PU037 Brief Technical Report

Lenard Dome

04 May, 2021

Contents


```
## data.table 1.14.0 using 4 threads (see ?getDTthreads).  Latest news: r-datatable.com

## ************

## Welcome to BayesFactor 0.9.12-4.2. If you have questions, please contact Richard Morey (richarddmorey

##

## Type BFManual() to open the manual.

## ************

##

## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
```

```
##
##     between, first, last
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
##
## Attaching package: 'ggthemes'
## The following object is masked from 'package:cowplot':
##
##     theme_map
```

Method

Participants

42 PPT (21 PPT in each condition).

Stimuli and Category Structures

36 stimuli, grey squares that varied in brightness and size, were displayed on a white background. There were two conditions: unidimensional (UD) and information-integration (ID) category structures. Conditions were counterbalanced, so that UD structures included both horizontal and vertical category boundaries and II structures involved diagonal category boundaries with both positive and negative gradients. Figure 1 shows how these stimuli are distributed in a 2D psychological space.

Training

PPT completed 360 supervised training trials in (3) blocks of 120. In each block, 24 stimuli randomly picked from the original 36 were shown 5 times. In each trial, PPT had to make a category judgement after the stimuli was displayed for 500ms. The stimuli stayed on the screen with Category A and B shown at the bottom of the screen until PPTs made a response. The response deadline was set to 5000ms.

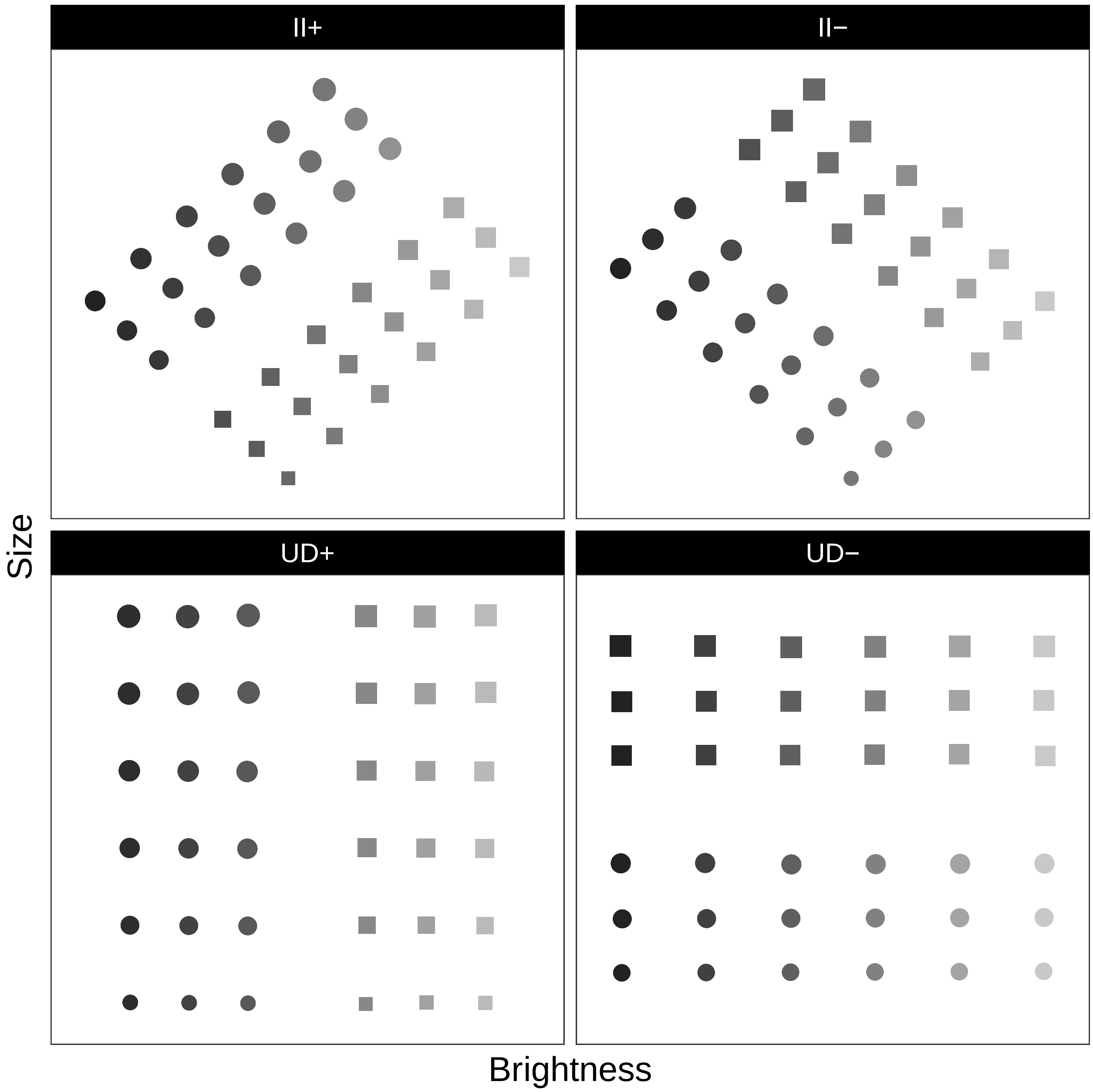


Figure 1: The 2D psychological stimuli space

Recognition Phase

In this phase, PPT judged items as either new or old. This phase included all 36 stimuli in one block and was repeated three times. Each old-new judgement were followed by a confidence rating (1-5 likert scale). The timing was the same as for training trials.

Category Test Phase

Participants needed to make category judgement without feedback of the 36 stimuli. Each stimulus was shown 3 times. The timing and trial structure was the same as training trials - with the exception of the lack of feedback.

Modelling

SUSTAIN's parameters were adjusted to minimize the sum of squared errors between the mean human categorization performance and the mean of SUSTAIN's categorization performance during the test phase. The behaviour of SUSTAIN is not highly sensitive to its parameters, therefore I simultaneously fitted SUSTAIN to all problems. The trial-order was randomized on each iterations with a random seed sampled in $(0, 1000]$. The model was fitted with a differential evolutionary algorithm, as implemented in DEoptim packages. The algorithm reiterated 1000 times to find the best fitting parameters. The speed of crossover was set to $c = 0.5$, which gave larger weights to succesful mutations. Every new population was generated from the top 30% best solution of the previous population. These settings helped to speed up the parameter search and find the best parameter for different trial orders. The best fitting parameters with $SSE = 3.8251454 \times 10^{-10}$ are presented in Table 3.

Table 1: Best fitting parameters for SUSTAIN

	Best Fitting Parameters	Upper Parameter Bounds
attentional focus (r)	4.13012180310834	10
lateral inhibition (beta)	8.3272805517042	10
decision consistency (d)	1.98827913858899	20
learning rate (eta)	0.0626198028612849	0.2

Simulation Analysis

Cluster Recruitment

Table 2: The number of clusters recruited by SUSTAIN

Condition	Mean	SD	Max	Min
II	5.5905	1.189335	12	4
UD	3.0070	1.183068	12	2

The mean number of clusters recruited are $M^c_{ii} = 5.59$, $SD^c_{ii} = 1.20$ for II and $M^c_{ud} = 3.01$, $SD^c_{ud} = 1.18$ for UD. SUSTAIN solves II with a minimum of 4 clusters and a maximum of 12 clusters. SUSTAIN solves UD with a minimum of 2 and a maximum of 12 clusters. Example clusters populating the psychological space are shown in Figure 2.

The number of clusters recruited reflects how trial-order interacts with the following mechanisms: similarity, attention and prediction error.

Simple problems also on average result in fewer clusters, while harder problems require the recruitment of more clusters.

In all cases, SUSTAIN recruits clusters and by proxy constructs category boundaries, and tries to recruit enough clusters so that it can compute similarity to minimize prediction errors given the current trial order.

In a scenario, where SUSTAIN recruits 14 clusters in UD conditions, attentional tuning was moving away from the relevant dimension: it judges stimuli to be closer to the wrong clusters due to their distance on the irrelevant dimension. In II, SUSTAIN adjusts receptive field tunings, so that it pays more attention to one of the dimension, even though both are equally important. In this scenario,

Attentional tuning

```
##
##  Pearson's product-moment correlation
##
## data:  abs(diff) and cluster
## t = 19.332, df = 1998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
```

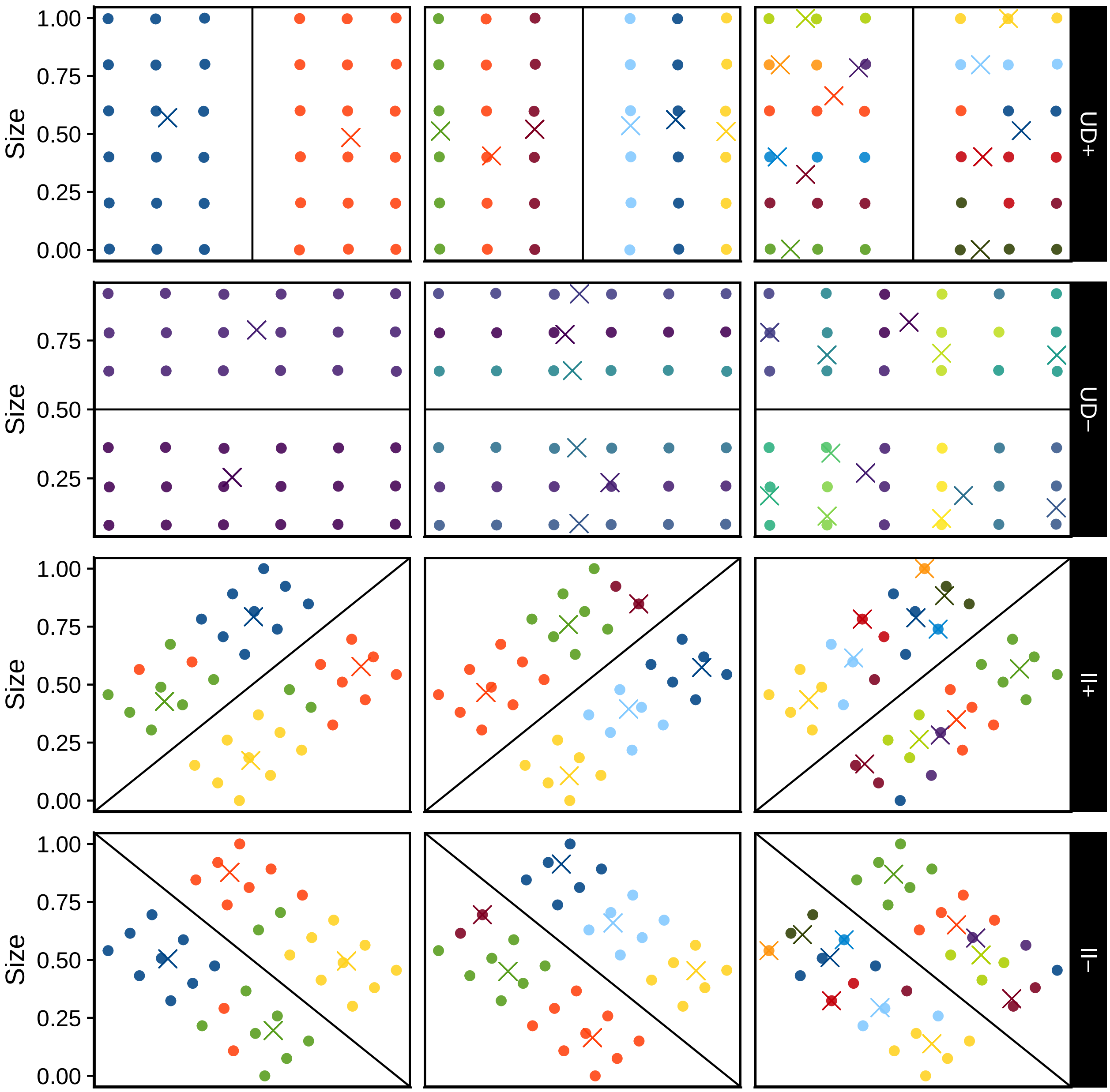
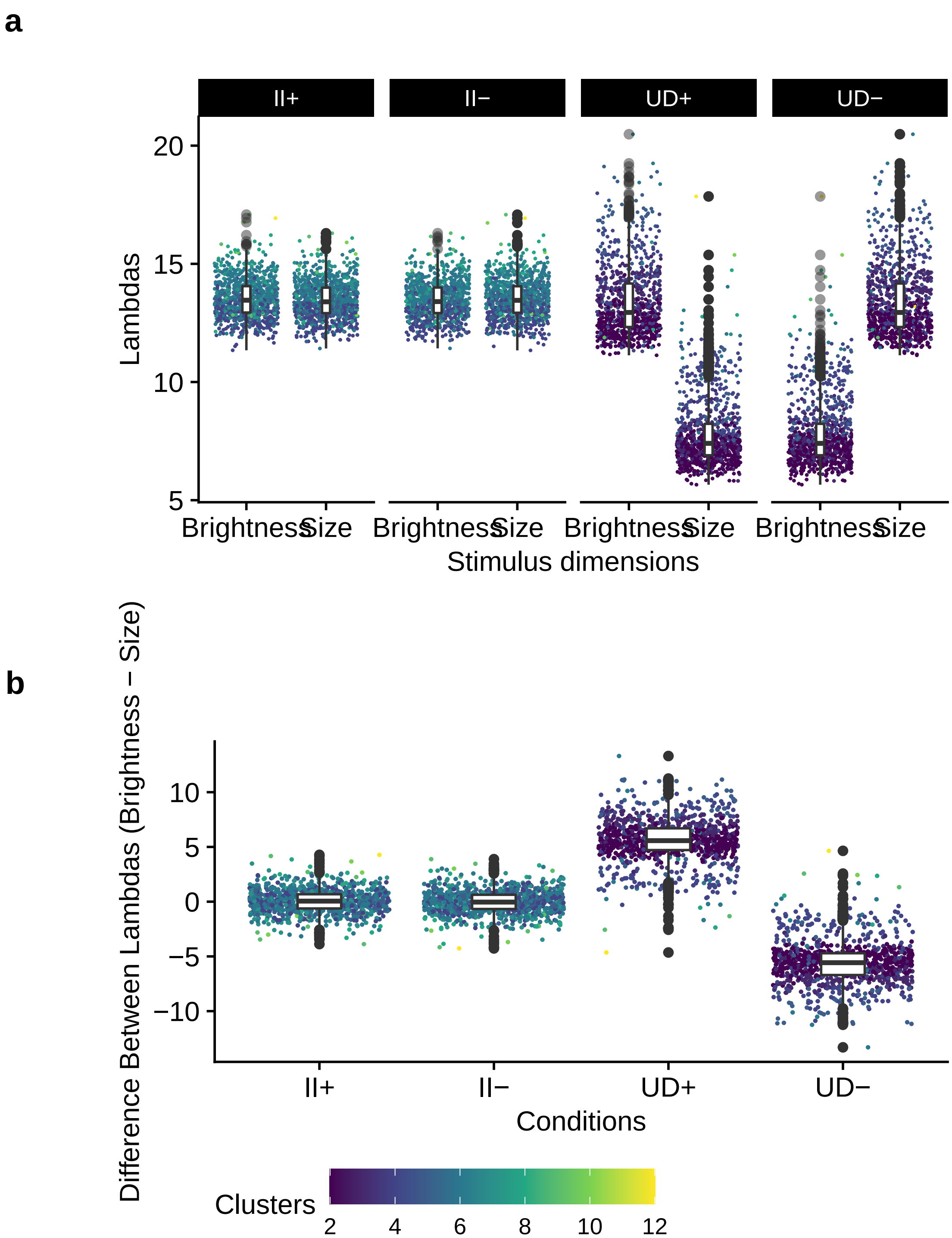


Figure 2: Cluster positions for 4 simulated participants from each condition. Shapes represent categories, while each colour is a separate cluster.

```
## 95 percent confidence interval:
##  0.3593727 0.4332440
## sample estimates:
##      cor
## 0.396951
```

```
## Bayes factor analysis
## --------------
## [1] Alt., r=0.333 : 7.749196e+72 ±0%
##
## Against denominator:
##   Null, rho = 0
## ---
## Bayes factor type: BFcorrelation, Jeffreys-beta*
```



col	II+	II-	UD+	UD-
x	13.5209719	13.4691795	13.378412	7.799121
y	13.4628020	13.5140328	7.799121	13.378412
xd	0.8100859	0.7776072	1.481036	1.408354
yd	0.7786433	0.8119450	1.408354	1.481036

Attentional tuning plays an important role in cluster recruitment. We see that in the II condition, bigger difference between tuning of the receptive fields will result in more clusters being recruited. If SUSTAIN pays more attention to one of the dimensions, it will need to densely populate the psychological space. UD conditions show a similar problem. If the tuning of the receptive fields moves away from relevant dimension to weigh in each equally or put more weigh in on irrelevant dimension, clusters are more densely populate the psychological space. These trends are observable on Figure 4. On the other hand, Figure 5 shows dimensional tuning of the receptive fields as indicated by lambdas for each condition. On both figures, we see that higher number of clusters have ineffective attentional tuning.

Category Test

SUSTAIN's categorization performance is qualitatively similar to what we observed from humans. Performance is worse in II than in UD.

```
## t is large; approximation invoked.
```

```
## Bayes factor analysis
## --------------
## [1] Alt., r=0.707 : 1.883353e+776 ±0%
##
## Against denominator:
##   Null, mu1-mu2 = 0
## ---
## Bayes factor type: BFindepSample, JZS
```

Table 3: Categorization Accuracy

Category Structures	SUSTAIN Mean	SUSTAIN SD	human
II	0.7834079	0.0271481	0.7751323
UD	0.8477706	0.0262727	0.8680556

SUSTAIN's performance is close to humans. SUSTAIN matches human-level performance with a mean difference of NA. See Table 3.

Table 4: Recognition Scores from Equation A6

Category Structure	M_{R}	SD_{R}
II	0.6170649	0.0220347
UD	0.6087618	0.0629418

Recognition

We can see a slightly larger bias to respond with old in the II problem compared to an UD problem, see Table 4. To get an approximate d' measure from recognition scores, I applied Equation A11 from Love and Gureckis (2007) to turn these recognition scores into choice probabilities:

$$P(old) = \frac{R}{R + k}$$

Table 5: Descriptives for d' for both conditions

Conditions	$M^{d'}$	$SD^{d'}$
II	0.0396369	0.0557282
UD	-0.0155043	0.1354422

```
## t is large; approximation invoked.

## Bayes factor analysis

## --------------

## [1] Alt., r=0.707 : 7.057373e+57 ±0%

##

## Against denominator:

##   Null, mu1-mu2 = 0

## ---

## Bayes factor type: BFindepSample, JZS
```

where k is a response threshold paramater. We found that the best-fitting threshold was 0.571.

We calculated the mean probability of a hit $P(H) = P(old \mid item^{old})$ and false alarms $P(F) = P(old \mid item^{new})$ for each participant. Then we continued to determine d' for each participant using the z-transformed $P(H)$ and $P(F)$, see Table 6.

As we see in Table 5, SUSTAIN recreates the superior recognition performance observed in the II by Edmunds et al. (2016), but predicts higher false alarm rates for UD structures. A comparison of d' between SUSTAIN and human data yields a mean difference of 0.0070663.

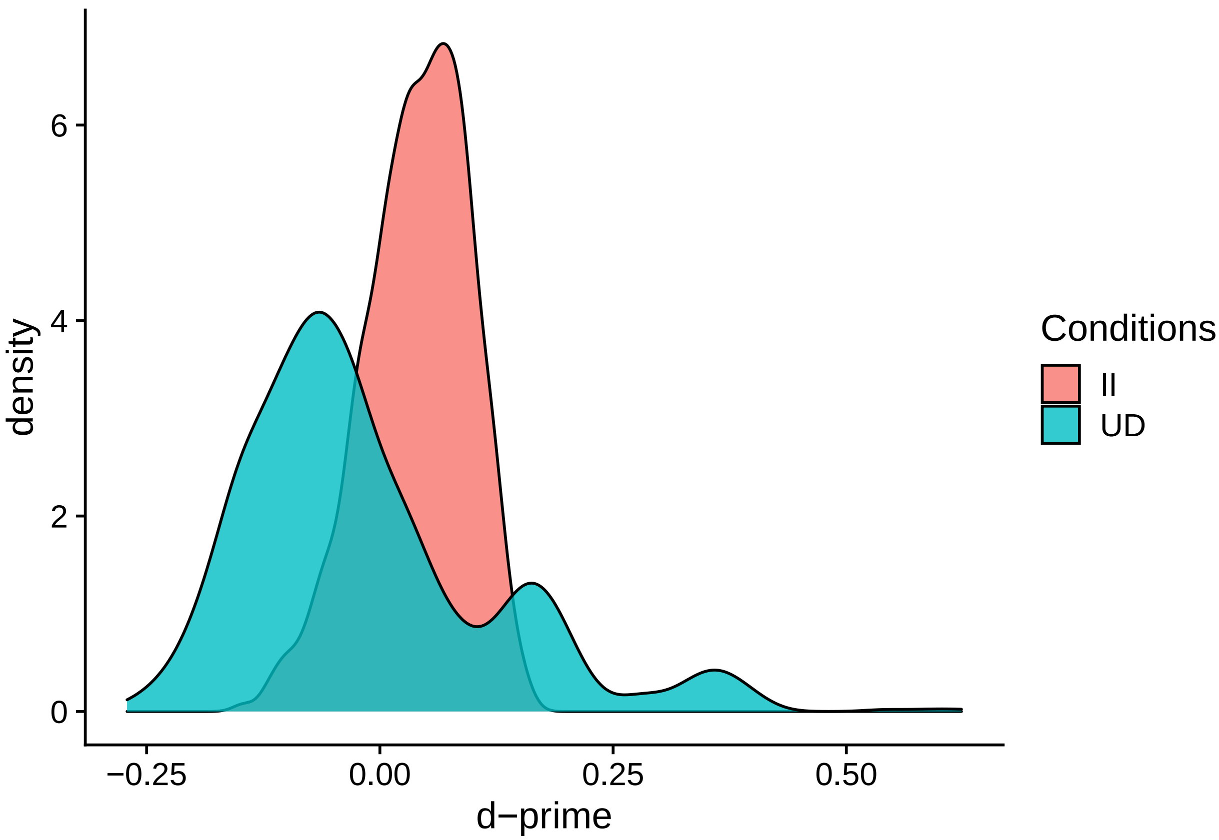


Figure 3: d-prime distributions for each condition.

Conclusion

SUSTAIN can accomodate the results.