

**Bayes Factor Design Analysis for the Cover Story Effects in Causality Judgment under  
Positive Contingency**

**Bayes Factor Design Analysis for the Cover Story Effects in Causality Judgment under  
Positive Contingency**

**Table of contents**

<b>Introduction</b>	<b>3</b>
<b>1. Rationale for Experiment 2</b>	<b>3</b>
<b>2. Hypothesized Differences</b>	<b>3</b>
<b>3. Bayesian Generalized Linear Model</b>	<b>5</b>
<b>4. Design (Point) Priors</b>	<b>7</b>
<b>5. Analysis Priors</b>	<b>7</b>
5.3 Prior sensitivity check . . . . .	11
<b>6. Power and False Positive Rate</b>	<b>14</b>
<b>References</b>	<b>22</b>

## **Bayes Factor Design Analysis for the Cover Story Effects in Causality Judgment under Positive Contingency**

### **1. Rationale for Experiment 2**

Building on the results of Experiment 1, the present experiment examines whether the pattern of responses observed in the absence of a true contingency (i.e., the illusion of causality) generalizes to a context in which a positive causal contingency exists between a potential cause and an outcome. Specifically, we aim to determine whether the effects of contextual framing, and in particular the activation of prior knowledge, emerge similarly when participants evaluate a positive causal scenario.

In Experiment 1, no contingency was present (illusory contingency), yet clear differences emerged across cover stories in the proportion of participants who judged the cause–effect relationship to be real: approximately 62% in the medicine task, 51% in the lever-and-machine task, and 29% in the alien task. This pattern suggests a strong influence of prior knowledge on causal judgments. Notably, the alien scenario, designed to minimize pre-existing beliefs, yielded the lowest rate of erroneous causal judgments. By contrast, although the lever scenario was framed as more deterministic, participants’ familiarity with mechanical cause–effect relations may have reintroduced expectations that partially counteracted the reduction in illusory judgments expected from deterministic framing alone.

In the present experiment, we introduce a positive causal contingency while retaining the same three cover stories. The numerosity and overall design structure are matched to those of Experiment 1 ( $N = 195$ ), thereby preserving an equivalent level of informational evidence. This design enables a direct assessment of whether the previously observed pattern of contextual effects persists when participants are exposed to a genuine causal relationship.

### **2. Hypothesized Differences**

For Experiment 2, we hypothesize heuristically that introducing a positive causal contingency will result in an overall increase of approximately 25 percentage points in “yes” responses across all conditions. Based on the observed proportions in Experiment 1, this

corresponds to expected rates of approximately 87% in the Medicine condition, 76% in the Lever-and-Machine condition, and 54% in the Alien condition.

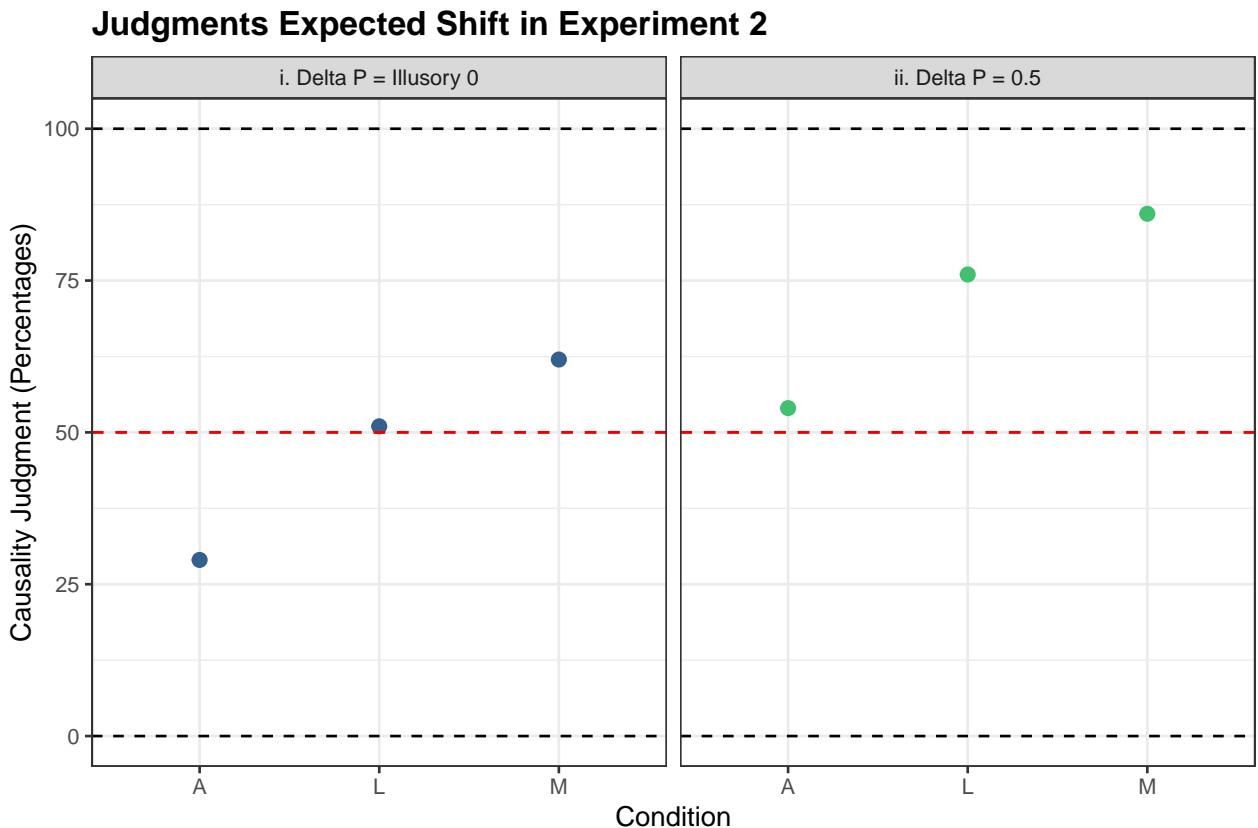
```
# Plot difference between Exp 1 and 2

library(ggplot2)

causality <- data.frame(
  "Cover Story" = as.factor(rep(c("M","L","A"), times = 2)),
  "Delta P" = as.factor(rep(c("i. Delta P = Illusory 0",
                             "ii. Delta P = 0.5"),
                            each = 3)),
  "Percent Estimates" = round(c(c(61.54, 50.77, 29.23),
                                 c(61.5 ,50.77,29.23)+25)))
  )

ggplot(causality, aes(x = Cover.Story,
                      y = Percent.Estimates,
                      color = Delta.P)) +
  geom_point(size = 3) +
  geom_hline(yintercept = 50, color = "red2", linetype = "dashed",
             lwd=.7) +
  theme_bw(base_size = 14) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  geom_hline(yintercept = 100, linetype = "dashed") +
  theme(
    plot.title = element_text(face = "bold"),
    legend.position = "right"
  ) +
  facet_grid(~Delta.P) +
  scale_color_viridis_d(begin = .3, end = .7) +
```

```
labs(title =
  "Judgments Expected Shift in Experiment 2",
  y = "Causality Judgment (Percentages)",
  x = "Condition") +
theme(legend.position = "none")
```



### 3. Bayesian Generalized Linear Model

We plan to recruit a total of 195 participants, equally distributed across the three experimental conditions (65 per condition). Data will be analyzed using a Bayesian generalized linear model with a binomial likelihood and a logit link function. The model includes cover story as a categorical predictor, and inference will focus on two pre-specified contrasts.

- H1 (General prior-knowledge effect). The Alien Task condition will be contrasted against the average of the Medicine and Lever-and-Machine Task conditions, capturing the overall influence of prior knowledge on causal judgments.

- H2 (Probabilistic prior-knowledge effect). The Alien Task condition will be contrasted specifically against the Medicine Task condition, capturing the influence of prior knowledge in a probabilistic context.

```
# Custom contrasts

dummy <- data.frame(
  y = rep(c(0, 1, 0), 65),
  group = factor(rep(
    c("Alien", "Lever", "Medicine"), 65),
    levels = c("Alien", "Lever", "Medicine")
  )
)

C <- matrix(
  c( 1,  1,
    -0.5, 0,
    -0.5, -1),
  ncol = 2,
  byrow = TRUE
)

colnames(C) <- c("H1. Alien vs Others", "H2. Alien vs Medicine")
contrasts(dummy$group) <- C
knitr::kable(contrasts(dummy$group))
```

The two planned contrasts will be evaluated within an overall Bayesian generalized linear model with a binomial likelihood and a logit link function:

$$\text{logit}(\Pr(y = 1)) = \beta_0 + \beta_1 \cdot \text{Contrast}_1 + \beta_2 \cdot \text{Contrast}_2$$

Model comparison will be conducted using Bayes Factors (BF), defined as the ratio of the marginal likelihoods of the competing models:

$$BF_{10} = \frac{p(\mathbf{y} | M_1)}{p(\mathbf{y} | M_0)}$$

where  $M_1$  denotes the full model including the planned contrasts, and  $M_0$  denotes the null model including only the intercept term:

$$\text{logit}(\Pr(y = 1)) = \beta_0$$

In addition to model-level comparisons, we will examine the posterior distributions of the regression coefficients  $\beta_1$  and  $\beta_2$  of  $M_1$ .

#### 4. Design (Point) Priors

As asserted before, we assume that introducing a positive causal contingency will result in an increase of approximately 25 percentage points in “yes” responses across conditions. Under this assumption, the expected rates are approximately 87% in the Medicine condition, 76% in the Lever-and-Machine condition, and 54% in the Alien condition. These expected point estimate differences will serve as point-targeted design priors for BFDA, functioning as the key simulation parameters.

```
knitr::kable(causality[causality$Delta.P=="ii. Delta P = 0.5",])
```

#### 5. Analysis Priors

Consistent with Experiment 1, weakly informative priors were specified for all regression coefficients on the log-odds scale. The intercept term was assigned a Student-(t) prior with 7 degrees of freedom, location 0, and scale 1.5:

$$\beta_0 \sim \text{Student-}t(7, 0, 1.5)$$

The coefficients associated with the planned contrasts were assigned broader Student-(t)

priors with 3 degrees of freedom, location 0, and scale 2.5:

$$\beta_1, \beta_2 \sim \text{Student-}t(3, 0, 2.5)$$

```
library(brms)
library(bayestestR)
library(bayesplot)

# Priors
prior <- c(
  # Intercept
  prior(student_t(7, 0, 1.5), class = "Intercept"),
  # Beta(s)
  prior(student_t(3, 0, 2.5), class = "b")
)

# Intercept prior
x <- seq(-10, 10, length.out = 1000)
df_intercept <- 7
scale_intercept <- 1.5
y_intercept <- dt((x - 0)/scale_intercept, df=df_intercept)/scale_intercept
intercept_df <- data.frame(x = x, density = y_intercept)

# Coefficient prior
df_beta <- 3
scale_beta <- 2.5
y_beta <- dt((x - 0)/scale_beta, df=df_beta)/scale_beta
beta_df <- data.frame(x = x, density = y_beta)
```

```
# Log scale

p_logodds <- ggplot() +
  geom_line(data = intercept_df, aes(x = x, y = density),
            color = "firebrick", size = 1.2) +
  geom_line(data = beta_df, aes(x = x, y = density),
            color = "darkblue", size = 1.2) +
  labs(title = "Priors on Log-Odds Scale",
       x = "Log-Odds",
       y = "Density") +
  theme_bw() +
  annotate("text", x = 6, y = 0.15, label = "Intercept",
           color = "firebrick") +
  annotate("text", x = 6, y = 0.08, label = "Coefficient",
           color = "darkblue")

# Probability scale

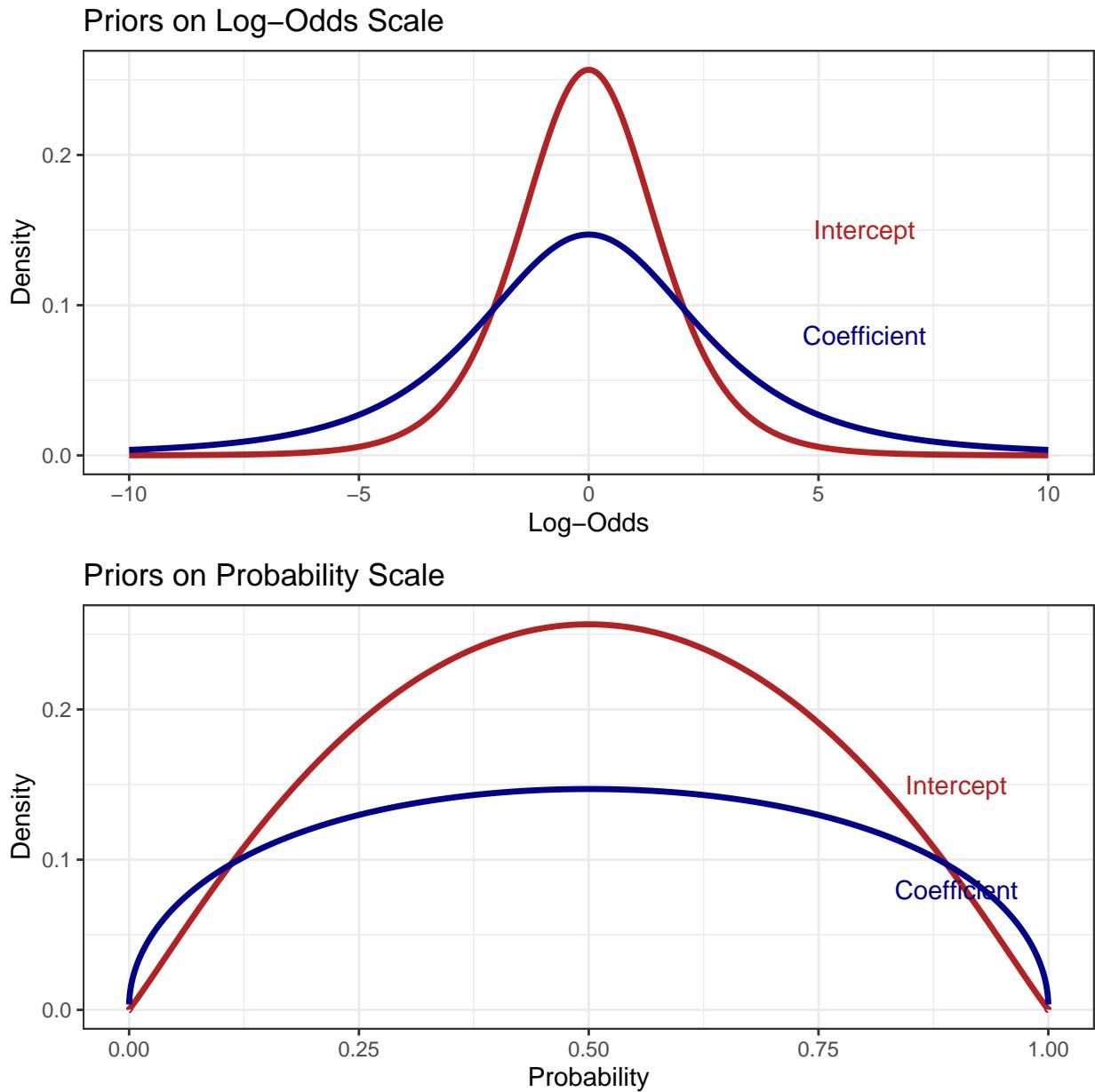
theta <- plogis(x)

intercept_prob <- data.frame(theta = theta, density = y_intercept)
beta_prob <- data.frame(theta = theta, density = y_beta)

p_prob <- ggplot() +
  geom_line(data = intercept_prob, aes(x = theta, y = density),
            color = "firebrick", size = 1.2) +
  geom_line(data = beta_prob, aes(x = theta, y = density),
            color = "darkblue", size = 1.2) +
  labs(title = "Priors on Probability Scale",
```

```
x = "Probability",
y = "Density") +
theme_bw() +
annotate("text", x = 0.9, y = 0.15, label = "Intercept", color = "firebrick") +
annotate("text", x = 0.9, y = 0.08, label = "Coefficient", color = "darkblue")

# Visualization
library(ggpubr)
ggarrange(p_logodds, p_prob, ncol = 1, common.legend = TRUE)
```



### 5.3 Prior sensitivity check

A prior-only model was fit to conduct a sensitivity analysis. This allows us to evaluate the implications of the analysis priors via prior predictive check.

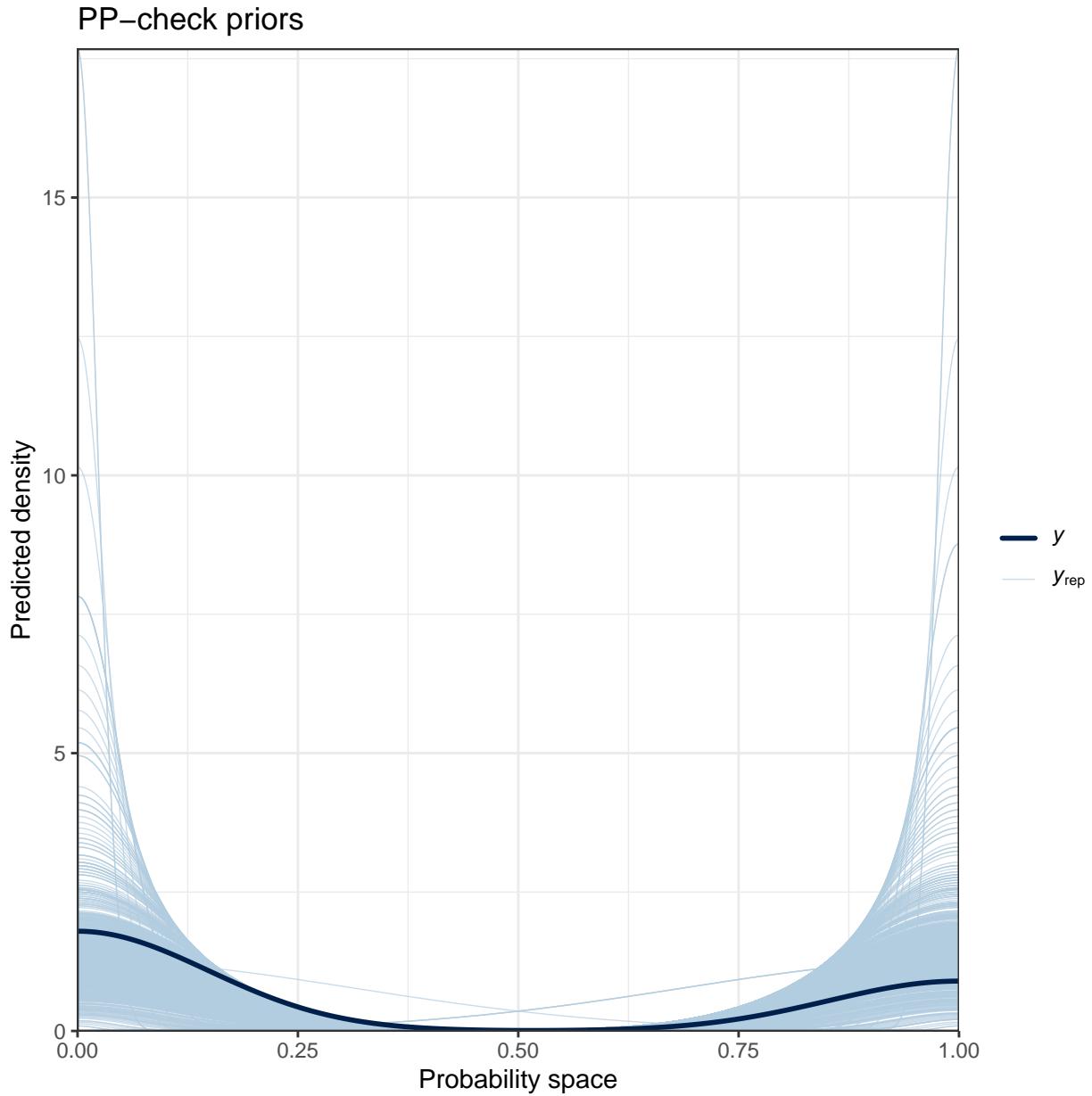
```
# Model specification

mod_h1_prior <-
  brm(y ~ 1 + # Intercept
       group, # Fixed effect
```

```
family = bernoulli(link="logit"), # Logit link  
data = dummy, save_pars = save_pars(all = TRUE), # save par.  
iter = 4000, warmup = 1000, #4000 iterations and 1000 warm ups  
prior = prior,  
seed = 2025,  
sample_prior = "only",  
file = 'H1modelwithprior-function.rds')  
  
knitr::kable(prior_summary(mod_h1_prior), caption = "Priors over M1")
```

```
# Posterior predictive check  
p1 <- pp_check(mod_h1_prior, ndraws = 1000) + theme_bw() +  
  labs(title = "PP-check priors", x ="Probability space",  
    y = "Predicted density")
```

```
p1
```



```
# Conditional effects

ce <- conditional_effects(mod_h1_prior, effects = "group")
invisible(capture.output(p2 <- plot(ce)[[1]] + theme_bw() +
  labs(title = "Conditional effects", x ="Probability space",
       y = "Condition")))

```

Ignoring unknown labels:

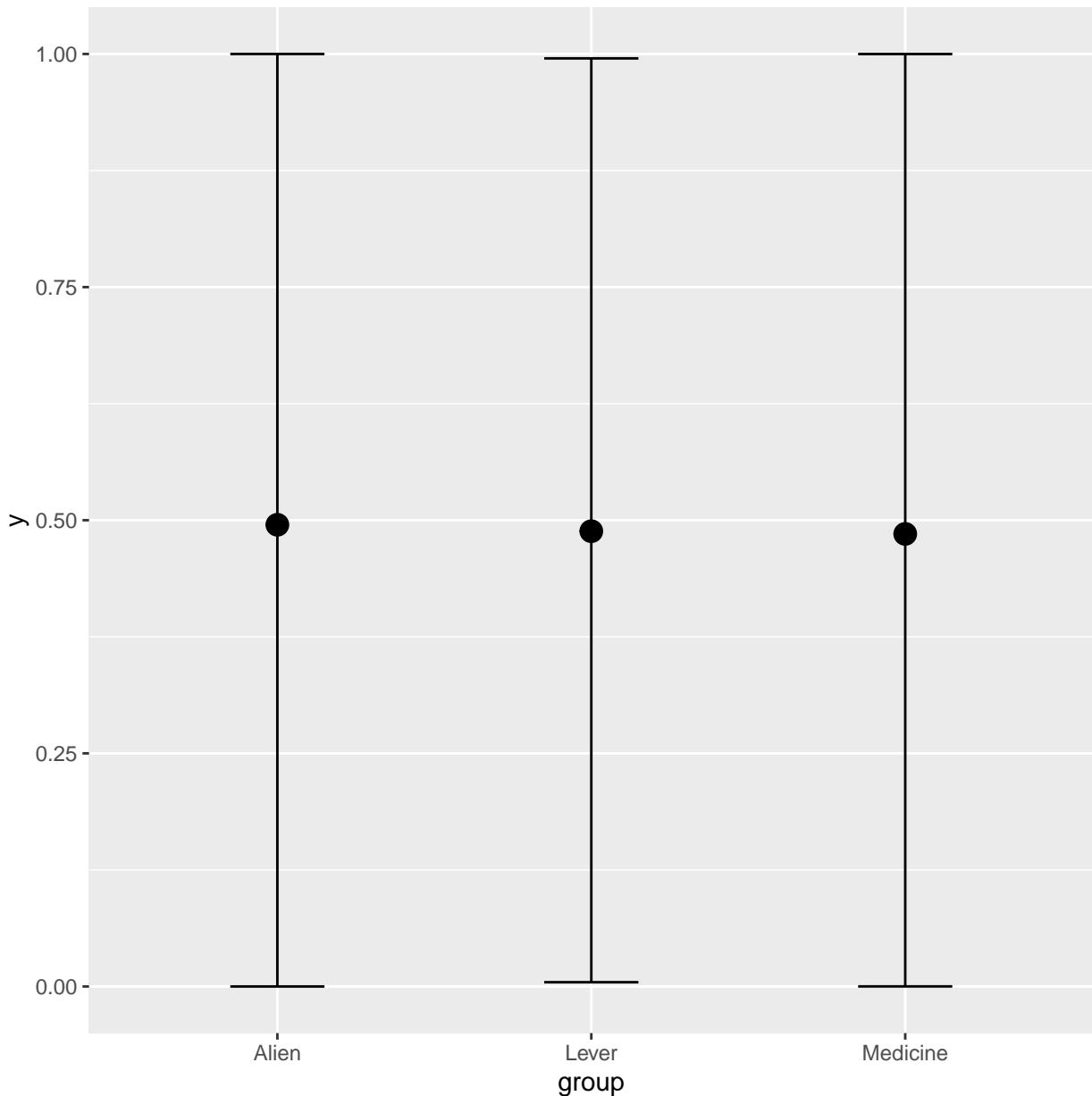
\* fill : "NA"

```
* colour : "NA"
```

Ignoring unknown labels:

```
* fill : "NA"
```

```
* colour : "NA"
```



## 6. Power and False Positive Rate

We conducted a Bayes Factor Design Analysis (BFDA) via simulation. We varied the true effect size, defined as the difference in probabilities between the Alien condition and the average

of Lever and Medicine, while holding the total sample size constant at  $N = 195$  (65 participants per condition). For each effect size, 2,000 simulated datasets were generated. Each dataset was analyzed using the planned Bayesian binomial GLM, and evidence was quantified by comparing the full model to a corresponding intercept-only null model using BF. Statistical power was defined as the proportion of simulations in which the BF exceeded the decision threshold ( $\text{BF} > 3.5$ ). When the true effect size was zero, this proportion corresponds to the false positive rate.

```
# # Parameters
#
# sample_sizes <- 195
# nsim <- 2000
# bf_threshold <- 3.5
# effect_sizes <- seq(0, 0.275, by = 0.025)
#
# # Dummy data and contrasts
#
# dummy <- data.frame(
#   y = c(0, 1, 0),
#   group = factor(
#     c("Alien", "Lever", "Medicine"),
#     levels = c("Alien", "Lever", "Medicine")
#   )
# )
#
#
#
# C <- matrix(
#   c( 2, -1,
```

```
#      -1,  0,
#      -1,  1),
# ncol = 2, byrow = TRUE
# )
#
# contrasts(dummy$group) <- C
#
#
#
# prior_full <- c(
#   prior(student_t(7, 0, 1.5), class = "Intercept"),
#   prior(student_t(3, 0, 2.5), class = "b")
# )
#
#
#
# fit_template <- brm(
#   y ~ group,
#   data = dummy,
#   family = bernoulli(),
#   prior = prior_full,
#   seed = 2025,
#   chains = 4, iter = 500, refresh = 0
# )
#
#
# fit0_template <- brm(
#   y ~ 1,
#   data = dummy,
```

```
#   family = bernoulli(),
#
#   prior = prior(student_t(7, 0, 1.5), class = "Intercept"),
#
#   seed = 2025,
#
#   chains = 4, iter = 500, refresh = 0
#
# )
#
#
#
#
#
# set.seed(2025)
#
#
# for (eff in effect_sizes) {
#
#
#   n_group <- sample_sizes / 3
#
#
#   results_eff <- data.frame(
#
#     eff       = numeric(nsim),
#
#     sim       = integer(nsim),
#
#     bf        = numeric(nsim),
#
#     Rhat_max = numeric(nsim),
#
#     ESS_min  = numeric(nsim),
#
#     check     = numeric(nsim),
#
#     stringsAsFactors = FALSE
#
#   )
#
#
#   for (i in seq_len(nsim)) {
#
#     #
```

```
#     eta_medicine <- qlogis(0.62 + 0.25)
#     eta_lever      <- qlogis(0.51 + 0.25)
#
#     p_medicine <- plogis(eta_medicine)
#     p_lever      <- plogis(eta_lever)
#
#     p_ref <- (p_lever + p_medicine) / 2
#
#     p_alien <- p_ref - eff
#
#     Alien      <- rbinom(n_group, 1, p_alien)
#     Lever      <- rbinom(n_group, 1, p_lever)
#     Medicine   <- rbinom(n_group, 1, p_medicine)
#
#     df <- data.frame(
#         y = c(Alien, Lever, Medicine),
#         group = factor(
#             rep(c("Alien", "Lever", "Medicine"), each = n_group),
#             levels = c("Alien", "Lever", "Medicine")
#         )
#     )
#
#     contrasts(df$group) <- C
#
#     fit  <- update(fit_template, newdata = df, chains = 4,
# #iter = 2000, refresh = 0)
#     fit0 <- update(fit0_template, newdata = df, chains = 4,
```

```
#iter = 2000, refresh = 0)

#
#      BF <- bayes_factor(
#          bridge_sampler(fit),
#          bridge_sampler(fit0)
#      )

#
#
#      rhats <- rhat(fit)
#      ess     <- neff_ratio(fit)
#
#
#      Rhat_max <- max(rhats, na.rm = TRUE)
#      ESS_min   <- min(ess, na.rm = TRUE)
#
#
#      check1 <- as.numeric(Rhat_max < 1.01 & ESS_min > 0.1)
#
#
#      results_eff[i, ] <- list(
#          eff       = eff,
#          sim       = i,
#          bf        = BF$bf,
#          Rhat_max = Rhat_max,
#          ESS_min  = ESS_min,
#          check     = check1
#      )
#
#  }

#
#  write.csv(
```

```
#     results_eff,
#     file = paste0("effesz_", eff, ".csv"),
#     row.names = FALSE
#   )
# }
```

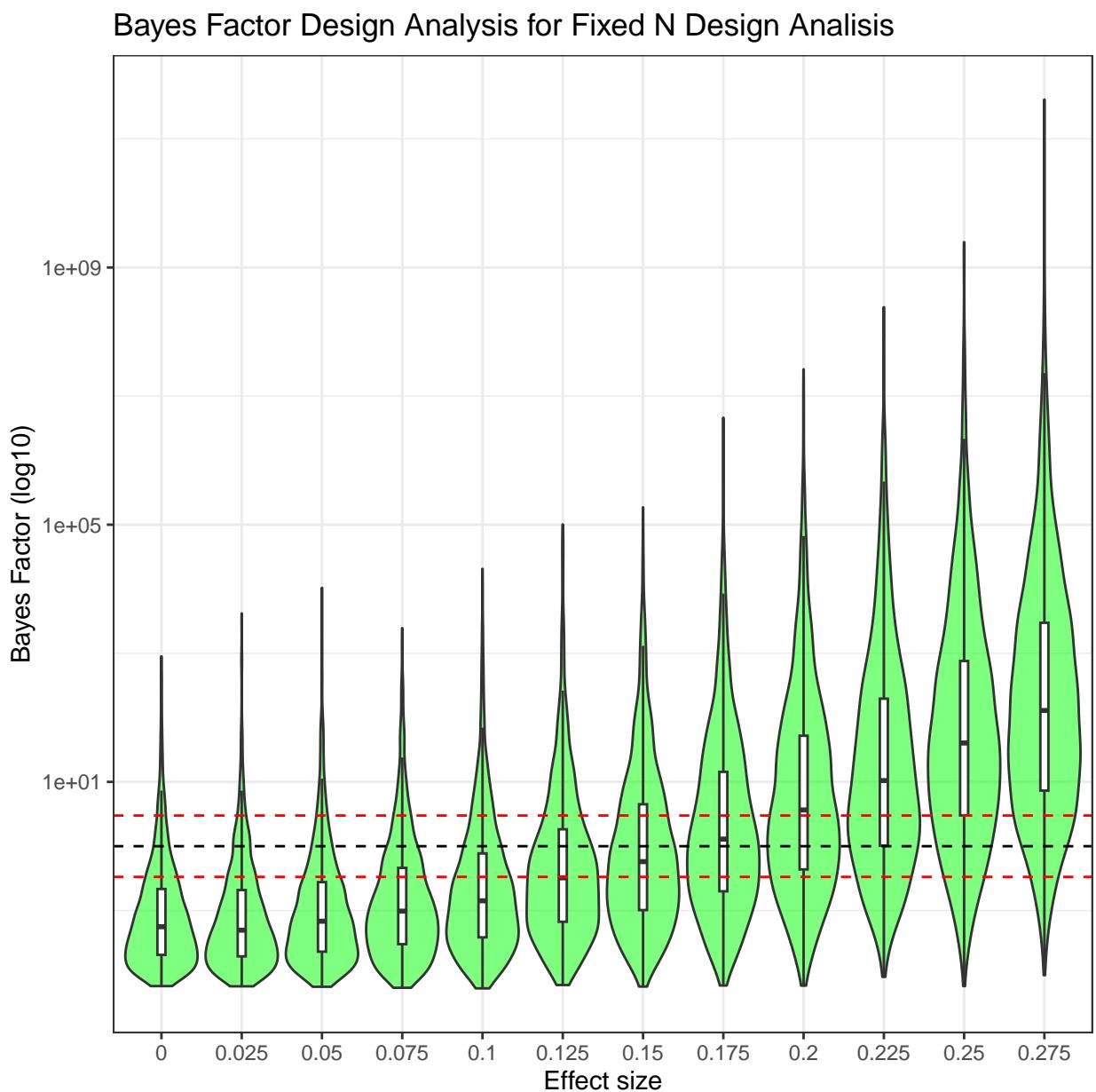
BFDA results were aggregated across all simulated effect sizes. For each effect size, we visualize the distribution of BF across simulations.

```
files <- list.files(
  pattern = "^effesz_.*\\.csv$",
  full.names = TRUE
)

results_all <- do.call(
  rbind,
  lapply(files, read.csv, stringsAsFactors = FALSE)
)

ggplot(results_all, aes(x = factor(eff), y = bf)) +
  geom_violin(fill = "green", alpha = 0.5, scale = "width") +
  geom_boxplot(width = 0.1, outlier.alpha = 0, fill = "white") +
  scale_y_log10() +
  geom_hline(yintercept = 3, linetype = "dashed", color = "red") +
  geom_hline(yintercept = 1, linetype = "dashed", color = "black") +
  geom_hline(yintercept = 1/3, linetype = "dashed", color = "red") +
  labs(
```

```
x = "Effect size",
y = "Bayes Factor (log10)",
title = "Bayes Factor Design Analysis for Fixed N Design Analisis"
) +
theme_bw()
```



```
# Table (Power and FPR)

power_table <- aggregate(
  results_all$bf,
  by = list(Effect_Size = results_all$eff),
  FUN = function(x) mean(x > 3.5)
)

colnames(power_table)[2] <- "Power"

knitr::kable(
  power_table,
  digits = 3)
```

## References

References marked with an asterisk indicate studies included in the meta-analysis.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>

Masked Citation. (n.d.). *Masked Title*.

Schönbrodt, F. D., & Wagenmakers, E. (2016). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 23(2), 254–271.  
<https://doi.org/10.3758/s13423-017-1230-y>

Stan Development Team. (2024). *RStan: The R interface to Stan*. <https://mc-stan.org/>

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E. (2017). A tutorial on bayes factor design analysis using an informed prior. *Behavior Research Methods*, 49(2), 413–428.

<https://doi.org/10.3758/s13428-018-01189-8>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

---

	H1. Alien vs Others	H2. Alien vs Medicine
Alien	1.0	1
Lever	-0.5	0
Medicine	-0.5	-1

---

Cover.Story	Delta.P	Percent.Estimates
4 M	ii. Delta P = 0.5	86
5 L	ii. Delta P = 0.5	76
6 A	ii. Delta P = 0.5	54

**Table 1**

### *Priors over M1*

---

**Effect\_Size    Power**

---

0.000    0.050

0.025    0.044

0.050    0.060

0.075    0.076

0.100    0.120

0.125    0.197

0.150    0.274

0.175    0.380

0.200    0.501

0.225    0.607

0.250    0.736

0.275    0.812

---

---