26.5, *p* < .001, *d* = 1.41, 95% CI [1.26, 1.56]) and the most efficient allocation (*M* = 2.96, *SD* = 1.36, *t*350 = 27.0, *p* < .001, *d* = 1.44, 95% CI [1.29, 1.59]) (both against a lowest point of 1, with 1 being all to the more effective charity and 5 being all to the less effective charity).

We conducted a paired t-test and found no support for differences between the responses to the two questions (*t*350 = -0.14, *p* = .888, *d* = 0.00, 95% CI [-0.11, 0.10]; summarized and plotted in Figure 9).

**Figure 9**

*Diversification in unequal efficiency, Study 4: Comparison of allocation (several projects versus one)*

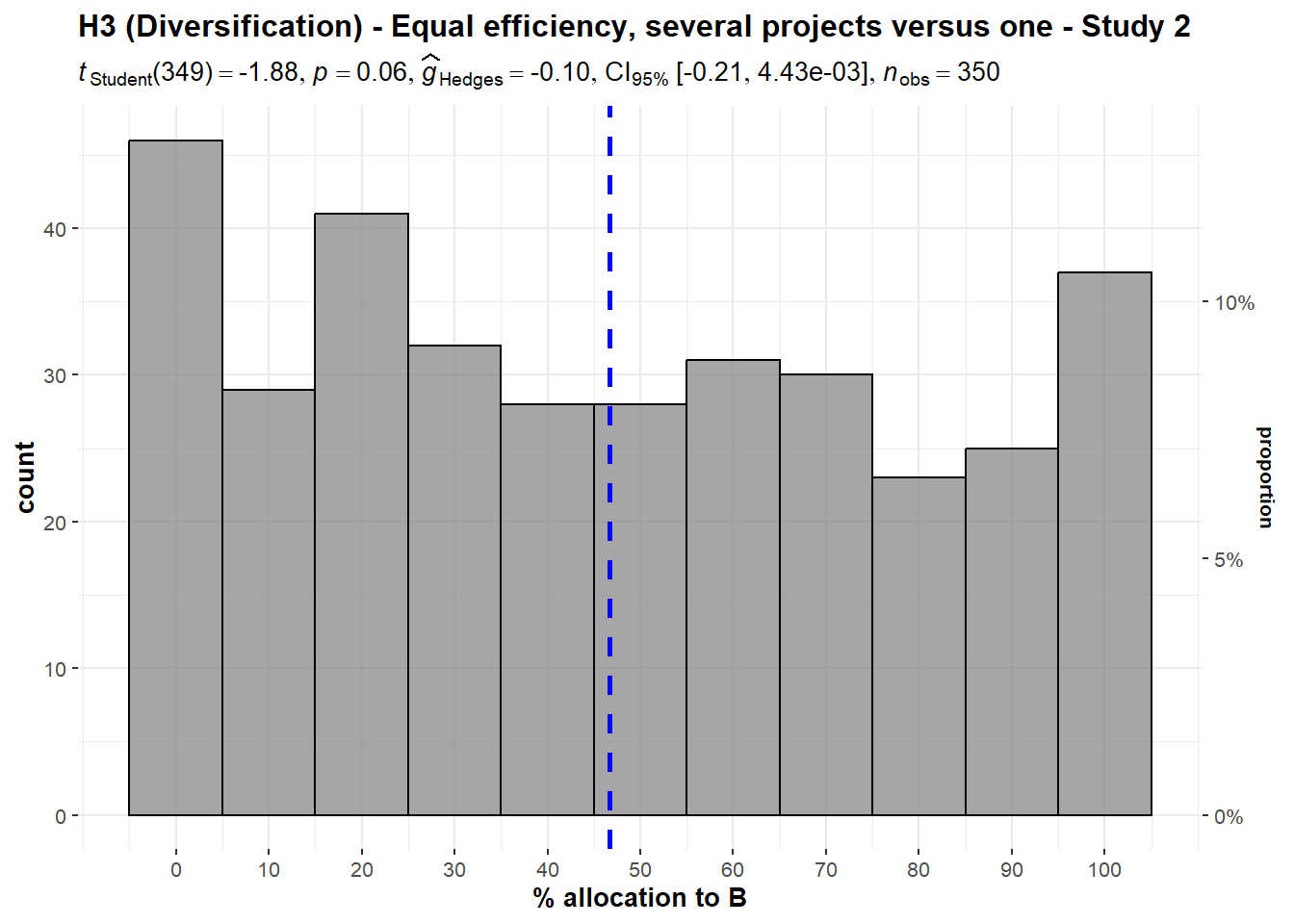
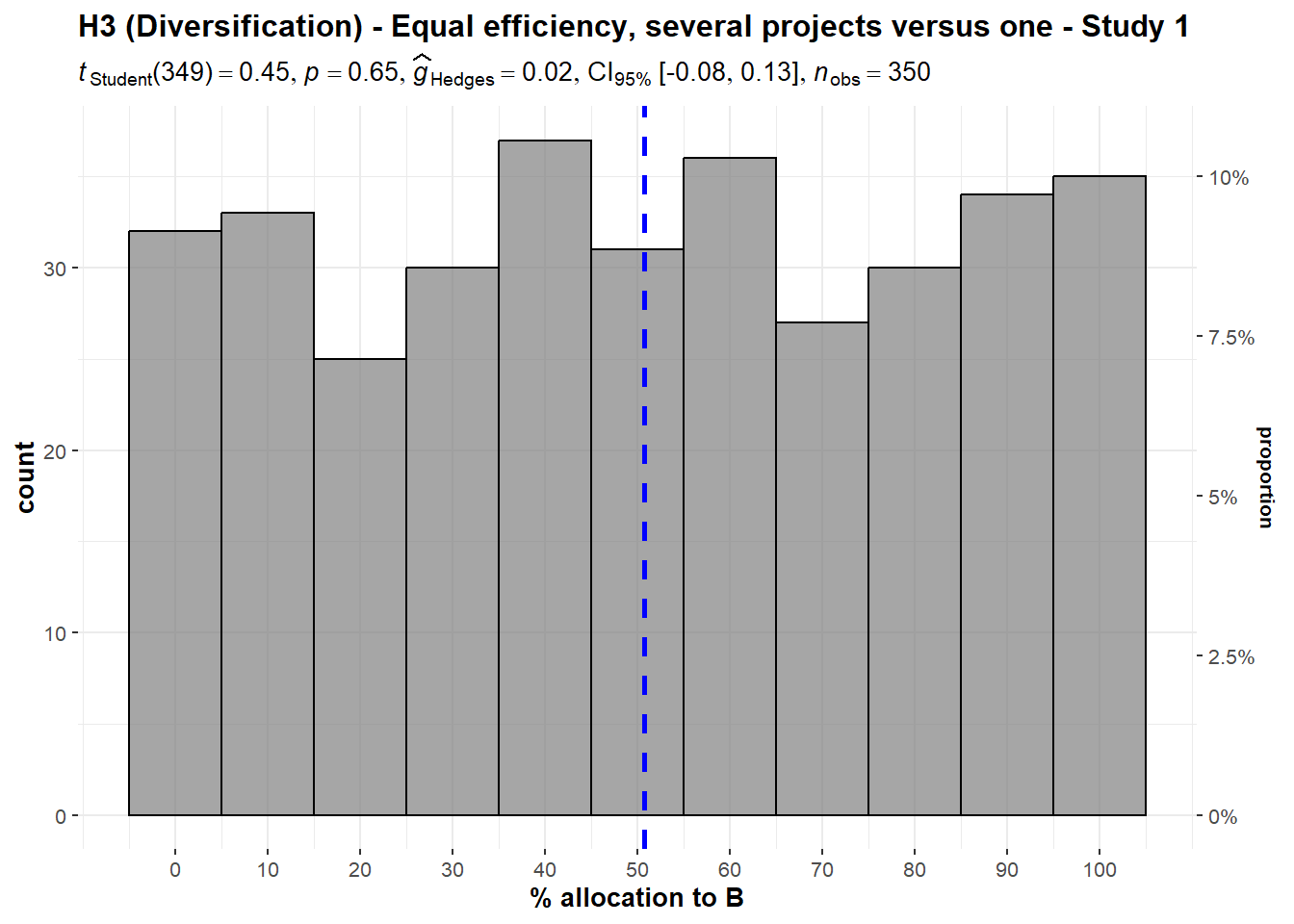


*Note*. Scale is ordinal; jitter was added for visualization purposes.  
Scenario: “**A** puts $1,000,000 into one project, which has a 75% chance of helping 10,000 children, and a 25% chance of doing no good. **B** puts $200,000 into each of 5 projects ($1,000,000 total). Each of the 5 has a **70%** chance of helping 2,000 children and a 30% chance of doing no good. (If all 5 succeed, then the total benefit is 10,000 children, the same as **A**.)”

#### Equal efficiency

We conducted two one-sample t-tests for Studies 1 and 2, summarized and plotted in Figure 10. We found no support for differences in allocation of money in Study 1 (*M* = 50.77, *SD* = 31.85, *t*349 = 0.45, *p* = .651, *d* = 0.02, 95% CI [-0.08, 0.13]) and Study 2 (*M* = 46.69, *SD* = 32.93, *t*349 = -1.88, *p* = .061, *d* = -0.10, 95% CI [-0.21, 0.00]), compared to an equal allocation of 50% each.

**Figure 10**

*Diversification with equal efficiency, Studies 1 and 2: Allocation* 

*Note*. Scenario: “**A** puts all the money into one project, which will help 100,000 children. **B** puts the money into five different projects, each of which will help 20,000 children. (The benefit per child will be the same.)”

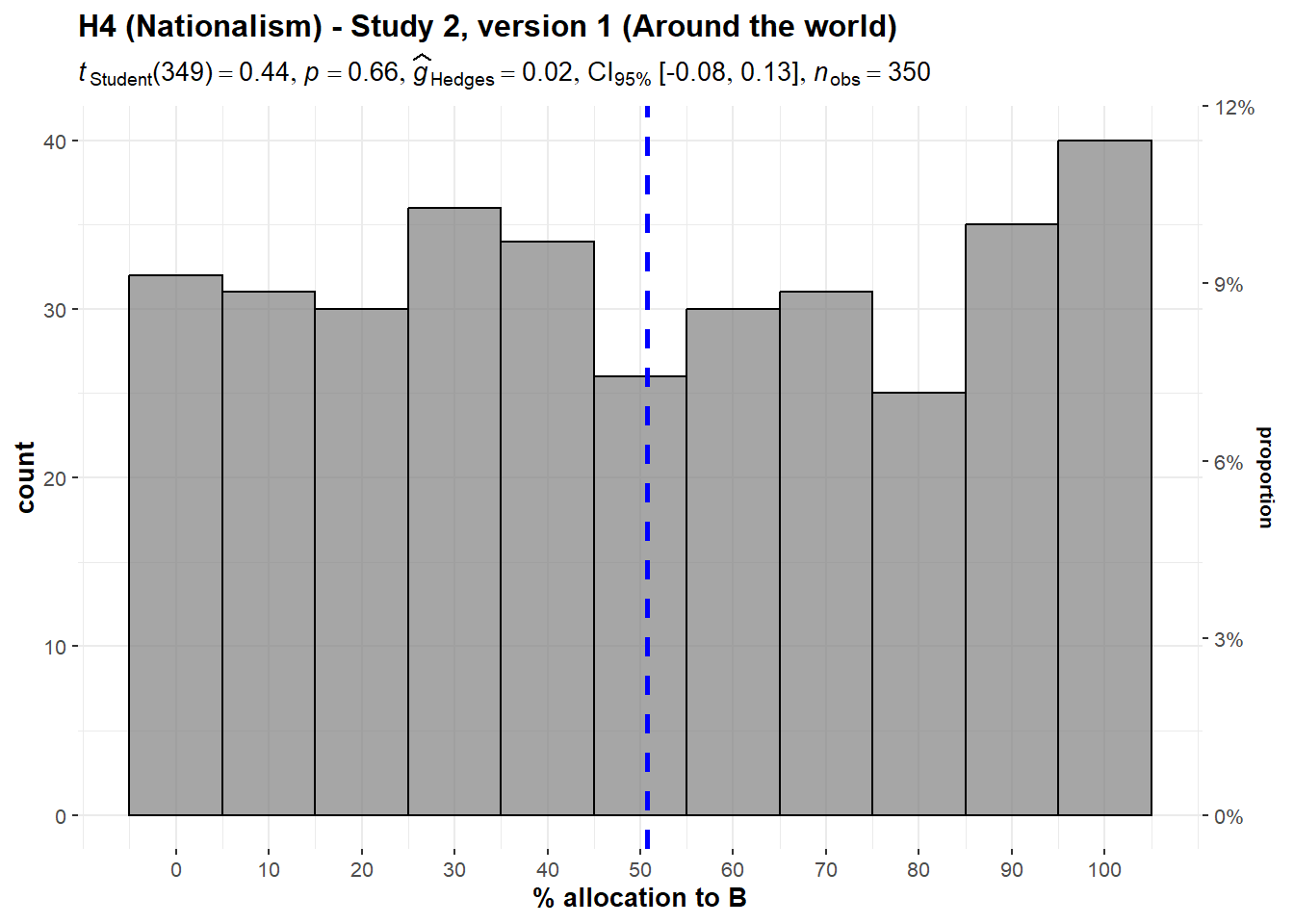
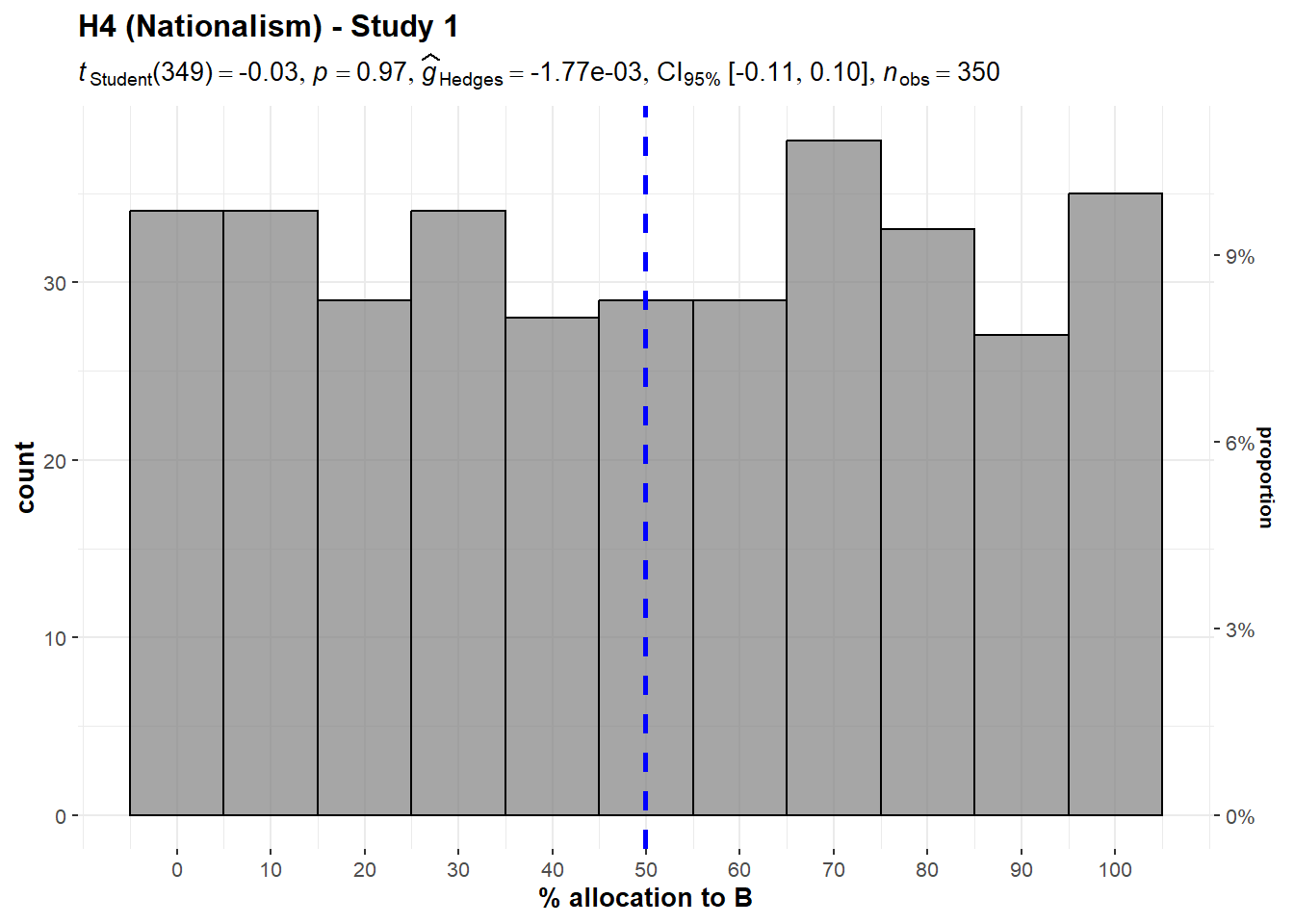
### 

### 4. Nationalism/Ingroup effect (Hypothesis 4)

We conducted two one-sample t-tests for Studies 1 and 2, summarized and plotted in Figure 11. We failed to find support for increased participant willingness to donate to charities that help children in their own country over children around the world in Study 1 (*M* = 49.94, *SD* = 32.14, *t*349 = -0.03, *p* = .974, *d* = 0.00, 95% CI [-0.11, 0.10]) and in Study 2 (*M* = 50.77, *SD* = 32.48, *t*349 = 0.44, *p* = .657, *d* = 0.02, 95% CI [-0.08, 0.13]) (both against a 50% midpoint).

**Figure 11**

*Ingroup effect, Studies 1 and 2: Allocation*



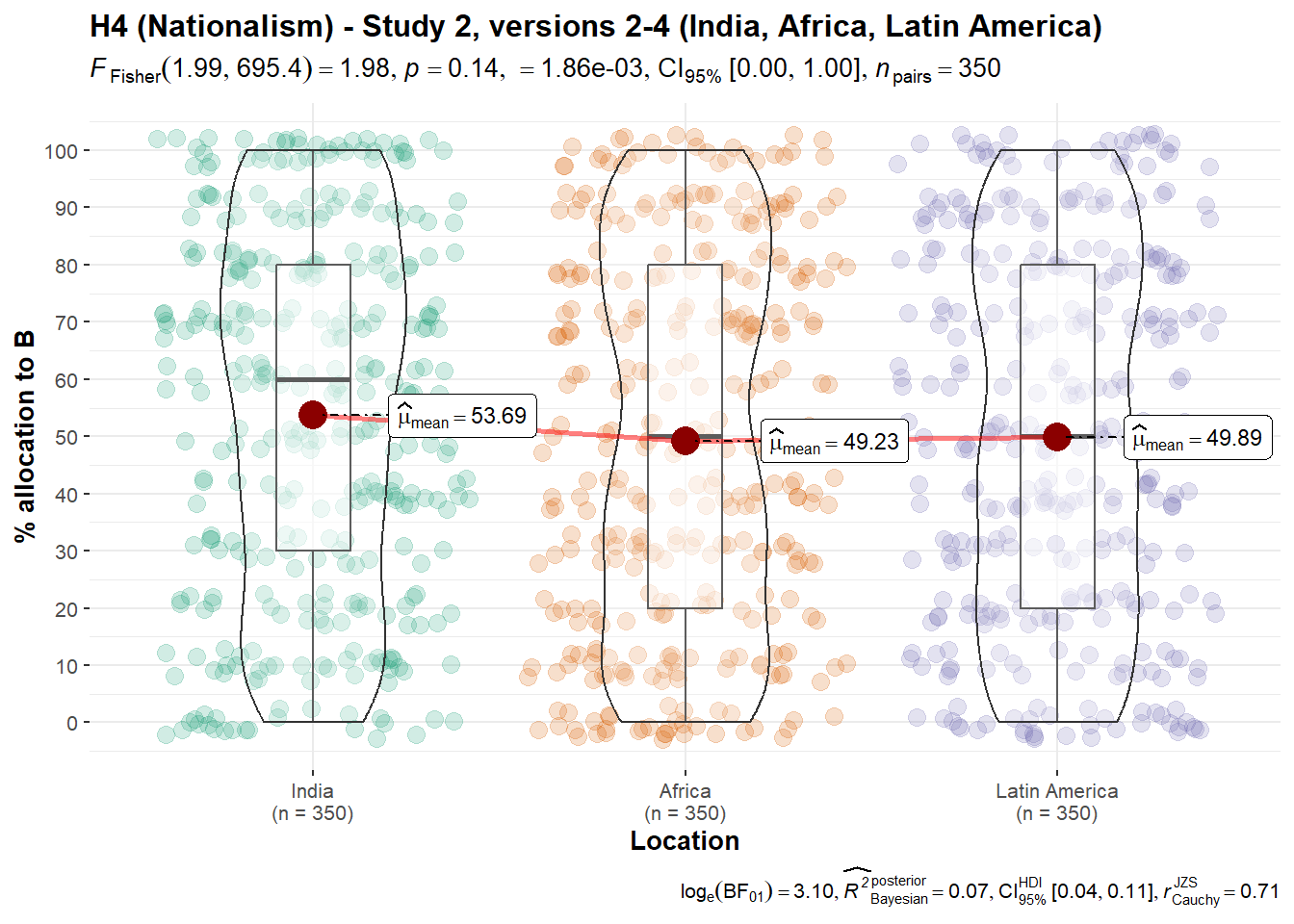
*Note*. Scenarios:  
Study 1: **A** helps children who are in your own country. **B** helps children around the world. The children are equally needy.  
Study 2, version 1: **A** helps children who are in your own country. **B** helps children around the world. The children are equally needy, and the benefits are the same for each child.

We conducted three additional one-sample t-tests for Study 2 for specific locations, and found support for increased participant willingness to donate to charities that help children in their own country over a given foreign country or region in the India condition (*M* = 53.69, *SD* = 31.52, *t*349 = 2.19, *p* = .029, *d* = 0.12, 95% CI [0.01, 0.22]), but failed to find support in the Africa (*M* = 49.23, *SD* = 32.49, *t*349 = -0.44, *p* = .657, *d* = -0.02, 95% CI [-0.13, 0.08]) and Latin America conditions (*M* = 49.89, *SD* = 31.96, *t*349 = -0.07, *p* = .947 *d* = 0.00, 95% CI [-0.11, 0.10]) (all against a 50% mid-point).

We then conducted a one-way repeated measures ANOVA test and found no support for differences between the three location conditions (*F*(1.99, 695.4) = 1.98, *p* = .139, partial ω² = 0.00; summarized and plotted in Figure 12).

**Figure 12**

*Ingroup effect in Study 2, versions 2-4: Comparison of allocation*



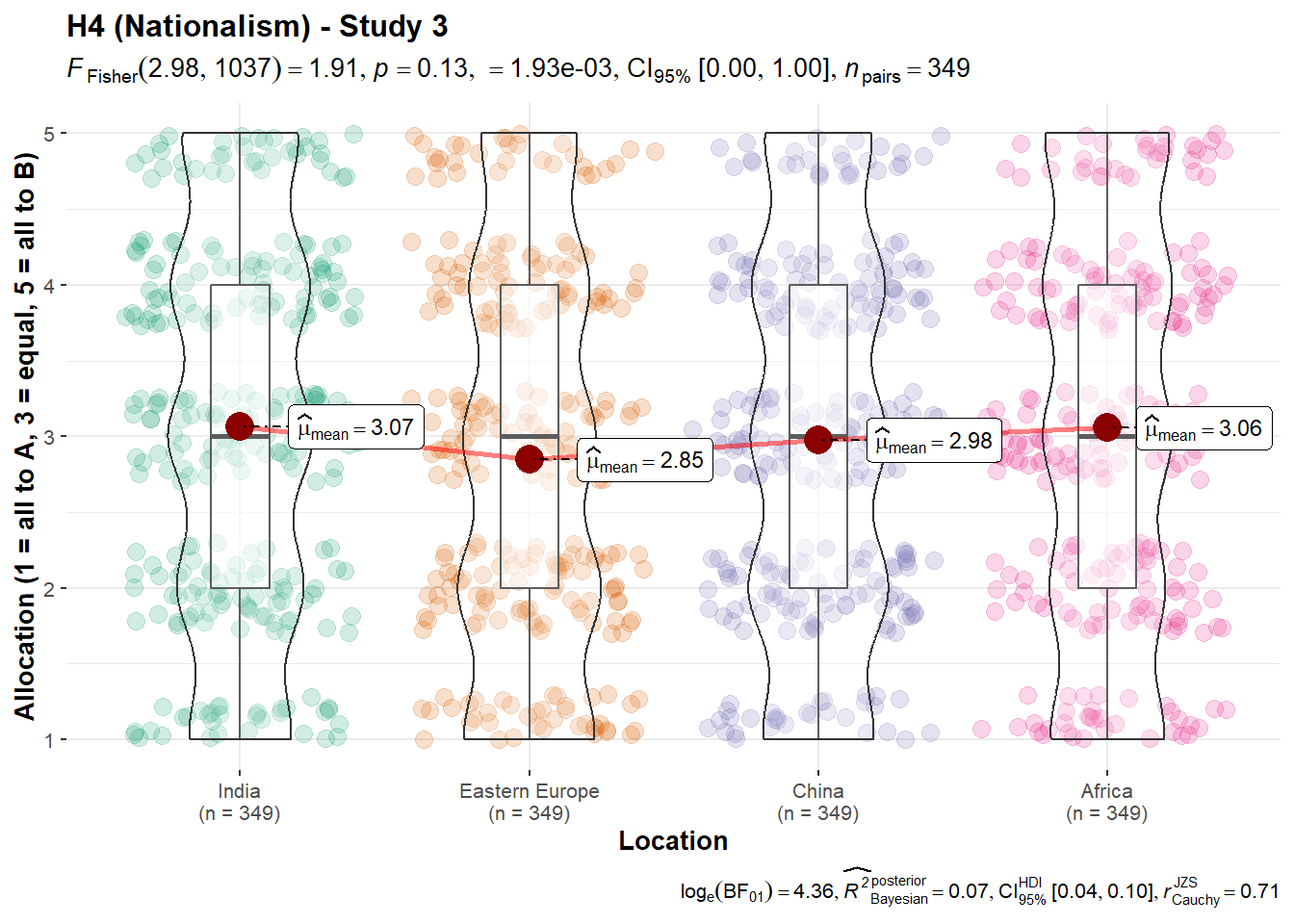
*Note*. Scale is ordinal; jitter was added for visualization purposes.  
Scenarios: “**A** helps children who are in your own country. **B** helps children in [Study 2, version 2: India; version 3: Africa; version 4: Latin America]. The children are equally needy, and the benefits are the same for each child.”

We conducted four one-sample t-tests for Study 3. We failed to find support for increased participant willingness to donate to charities that help children in their own country over a given foreign country or region in the Eastern Europe conditions (*M* = 2.85, *SD* = 1.37, *t*348 = -2.07, *p* = .039, *d* = -0.11, 95% CI [-0.22, -0.01]), but failed to find support in the India (*M* = 3.07, *SD* = 1.35, *t*348 = 0.91, *p* = .362, *d* = 0.05, 95% CI [-0.06, 0.15]), China (*M* = 2.98, *SD* = 1.36, *t*348 = -0.32, *p* = .752, *d* = -0.02, 95% CI [-0.12, 0.09]), and Africa (*M* = 3.06, *SD* = 1.41, *t*348 = 0.76, *p* = .447, *d* = 0.04, 95% CI [-0.06, 0.15]) (all against a mid-point of 3, summarized and plotted in Figure 13).

We also conducted a one-way repeated measures ANOVA test comparing the four conditions and found no support for differences between the four location conditions (*F*(2.98, 1037) = 1.91, *p* = .127, partial ω² = 0.00).

**Figure 13**

*Ingroup effect in Study 3, versions 1-4: Comparison of allocation*

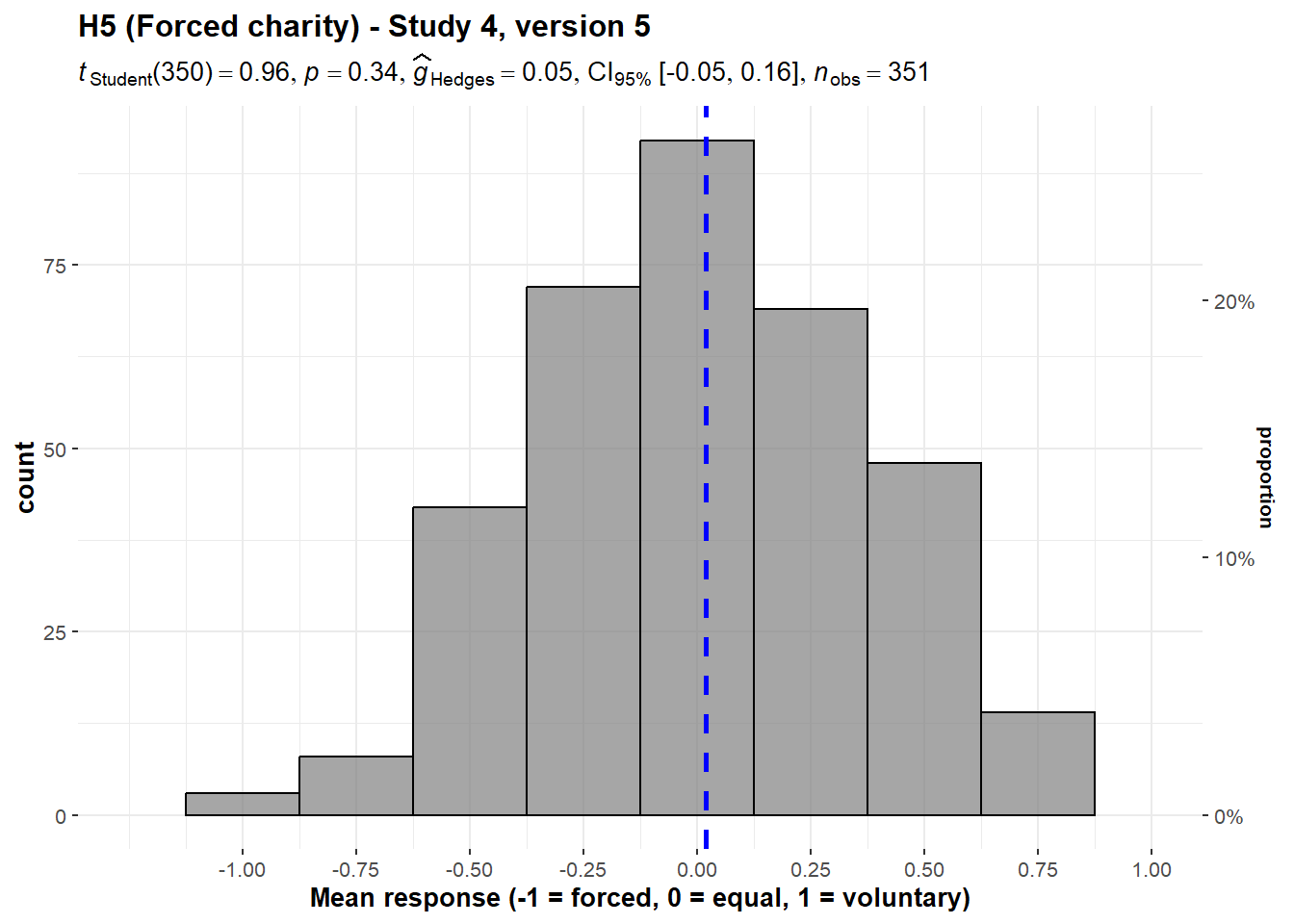
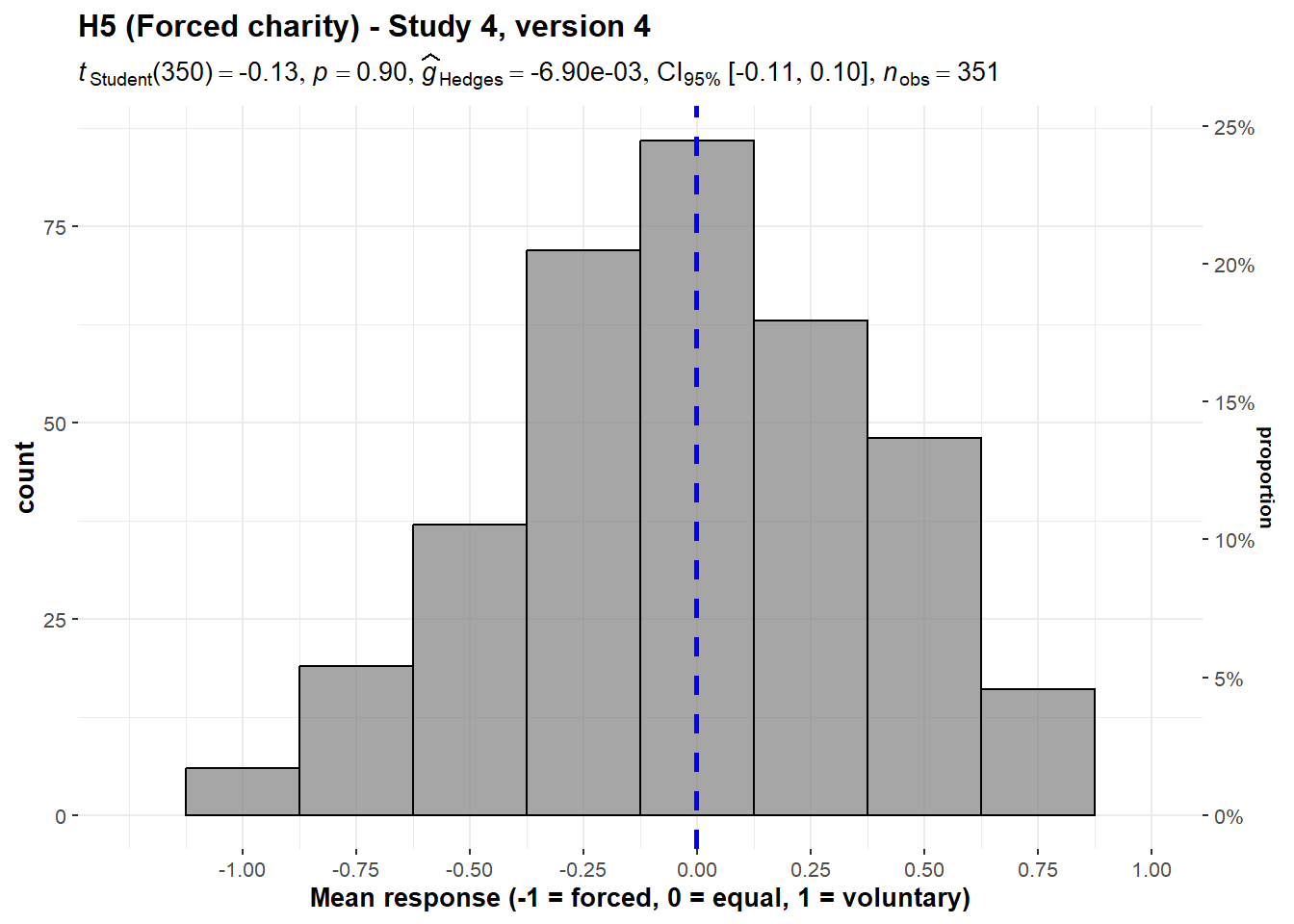
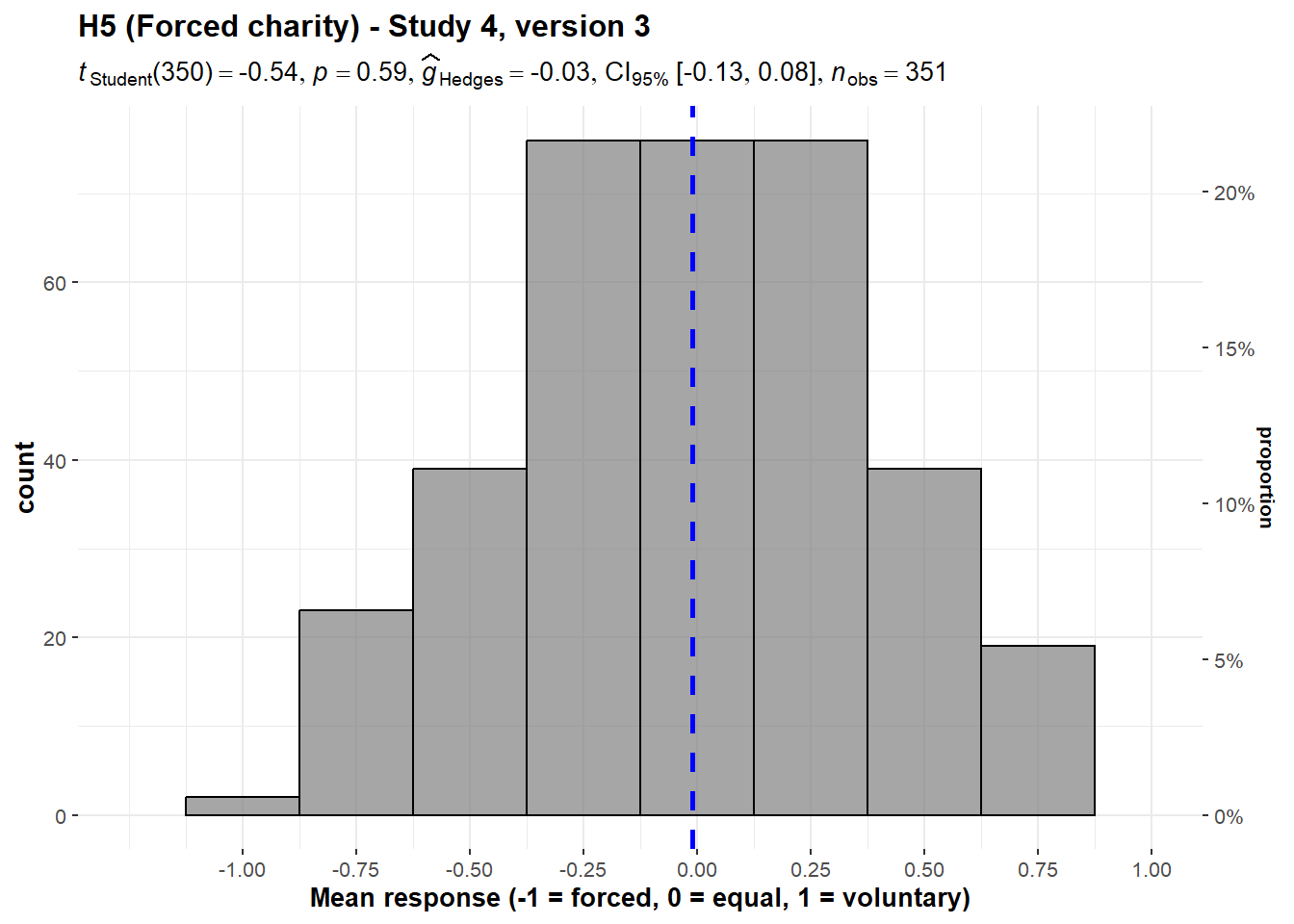
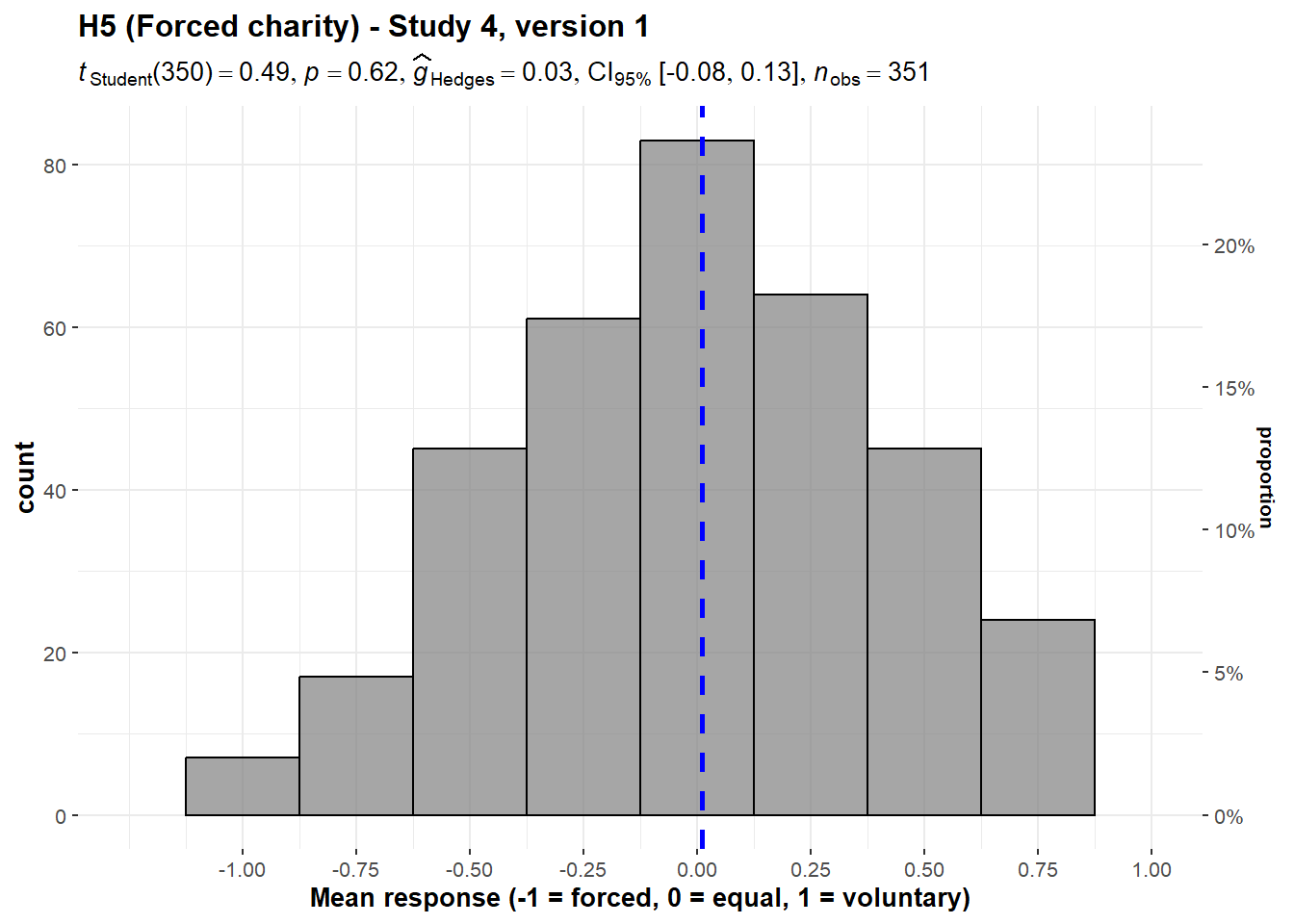


*Note*. Scale is ordinal; jitter was added for visualization purposes.  
Scenario: Study 3: **A** cures a disease in children who are in a distant part of your own country. **B** cures the same disease in children in [version 1: India; version 2: Eastern Europe; version 3: China; version 4: Africa]. **A** and **B** are equally efficient. You do not know any of the affected children, or any children who have had this disease.”

### 5. Forced-charity/Government-taxes effect (Hypothesis 5)

We conducted three one-sample t-tests for Study 5, versions 1 through 5 using the average of the four questions asked per version, which we summarized and plotted in Figure 14. We failed to find support for the hypothesis that participants would be biased against forced charity version 1 (*M* = 0.01, *SD* = 0.43, *t*350 = 0.49, *p* = .622, *d* = 0.03, 95% CI [-0.08, 0.13]), version 2 (*M* = -0.03, *SD* = 0.41, *t*350 = -1.19, *p* = .234, *d* = -0.06, 95% CI [-0.17, 0.04]), version 3 (*M* = -0.01, *SD* = 0.40, *t*350 = -0.54, *p* = .593, *d* = -0.03, 95% CI [-0.13, 0.08]), version 4 (*M* = 0.00, *SD* = 0.41, *t*350 = -0.13, *p* = .897, *d* = 0.00, 95% CI [-0.11, 0.10]), and version 5 (*M* = 0.02, *SD* = 0.38, *t*350 = 0.96, *p* = .339, *d* = 0.05, 95% CI [-0.05, 0.16]) (all against an equal response of 0).

**Figure 14**

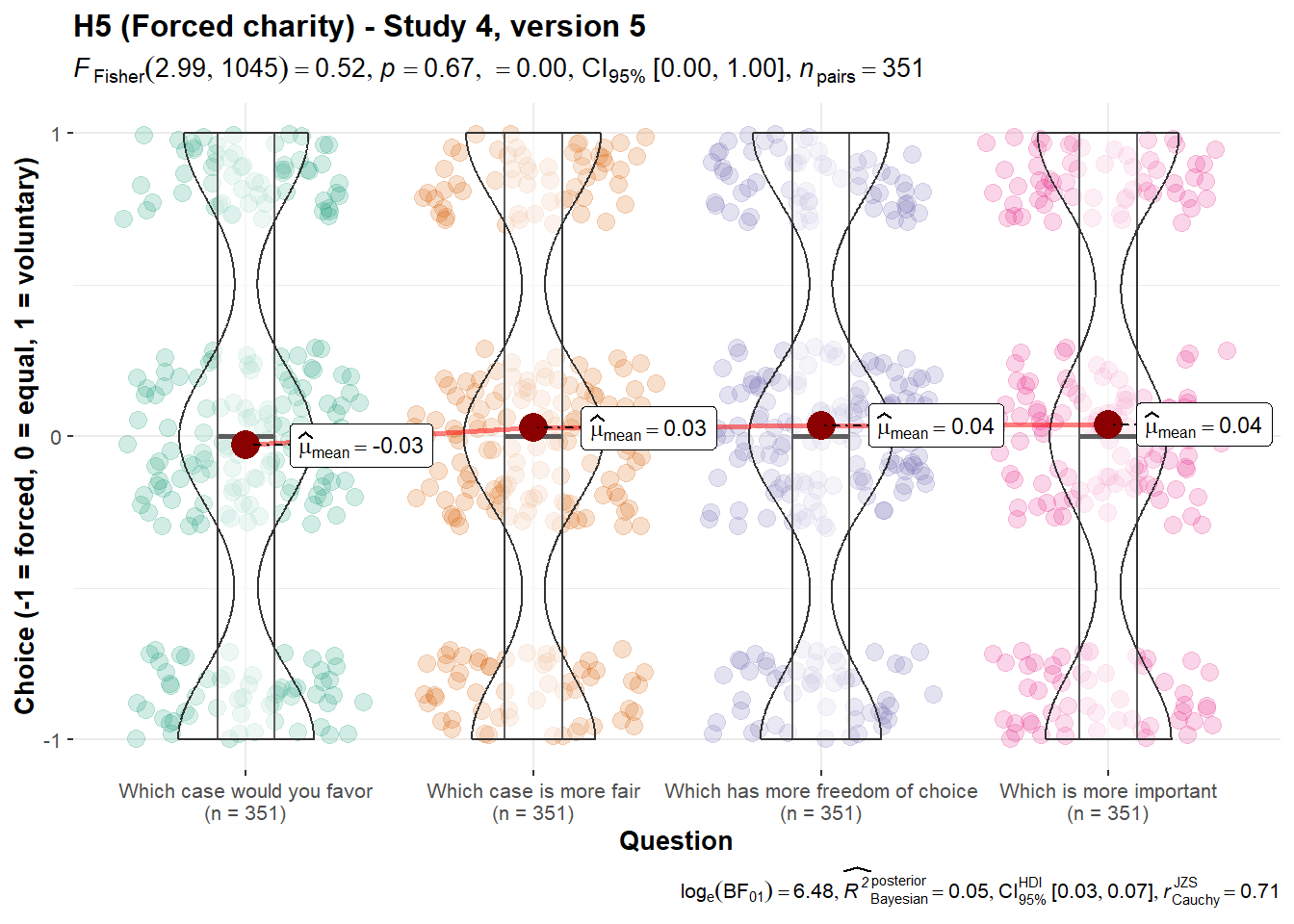
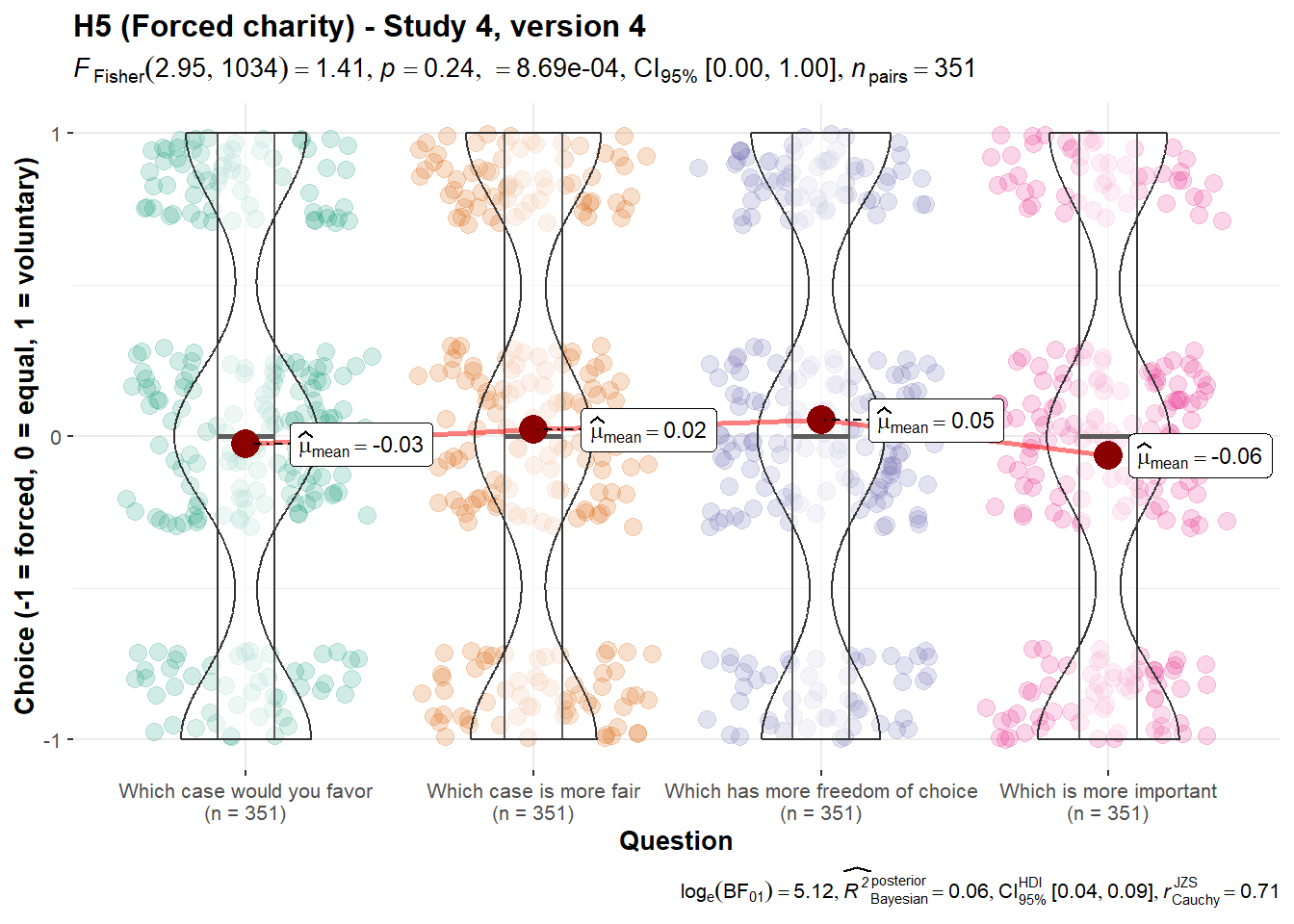
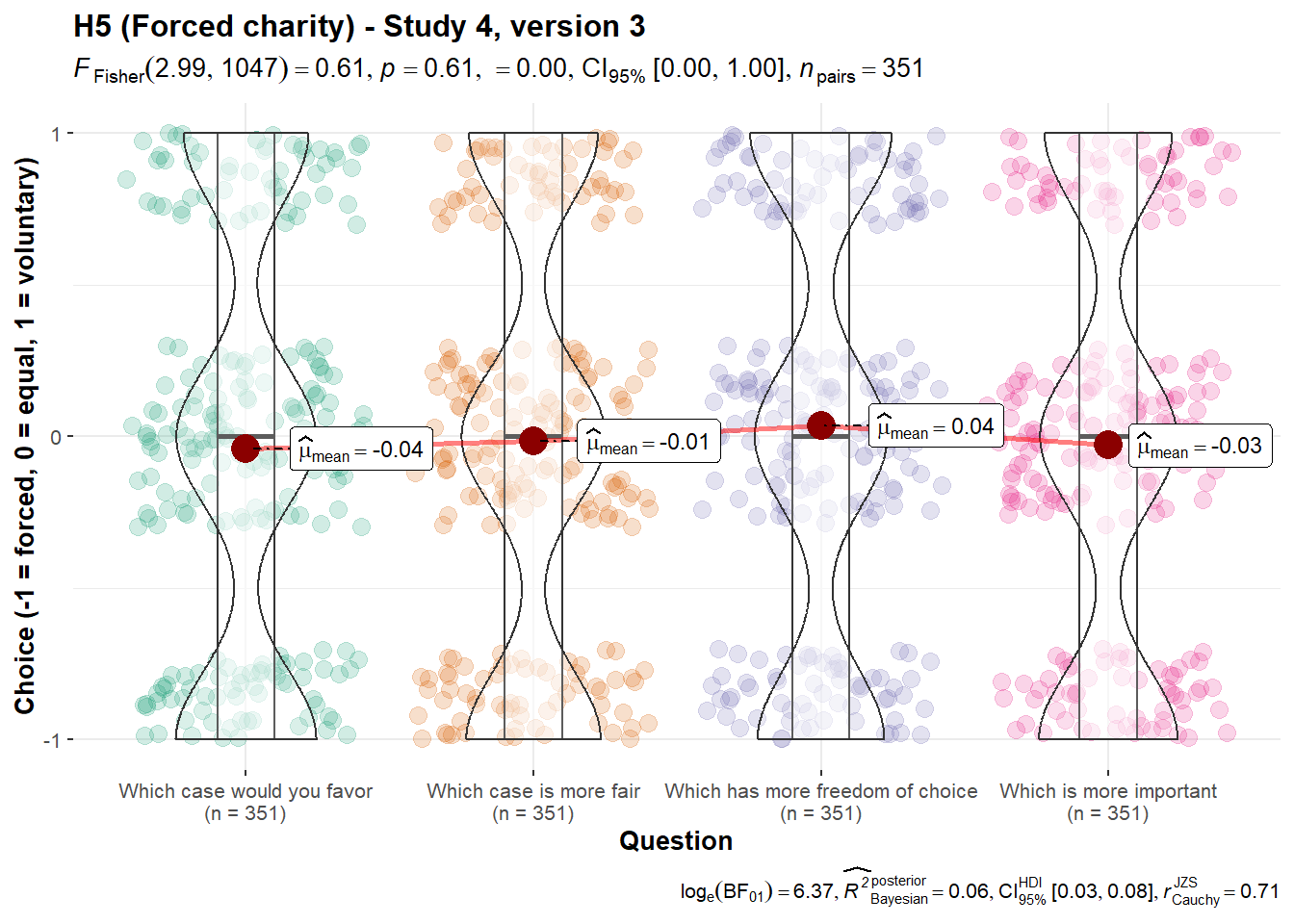
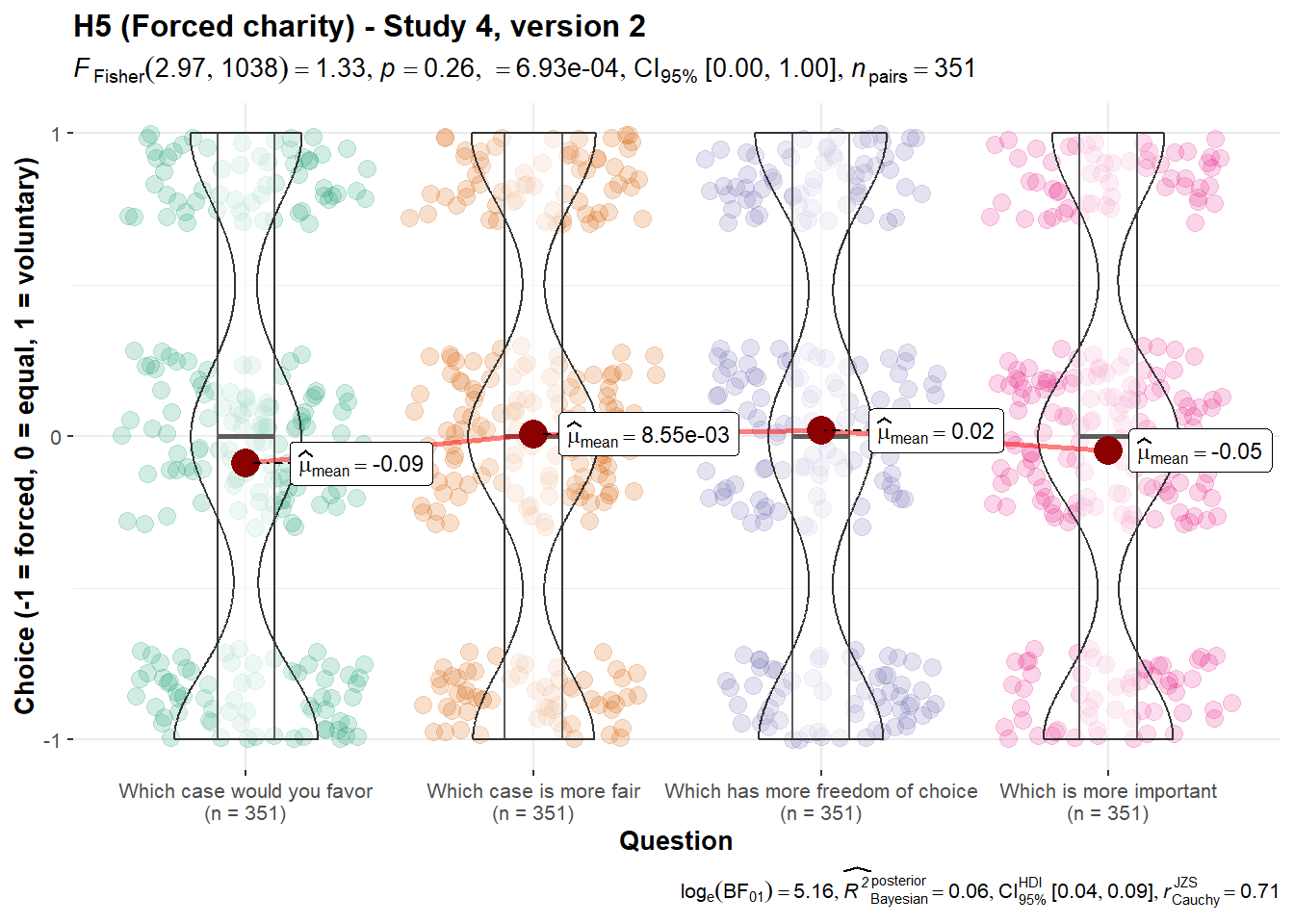
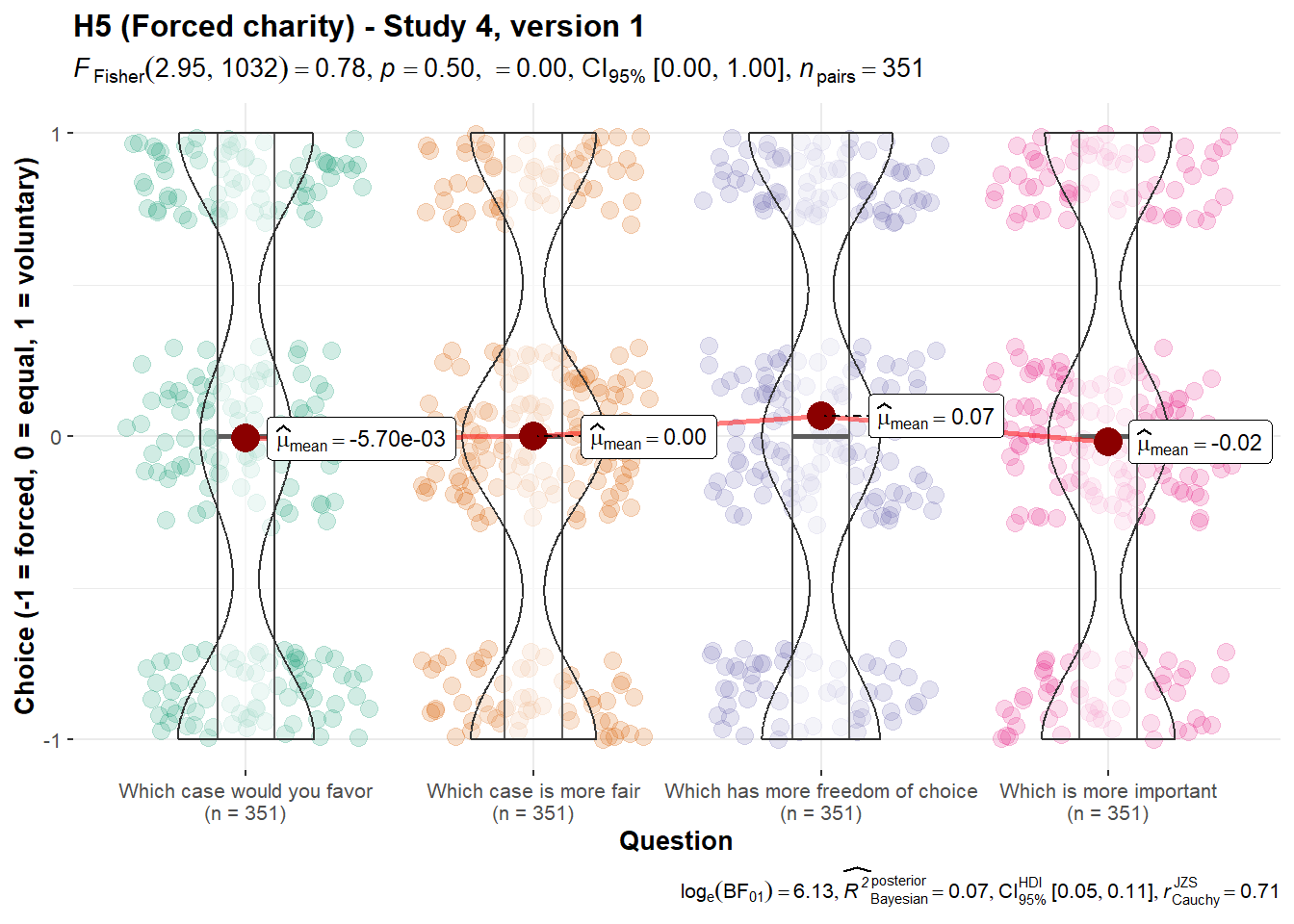
*Forced charity in Study 4, versions 1 to 5: Mean response*

*Note*. See the supplementary for differences between the five versions. This is the mean of the four dependent measures: "which case would you favor", "which case is more fair", "which has more freedom of choice" (-1 = the forced case; 0 = both cases are equal; 1 = the voluntary case), and "which [of fair cost allocation and freedom of choice] is more important" (-1 = fair cost allocation; 0 = both are equal; 1 = freedom of choice).

We then conducted five one-way repeated measures ANOVA tests and found no support for differences in responses between the questions in version 1 (*F*(2.95, 1032) = 0.781, *p* = .503, partial ω² = 0.00), version 2 (*F*(2.97, 1038) = 1.37, *p* = .264, partial ω² = 0.00), version 3 (*F*(2.99, 1047) = 0.607, *p* = .610, partial ω² = 0.00), version 4 (*F*(2.95, 1034) = 0.674, *p* = .239, partial ω² = 0.00), and version 5 (*F*(2.96, 1045) = 0.523, *p* = .665, partial ω² = 0.00). We summarized and plotted all analyses in Figure 15.

**Figure 15**

*Forced charity in Study 4, versions 1 to 5: Comparison of responses*

**

*Note.* Scale is ordinal; jitter was added for visualization purposes.

### Comparing replication to original findings

We provide a summary of replication statistical tests below in Tables 5 and 6. For tests corresponding to ones in the original study where enough details were provided to calculate an effect size, we interpret the results of our replication based on the criteria in LeBel et al. (2019) by comparing our replication effect sizes and confidence intervals to the original effect sizes in the target article.

**Table 5**

*Summary of replication statistical tests (one-sample t-tests)*

|  |  |  |  |  |  |  |  |  | Target article |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hypothesis | Study (Version/Question) | *M* | *SD* | t-test midpoint | *t* | *df* | *p* | Cohen’s *d* and 95% CI | Cohen’s *d* and 95% CI | Interpretation |
| 1 (Waste/ overhead) | 1 | 52.94 | 30.72 | 50 | 1.79 | 349 | .074 | 0.10 [-0.01, 0.20] |  |  |
| 2 | 48.80 | 31.43 | 50 | -0.71 | 349 | .476 | -0.04 [-0.14, 0.07] |  |  |
| 3 | 2.88 | 1.42 | 3 | -1.54 | 348 | .124 | -0.08 [-0.19, 0.02] | 0.37 [0.15, 0.56] | No signal — inconsistent |
| 2 (Past costs) | 1 | 47.37 | 31.18 | 50 | -1.58 | 349 | .116 | -0.08 [-0.19, 0.02] |  |  |
| 2 | 49.20 | 31.96 | 50 | -0.47 | 349 | .639 | -0.03 [-0.13, 0.08] | 0.42 [0.19, 0.64] | No signal — inconsistent |
| 3 (Diversification effect) (Unequal efficiency) | 2 (version 1) | 49.57 | 30.49 | 0 | 30.4 | 349 | <.001 | 1.63 [1.47, 1.79] |  |  |
| 2 (version 2) | 49.54 | 31.39 | 0 | 29.5 | 349 | <.001 | 1.58 [1.42, 1.73] |  |  |
| 2 (version 3) | 50.66 | 31.04 | 0 | 30.5 | 349 | <.001 | 1.63 [1.47, 1.79] |  |  |
| 4 (right allocation) | 2.98 | 1.43 | 1 | 44.9 | 1052 | <.001 | 1.38 [1.30, 1.47] |  |  |
| 4 (feel best) | 3.03 | 1.41 | 1 | 46.8 | 1052 | <.001 | 1.44 [1.36, 1.53] |  |  |
| 4 (most efficient) | 3.00 | 1.42 | 1 | 45.8 | 1052 | <.001 | 1.41 [1.33, 1.50] |  |  |
| 4 (most good done) | 3.05 | 1.43 | 1 | 46.5 | 1052 | <.001 | 1.43 [1.35, 1.52] |  |  |
| (Unequal efficiency, several projects versus one) | 1 | 52.34 | 31.73 | 0 | 30.9 | 349 | <.001 | 1.65 [1.49, 1.81] |  |  |
| 3 | 3.06 | 1.46 | 1 | 26.3 | 348 | <.001 | 1.41 [1.26, 1.56] |  |  |
| 4 (right allocation) | 2.95 | 1.38 | 1 | 26.5 | 350 | <.001 | 1.41 [1.26, 1.56] |  |  |
| 4 (efficient allocation) | 2.96 | 1.36 | 1 | 27.0 | 350 | <.001 | 1.44 [1.29, 1.59] |  |  |
| (Equal efficiency) | 1 | 50.77 | 31.85 | 50 | 0.45 | 349 | .651 | 0.02 [-0.08, 0.13] |  |  |
| 2 | 46.69 | 32.93 | 50 | -1.88 | 349 | .061 | -0.10 [-0.21, 0.00] |  |  |
| 4 (Ingroup effect) | 1 | 49.94 | 32.14 | 50 | -0.03 | 349 | .974 | 0.00 [-0.11, 0.10] |  |  |
| 2 (around the world) | 50.77 | 32.48 | 50 | 0.44 | 349 | .657 | 0.02 [-0.08, 0.13] |  |  |
| 2 (India) | 53.69 | 31.52 | 50 | 2.19 | 349 | .029 | 0.12 [0.01, 0.22] |  |  |
| 2 (Africa) | 49.23 | 32.49 | 50 | -0.44 | 349 | .657 | -0.02 [-0.13, 0.08] |  |  |
| 2 (Latin America) | 49.89 | 31.96 | 50 | -0.07 | 349 | .947 | 0.00 [-0.11, 0.10] |  |  |
| 3 (India) | 3.07 | 1.35 | 3 | 0.91 | 348 | .362 | 0.05 [-0.06, 0.15] |  |  |
| 3 (Eastern Europe) | 2.85 | 1.37 | 3 | -2.07 | 348 | .039 | -0.11 [-0.22, -0.01] |  |  |
| 3 (China) | 2.98 | 1.36 | 3 | -0.32 | 348 | .752 | -0.02 [-0.12, 0.09] |  |  |
| 3 (Africa) | 3.06 | 1.41 | 3 | 0.76 | 348 | .447 | 0.04 [-0.06, 0.15] |  |  |
| 5 (Forced-charity/ Government-taxes effect) | 4 (version 1) | 0.01 | 0.43 | 0 | 0.49 | 350 | .622 | 0.03 [-0.08, 0.13] |  |  |
| 4 (version 2) | -0.03 | 0.41 | 0 | -1.19 | 350 | .234 | -0.06 [-0.17, 0.04] |  |  |
| 4 (version 3) | -0.01 | 0.40 | 0 | -0.54 | 350 | .593 | -0.03 [-0.13, 0.08] | 0.32 [0.09, 0.55] | No signal — inconsistent |
| 4 (version 4) | 0.00 | 0.41 | 0 | -0.13 | 350 | .897 | 0.00 [-0.11, 0.10] | 0.70 [0.43, 0.93] | No signal — inconsistent |
| 4 (version 5) | 0.02 | 0.38 | 0 | 0.96 | 350 | .339 | 0.05 [-0.05, 0.16] |  |  |

*Note*. Outcome interpretations are based on LeBel et al. (2019) where target article Cohen’s d and 95% CI are available.

**Table 6**

*Summary of replication statistical tests (paired t-tests)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Condition 1 | | | Condition 2 | | |  |  |  |  | Target article |  |
| Hypothesis and study | Comparison | Label | *M* | *SD* | Label | *M* | *SD* | *t* | *df* | *p* | Cohen’s *d* and 95% CI | Cohen’s *d* and 95% CI | Interpretation |
| 3 (Diversifi- cation effect), Study 4 (Unequal efficiency) | Comparison of means | Right allocation/feel best | 3.00 | 0.57 | Most efficient use/most good done | 3.03 | 0.58 | -0.58 | 350 | .562 | -0.03 [-0.14, 0.07] | 0.36 [0.13, 0.59] | No signal — inconsistent |
| Comparison of proportions | Right allocation/feel best | 0.80 | 0.16 | Most efficient use/most good done | 0.81 | 0.16 | -0.20 | 350 | .840 | -0.01 [-0.12, 0.09] | 0.41 [0.18, 0.64 | No signal — inconsistent |
| (Unequal efficiency, several projects versus one) | Comparison of allocation | Right allocation | 2.95 | 1.38 | Most efficient use | 2.96 | 1.36 | -0.14 | 350 | .888 | 0.00 [-0.11, 0.10] |  |  |

*Note*. Outcome interpretations are based on LeBel et al. (2019) where target article Cohen’s d and 95% CI are available.

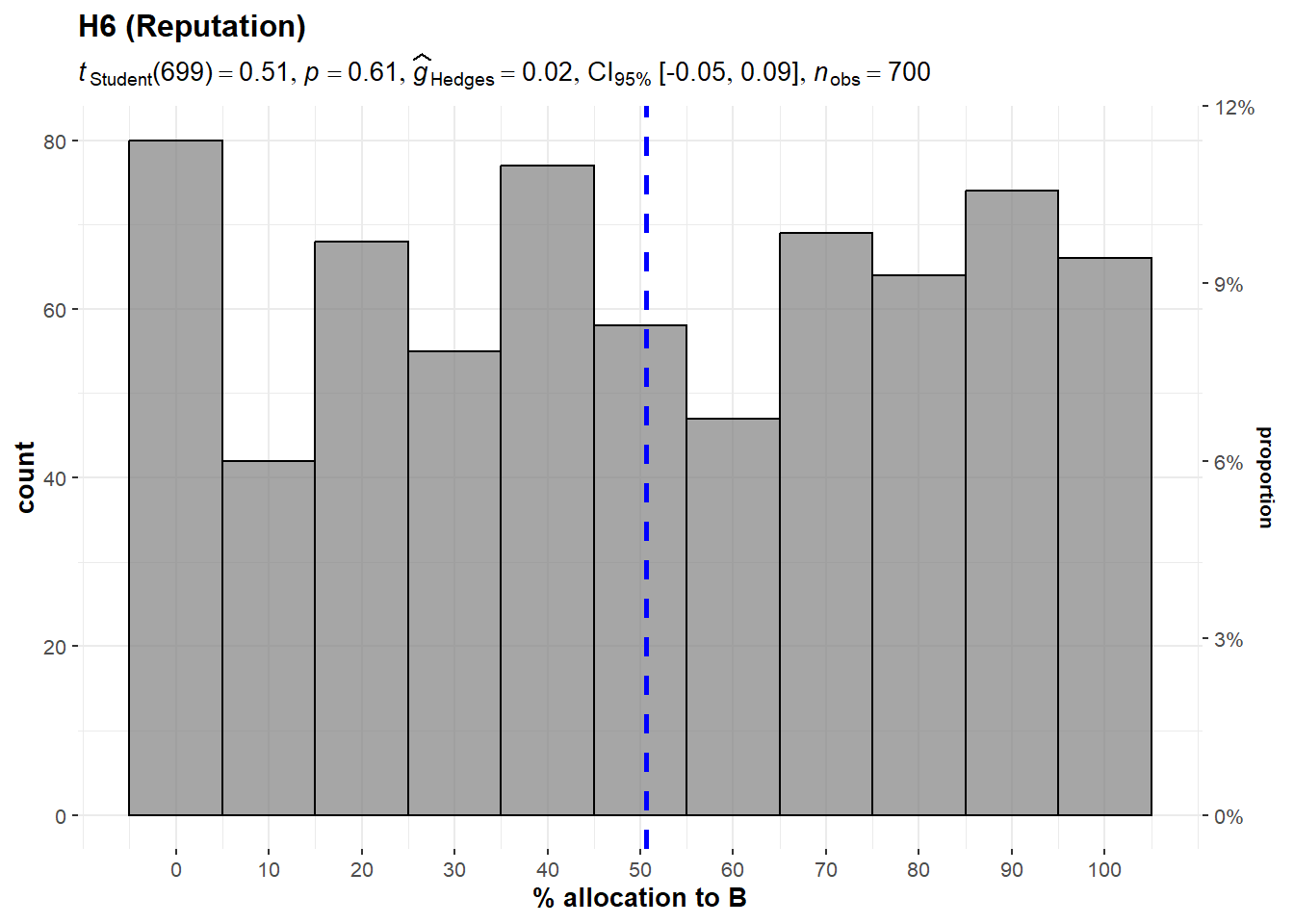
## Extensions

### 6. Reputation/publicity (Hypothesis 6)

We conducted a one-sample t-test and failed to find support for participant preference in donating to causes that could improve their reputation (against a 50% midpoint; *M* = 50.63, *SD* = 32.34, *t*699 = 0.51, *p* = .607, *d* = 0.02, 95% CI [-0.05, 0.09]; summarized and plotted in Figure 15).

**Figure 15**

*Reputation in Studies 1 and 2 combined: Allocation*

  
*Note*. Scenario: “Study 1, Study 2: **A** and **B** both help thousands of children. **A** publishes the names of donors and how much they donated on their website. **B** keeps donors anonymous.”

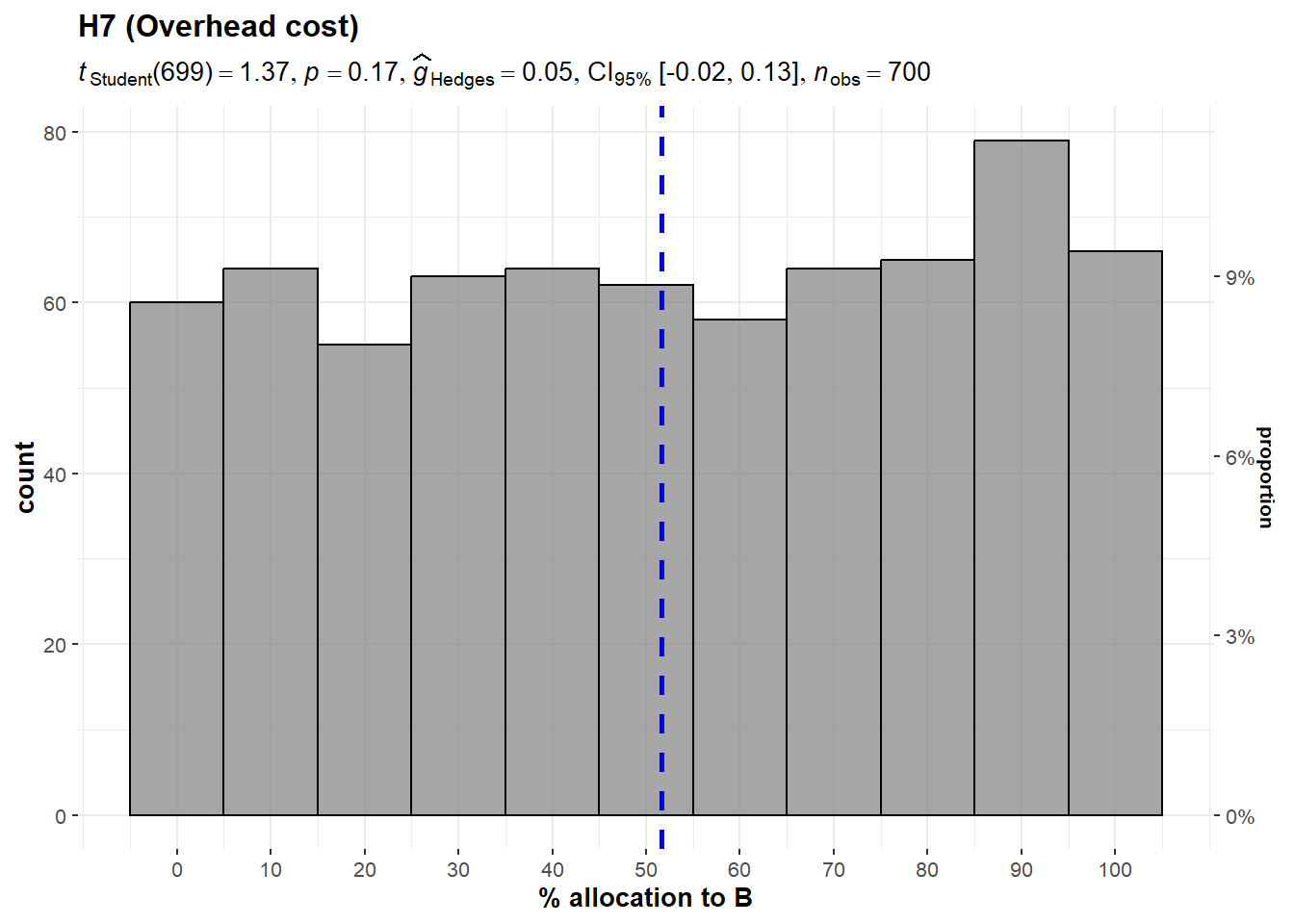
### 

### 7. Overhead costs external funding (Hypothesis 7)

We conducted a one-sample t-test and failed to find support for participant preference in donating to causes for which overhead has been paid for by another donor (against a 50% mid-point; *M* = 51.66, *SD* = 31.94, *t*699 = 1.37, *p* = .170, *d* = 0.05, 95% CI [-0.02, 0.13]; summarized and plotted in Figure 16).

**Figure 16**

*Overhead funding in Studies 1 and 2 combined: Allocation*



*Note*. Scenario: “Study 1, Study 2: **A** and **B** both help thousands of children. Both charities spend 50% of the donations they receive on administrative costs. For each $100 contribution to **A**, $50 will go to helping children and $50 will be used to cover administrative costs. For each $100 contribution to **B**, all $100 will go to helping children; another donor will cover the corresponding $100 administrative cost of this contribution.”

### Summary of extension findings

We provide a summary of replication statistical tests below in Table 7.

**Table 7**

*Summary of extension statistical tests (One-way t-tests)*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *M* | *SD* | Mid-  point | t-stat | df | *p* | Cohen's *d* and 95% CI |
| Reputation  (Hypothesis 6) | | | | | | | |
| Studies 1 and 2 | 50.63 | 32.34 | 50 | 0.51 | 699 | .607 | 0.02 [-0.05, 0.09] |
| Overhead funding  (Hypothesis 7) | | | | | | | |
| Studies 1 and 2 | 51.66 | 31.94 | 50 | 1.37 | 699 | .170 | 0.05 [-0.02, 0.13] |

## 

## Exploratory analyses

[To be completed in stage 2.]

# Discussion

[Please note that the discussion and conclusion sections are only to be completed in Stage 2 following data collection.]

**Replication**

## Extensions

## Implications

## Limitations and directions for future research

[Planned discussion for Stage 2: Following on Dr./Prof. Jonathan Berman’s comment: Studies are within-subject, as per the target article we aim to replicate; Yet, we will discuss advantages and disadvantages of that approach and comparisons between a within and a between subject design. People may react in ways they wouldn’t if the experiments were conducted using a between-subject design . We will also note this as directions for future research.

Planned discussion for Stage 2: Following on Dr./Prof. Jonathan Berman’s comment: There seems to be a methodological weakness in the test of Hypothesis 3, such that few participants may impact results.

Planned discussion for Stage 2: Following Dr./Prof. Amanda Geiser’s suggestion, discussing the generalizability of results and expanding on these in future studies in real-life and the field.

Planned discussion for Stage 2: Results not representative of all possible impediments to effective altruism, we’ll discuss suggestions for future replications and work referring to recommendations made by reviewer Dr./Prof. Amanda Geiser for Hsee et al. (2013) and Caviola, Schubert, & Greene (2021).

Planned discussion for Stage 2: Based on the Dr./Prof. Romain Espinosa’s comment on multiple analyses, we will discuss our adjusted alpha threshold with advantages and disadvantages, with references to supplementary Bayes analyses, and suggestions for replication work.

Planned discussion for Stage 2: Following on Dr./Prof. Jonathan Berman’s comment we will discuss the differences between Gneezy et al. (2014) and Camerer et al. (2018) and our conceptual replication implementation in our extension, discussing between versus within designs and other elements. ]

**Conclusion**

# References

Alpizar, F., Carlsson, F., & Johansson-Stenman, O. (2008). Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics*, *92*(5-6), 1047–1060. <https://doi.org/10.1016/j.jpubeco.2007.11.004>

Baron, J., & Szymanska, E. (2011). Heuristics and Biases in Charity. In D. M. Oppenheimer & C. Y. Olivola (Eds.), *The Science of Giving: Experimental Approaches to the Study of Charity* (pp. 215–235). Psychology Press. <https://doi.org/10.4324/9780203865972-24>

Bekkers, R., & Wiepking, P. (2010). A Literature Review of Empirical Studies of Philanthropy. *Nonprofit and Voluntary Sector Quarterly*, *40*(5), 924–973. <https://doi.org/10.1177/0899764010380927>

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software, 5*(56), 2815. <https://doi.org/10.21105/joss.02815>

Berman, J. Z., Barasch, A., Levine, E. E., & Small, D. A. (2018). Impediments to Effective Altruism: The Role of Subjective Preferences in Charitable Giving. *Psychological Science*, *29*(5), 834–844. <https://doi.org/10.1177/0956797617747648>

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1875), 20180116. <https://doi.org/10.1098/rspb.2018.0116>

Burtch, G., Ghose, A., & Wattal, S. (2014). Cultural Differences and Geography as Determinants of Online Prosocial Lending. *MIS Quarterly*, *38*(3), 773–794. <https://doi.org/10.25300/misq/2014/38.3.07>

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes*, *151*, 16–33. <https://doi.org/10.1016/j.obhdp.2018.12.006>

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., & Isaksson, S. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>

Caviola, L., Faulmüller, N., Everett, Jim. A. C., Savulescu, J., & Kahane, G. (2014). The evaluability bias in charitable giving: Saving administration costs or saving lives? *Judgment and Decision Making*, *9*(4), 303–315. <https://doi.org/10.1017/s1930297500006185>

Caviola, L., Schubert, S., & Greene, J. D. (2021). The Psychology of (In)Effective Altruism. *Trends in Cognitive Sciences*, *25*(7), 596–607. <https://doi.org/10.1016/j.tics.2021.03.015>

Fox, C. R., Ratner, R. K., & Lieb, D. S. (2005). How Subjective Grouping of Options Influences Choice and Allocation: Diversification Bias and the Phenomenon of Partition Dependence. *Journal of Experimental Psychology: General*, *134*(4), 538–551. <https://doi.org/10.1037/0096-3445.134.4.538>

Gneezy, U., Keenan, E. A., & Gneezy, A. (2014). Avoiding overhead aversion in charity. *Science*, *346*(6209), 632–635. <https://doi.org/10.1126/science.1253932>

Hsee, C. K. (1996). The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257. <https://doi.org/10.1006/obhd.1996.0077>

James, T. K., & Zagefka, H. (2017). The effects of group memberships of victims and perpetrators in humanly caused disasters on charitable donations to victims. *Journal of Applied Social Psychology*, *47*(8), 446–458. <https://doi.org/10.1111/jasp.12452>

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, 1-15.

Larmarange, J. (2022). *labelled: Manipulating Labelled Data*. R package version 2.10.0. [https://CRAN.R-project.org/package=labelled](https://cran.r-project.org/package=labelled)

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*, 389-402. <https://doi.org/10.1177/2515245918787489>

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, *3*, 1-9. <https://doi.org/10.15626/MP.2018.843>

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods*, *49*(2), 433-442. <https://doi.org/10.3758/s13428-016-0727-z>

Lüdecke, D. (2022). *sjlabelled: Labelled Data Utility Functions* (Version 1.2.0). <https://doi.org/10.5281/zenodo.1249215>

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103–122. <https://doi.org/10.1002/ejsp.2420220202>

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>

Patil, I., (2021a). statsExpressions: R Package for Tidy Dataframes and Expressions with Statistical Details. *Journal of Open Source Software, 6*(61), 3236. <https://doi.org/10.21105/joss.03236>

Patil, I. (2021b). Visualizations with statistical details: The “ggstatsplot” approach. *Journal of Open Source Software*, 6(61), 3167. <https://doi.org/10.21105/joss.03167>

Pohlert, T. (2022). *PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended*. R package version 1.9.6. [https://CRAN.R-project.org/package=PMCMRplus](https://cran.r-project.org/package=PMCMRplus)

Posit team (2022). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. <http://www.posit.co/>

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. [https://www.R-project.org](https://www.r-project.org)

Read, D., & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, *1*(1), 34–49. <https://doi.org/10.1037/1076-898x.1.1.34>

Revelle, W. (2022) *psych: Procedures for Personality and Psychological Research*. Northwestern University. R package version 2.2.9. [https://CRAN.R-project.org/package=psych](https://cran.r-project.org/package=psych)

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science*, *26*(5), 559–569. <https://doi.org/10.1177/0956797614567341>

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. (2023). *afex: Analysis of Factorial Experiments*. R package version 1.2-1. [https://CRAN.R-project.org/package=afex](https://cran.r-project.org/package=afex)

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software, 21*(12). <https://doi.org/10.18637/jss.v021.i12>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., & Takahashi, K. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. <https://doi.org/10.21105/joss.01686>

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.0. [https://CRAN.R-project.org/package=dplyr](https://cran.r-project.org/package=dplyr)

Wickham, H., Miller, E., & Smith, D. (2023). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. R package version 2.5.2. [https://CRAN.R-project.org/package=haven](https://cran.r-project.org/package=haven)

Wickham, H., & Sievert, C. (2016). *ggplot2: Elegant Graphics For Data Analysis*. Springer.

Xiao, Q., Yeung, S. K., Dunleavy, D. J., Röseler, L., Elsherif, M., & Feldman, G. (2023) Effect sizes and confidence intervals guide. DOI: 10.17605/OSF.IO/D8C4G . Retrieved from: <https://osf.io/d8c4g/>