Applied Analysis 1

Rohan Hore

August 2024

1 Relativism of moral judgements

Certain behaviours such as plagiarism or copying another student's homework, texting while driving or driving while intoxicated, corruption or abuse of public office for private gain, encouraging or offering support for hate-groups, and so on, are assessed by the majority in today's society as improper or immoral to varying degrees. Morality is, in part, a code of conduct for acceptable behaviour of individuals: see `https://plato.stanford.edu/entries/morality-definition` for further discussion and definitions of concepts. The study described below is concerned with morality judgements or assessments made by individuals about the activities of other individuals in certain settings.

Five skeleton scenarios were constructed that could be tweaked in various ways to highlight certain philosophical concepts. Each scenario is a short piece of text labelled by the protagonist, Driver, Construction worker, Footballer, Firefighter and Train rider, whose activity is to be judged by the assessor. The philosophical concepts to be studied are moral luck (ML), culpable causation (CC) identification (Id), hedonic state (HS), and impulsive versus deliberative action (IvD). For each concept, two versions of each scenario were constructed to highlight the distinction associated with that concept.

For example, the Driver scenario involves an individual who is drunk, falls asleep at the wheel and has a serious accident. Moral luck arises in the contrast between outcomes: in version ML-A, the driver hits a tree and there are no serious injuries; in version ML-B , the driver strikes a pedestrian who dies at the scene. The hedonic state concept arises in the contrast between attitudes following the event: in version HS-A, the driver is remorseful following the accident; in version HS-B, the driver shows no remorse for his actions. Culpable causation arises in the contrast between causes or reasons for the activity: in one version the driver is on his way home to conceal a stash of marijuana from his parents; in the other to conceal a present that he had bought for their wedding anniversary. Moral luck is a contrast solely between outcomes; hedonic state is a contrast between mental attitudes following the event, and culpable causation is a contrast between reasons or purposes leading to the event. Each version is a short piece of text called a vignette, so there are five scenarios, five philosophical concepts, and 25 combinations illustrated by 25 pairs of vignettes.

One of the issues that arises here is the extent to which moral luck, culpable causation and so on, play a role in our assessments of morality. Is driver ML-B judged by society to be morally more reprehensible because his or her actions led to the death of a pedestrian, while the same actions by ML-A having relatively less serious consequences are judged to be less reprehensible? Similar questions arise for culpable causation, and the other philosophical constructs.

For technical discussion of philosophical terms, see the Stanford Encyclopedia of Philosophy, which is available online. The entry for moral luck commences as follows: Moral luck occurs when an agent can be correctly treated as an object of moral judgment despite the fact that a significant aspect of what she is assessed for depends on factors beyond her control. Bernard Williams writes, "when I first introduced the expression moral luck, I expected to suggest an oxymoron (Williams 1993, 251). Indeed, immunity from luck has been thought by many to be part of the very essence of morality. And yet, as Williams (1981) and Thomas Nagel (1979) showed in their now classic pair of articles, it appears that our everyday judgments and practices commit us to the existence of moral luck." The problem of moral luck arises because we seem to be committed to the general principle that we are morally assessable only to the extent that what we are assessed for depends on factors under our control (call this the Control Principle). At the same time, when it comes to countless particular cases, we morally assess agents for things that depend on factors that are not in their control. And making the situation still more problematic is the fact that a very natural line of reasoning suggests that it is impossible to morally assess anyone for anything if we adhere to the Control Principle. (Stanford Encyclopedia of Philosophy `https://plato.stanford.edu/entries/moral-luck/`).

For this study, 1068 participants were recruited for an online survey at the website Amazon Mechanical Turk. Each participant was asked for an assessment of the morality of the protagonist's activity in five pairs of vignettes, i.e., ten morality assessments by each participant. The menu of pairs was selected at random and independently for each participant subject to the constraint that each philosophical concept and each scenario be included exactly once. Morality assessments were made on a 1–7 scale, with 1 labelled 'Not at all immoral' and 7 labelled 'Extremely immoral'. For examination purposes, the scale is to be treated quantitatively.

For a randomly selected subset of 525 participants, labelled joint evaluation, the two versions of each vignette were presented side-by-side on the screen, and the response was entered simultaneously as an ordered pair. The five pairs were presented in random order, which was recorded. For the remaining 543 participants labelled single evaluation, or sequential evaluation, the ten vignettes were presented one at a time sequentially in random order. Participants were not given the opportunity to revise an earlier response. For a discussion of the difference between joint and separate evaluation, see the 1996 paper *The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives* by C.K. Hsee in *Organizational Behavior and Human Decision Processes 67, 247–257*

Data Reading

```r
morality_data=read.table("morality.dat.txt",header=TRUE)
head(morality_data)
```

```
##   subj mode order  scenario concept first Yp Yn age gender  race
## 1    1   je     3 linebacker      hs   neg  5  6  36 Female white
## 2    1   je     1     train     ivd   pos  7  7  36 Female white
## 3    1   je     2      fire      cc   pos  2  6  36 Female white
## 4    1   je     4    driver   ident   pos  7  7  36 Female white
## 5    1   je     5  construct      ml   pos  5  7  36 Female white
## 6    2   se     5      fire   ident   neg  7  6  23   Male white
```

2 Possible Questions

Problem 1

"Is driver ML-B judged by society to be morally more reprehensible because his or her actions led to the death of a pedestrian, while the same actions by ML-A having relatively less serious consequences are judged to be less reprehensible?"– How do you statistically answer this question?

Let $\mu^+_{(ML,D)}, \mu^-_{(ML,D)}$ be correspondingly the mean population response for positive and negative version of (driver, moral luck) vignette pair. The question can be then framed statistically as following hypothesis testing problem

$$H_0 : \mu^+_{(ML,D)} = \mu^-_{(ML,D)} \text{ vs } H_1 : \mu^+_{(ML,D)} < \mu^-_{(ML,D)}$$

Ignoring the assessment mode, we have paired-data responses and hence, we can do a paired t-test which turns out to be significant.

```r
pos_ML=morality_data[morality_data$concept=="ml" & morality_data$scenario=="driver",7]
neg_ML=morality_data[morality_data$concept=="ml" & morality_data$scenario=="driver",8]
t.test(pos_ML,neg_ML,paired=TRUE,alternative="less")
```

```
Paired t-test
data:  pos_ML and neg_ML
t = -10.993, df = 204, p-value < 2.2e-16
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
      -Inf -1.052783
sample estimates:
mean difference
      -1.239024
```

Statement 1: Does the assessor favor the positive version of moral luck scenarios? **Statement 2:** Does the assessor favor the positive version of moral luck stories for all scenarios?

Problem 2

For the moral luck concept only, tabulate the mean differences $Y_p - Y_n$ by assessment mode and scenario. Compute an approximate standard error, and summarize what this table tells you about the effect of assessment mode on moral-luck judgements.

```
SE=function(x) sd(x)/sqrt(length(x))
res=aggregate(Yp-Yn~scenario+mode, data = morality_data[morality_data$concept=="ml",], FUN= "mean")
colnames(res)[3]="value";dcast(setDT(res), scenario~mode, value.var = c('value'))
res=aggregate(Yp-Yn~scenario+mode, data = morality_data[morality_data$concept=="ml",], FUN= SE )
colnames(res)[3]="value";dcast(setDT(res), scenario~mode, value.var = c('value'))
res=aggregate(Yp-Yn~scenario+mode, data = morality_data[morality_data$concept=="ml",], FUN= length )
colnames(res)[3]="value";dcast(setDT(res), scenario~mode, value.var = c('value'))
```

2-way table for mean differences of assessment for moral luck:

```
##      scenario         je         se
## 1:  construct -0.3423423 -0.9278351
## 2:     driver -1.2300000 -1.2476190
## 3:       fire -0.3412698 -0.3271028
## 4: linebacker -0.3333333 -0.3884298
## 5:      train -0.6626506 -0.4070796
```

2-way table for standard error of differences of assessment for moral luck:

```
##      scenario         je         se
## 1:  construct 0.09788550 0.13922298
## 2:     driver 0.17047468 0.14934592
## 3:       fire 0.08060304 0.11482869
## 4: linebacker 0.10540926 0.09416291
## 5:      train 0.13086335 0.09376684
```

2-way table for sample size for assessment of moral luck:

```
##      scenario  je  se
## 1:  construct 111  97
## 2:     driver 100 105
## 3:       fire 126 107
## 4: linebacker 105 121
## 5:      train  83 113
```

Comment: Sample sizes are around 100, the standard errors range from 0.09 to 0.17. Other than construct scenario, we can't see much of an impact of assessment mode on moral luck judgements.

Problem 3

For the moral luck concept only, tabulate the mean differences $Y_p - Y_n$ by assessment mode and first-in-pair. Compute an approximate standard error, and summarize what this table tells you about the effect of assessment mode on moral-luck judgements.

```
res=aggregate(Yp-Yn~first+mode, data = morality_data[morality_data$concept=="ml",], FUN= "mean")
colnames(res)[3]="value";dcast(setDT(res), first~mode, value.var = c('value'))
res=aggregate(Yp-Yn~first+mode, data = morality_data[morality_data$concept=="ml",], FUN= SE )
colnames(res)[3]="value";dcast(setDT(res), first~mode, value.var = c('value'))
res=aggregate(Yp-Yn~first+mode, data = morality_data[morality_data$concept=="ml",], FUN= length )
colnames(res)[3]="value";dcast(setDT(res), first~mode, value.var = c('value'))
```

2-way table for mean differences of assessment for moral luck:

```
##    first         je         se
## 1:   neg -0.4941176 -0.4624060
## 2:   pos -0.6222222 -0.8158845
```

2-way table for standard error of differences of assessment for moral luck:

```
##    first         je         se
## 1:   neg 0.07348292 0.07206931
## 2:   pos 0.07918999 0.08100069
```

2-way table for sample size for assessment of moral luck:

```
##    first  je  se
## 1:   neg 255 266
## 2:   pos 270 277
```

Comment: The sample sizes are around 250, the standard errors are very close to 0.07. For single evaluation mode, whether the positive or negative scenario was been shown first seems to matter while not so much for the joint evaluation mode. That also matches with the general intuition.

Problem 4

For the subjects that were assigned to morality assessment in joint mode, tabulate the two-way tables of mean differences $Y_p - Y_n$ by philosophical concept, by scenario, and by left-right on-screen presentation order.

```
res=aggregate(Yp-Yn~scenario+concept, data = morality_data[morality_data$mode=="je",], FUN= "mean" )
colnames(res)[3]="value";dcast(setDT(res), scenario~concept, value.var = c('value'))
res=aggregate(Yp-Yn~first+concept, data = morality_data[morality_data$mode=="je",], FUN= "mean" )
colnames(res)[3]="value";dcast(setDT(res),concept~first, value.var = c('value'))
res=aggregate(Yp-Yn~first+scenario, data = morality_data[morality_data$mode=="je",], FUN= "mean" )
colnames(res)[3]="value";dcast(setDT(res), scenario~first, value.var = c('value'))
```

Mean difference $Y_p - Y_n$ by scenario and philosophical concept:

```
##      scenario          cc         hs    ident        ivd         ml
## 1:  construct  0.04201681 -0.5600000 2.600000 0.15000000 -0.3423423
## 2:     driver -1.74285714 -0.3272727 3.063158 0.38260870 -1.2300000
## 3:       fire -1.47058824 -2.6105263 2.185185 1.04255319 -0.3412698
## 4: linebacker -2.41052632 -1.4247788 2.839286 0.16000000 -0.3333333
## 5:      train -1.21153846 -0.5140187 2.026087 0.03448276 -0.6626506
```

Mean difference $Y_p - Y_n$ by concept and screen position:

```
##    concept         neg        pos
## 1:      cc -1.1239669 -1.4522968
## 2:      hs -1.0224719 -1.0968992
## 3:   ident  2.7790262  2.2596899
## 4:     ivd  0.5078740  0.1771218
## 5:      ml -0.4941176 -0.6222222
```

Mean difference $Y_p - Y_n$ by scenario and screen position:

```
##      scenario         neg        pos
## 1:  construct  0.47656250  0.1895911
## 2:     driver  0.21886792 -0.2500000
## 3:       fire  0.02008032 -0.4057971
## 4: linebacker -0.06130268 -0.2840909
## 5:      train  0.12204724 -0.1107011
```

Problem 5

For the subjects that were assigned to morality assessment in single mode, tabulate the two-way tables of mean difference $Y_p - Y_n$ by philosophical concept, by scenario and by within-pair order of presentation.

```
res=aggregate(Yp-Yn~scenario+concept, data = morality_data[morality_data$mode=="se",], FUN= "mean" )
colnames(res)[3]="value";dcast(setDT(res), scenario~concept, value.var = c('value'))
res=aggregate(Yp-Yn~first+concept, data = morality_data[morality_data$mode=="se",], FUN= "mean" )
colnames(res)[3]="value";dcast(setDT(res),concept~first, value.var = c('value'))
res=aggregate(Yp-Yn~first+scenario, data = morality_data[morality_data$mode=="se",], FUN= "mean" )
colnames(res)[3]="value";dcast(setDT(res), scenario~first, value.var = c('value'))
```

Mean difference $Y_p - Y_n$ by scenario and philosophical concept:

```
##      scenario         cc         hs    ident        ivd         ml
## 1:  construct  0.1009174 -0.2843137 2.239316 0.05084746 -0.9278351
## 2:     driver -1.4324324 -0.3032787 2.620370 0.21649485 -1.2476190
## 3:       fire -1.3448276 -1.9322034 1.693878 0.52884615 -0.3271028
## 4: linebacker -1.9700000 -1.4271845 2.384615 0.17647059 -0.3884298
## 5:      train -1.3551402 -0.4081633 1.854369 0.19672131 -0.4070796
```

Mean difference $Y_p - Y_n$ by concept and screen position:

```
##    concept         neg        pos
## 1:      cc -1.1643357 -1.2178988
## 2:      hs -1.0531915 -0.7049808
## 3:   ident  2.0430108  2.3143939
## 4:     ivd  0.2340426  0.2222222
## 5:      ml -0.4624060 -0.8158845
```

Mean difference $Y_p - Y_n$ by scenario and screen position:

```
##      scenario         neg         pos
## 1:  construct  0.36524823  0.21839080
## 2:     driver -0.13494810  0.06299213
## 3:       fire -0.37710438 -0.34959350
## 4: linebacker -0.22480620 -0.12631579
## 5:      train -0.04089219 -0.01824818
```

Problem 6

Comment briefly on the similarities and differences in the summary tables for joint-mode versus single-mode assessment. Are there any differences that appear surprising?

The joint- and single-mode patterns for philosophical concept are fairly similar with large positive values for identity, and large negative values for cc and hs. The patterns for scenario also appear to be quite similar, with a large mean positive difference for the construction-worker scenario, a large mean negative difference for the firefighter and linebacker scenarios, and smaller negative mean differences for the others (driver and train rider).

For the order within pair, the story is much interesting. One can observe for concepts cc, ident and ivd the effect of which order the positive and negative story has been presented matters more in the joint assessment mode. This is against our general interpretation, while for ml and hs concepts the results support our previous intuition and observations. (Is there a way of justifying this behaviour?) In the joint assessment mode, one can observe the mean differences are higher if negative scenario is shown earlier. Does this imply the assessor somehow favored the "right" vignette?

Problem 7

Each respondent i provided one pair of assessments for each of the matched pairs of vignettes presented. That means five differences $Y_p - Y_n$ for each respondent. Under what sort of statistical assumptions is it reasonable to treat the differences as independent random variables?

If we associate with each assessor i a random variable α_i such that the response is additive, i.e.

$$Y_i^+(x) = \alpha_i + u^+(x) + \epsilon_i^+(x), \qquad Y_i^-(x) = \alpha_i + u^-(x) + \epsilon_i^-(x)$$

where x denotes a vignette pair, where α_i is main effect for respondent i, $u^+(x), u^-(x)$ correspondingly denote the main effect of positive and negative story of vignette pair. Then the difference in responses can be expressed as

$$Y_i^+(x) - Y_i^-(x) = u^+(x) - u^-(x) + \epsilon_i^+(x) - \epsilon_i^-(x)$$

which is independent of α_i. Thus as long as $(\epsilon_i^+(x), \epsilon_i^-(x))$ is independent across different vignette pairs, we can treat differences to be independent under this simple additive model.

Problem 8

For the joint-evaluation design, let $n = 2625$ be the number of vignette-pair assessments, let D be the n-component vector of differences, and let $\mathcal{X} \subset \mathbb{R}^n$ be the subspace.

$$\mathcal{X} = \text{scenario * concept + order + left-panel+age}$$

where scenario, concept and order are five-level qualitative factors, left-panel is a binary factor, and age is a quantitative variable. Fit the additive iid random-effects model in which the mean lies in $\mathcal{X}$, and

$$\text{cov}(D) = \sigma_0^2 I_n + \sigma_1^2 B$$

is a linear combination of two matrices in which B is the block factor, or indicator matrix, for assessors. Report the REML variance-component estimates (not necessarily positive). Report also the fitted regression coefficients for age and left-panel with standard errors.

Repeat this analysis for the vector of sums $Y_p + Y_n$. Comment on any major differences between these analyses.

```
joint_data=morality_data[morality_data$mode=="je",]
attach(joint_data);order=as.factor(order)
#difference model
fit_D<-lmer(Yp-Yn~scenario*concept+order+first+age+(1|subj),REML=TRUE)
summary(fit_D)
```

```r
#sum model
fit_S<-lmer(Yp+Yn~scenario*concept+order+first+age+(1|subj),REML=TRUE)
summary(fit_S)
```

Results for the difference model

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Yp - Yn ~ scenario * concept + order + first + age + (1 | subj)
##
## REML criterion at convergence: 9589.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0435 -0.4037  0.0644  0.4634  4.2273
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subj     (Intercept) 0.00     0.00
##  Residual             2.22     1.49
## Number of obs: 2625, groups:  subj, 525
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                      0.107153   0.176390   0.607
## firstpos                        -0.237031   0.058432  -4.057
## age                              0.002960   0.002514   1.177
```

Results for the Sum model

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Yp + Yn ~ scenario * concept + order + first + age + (1 | subj)
##
## REML criterion at convergence: 12700.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4586 -0.5679  0.0915  0.5880  3.0124
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subj     (Intercept) 1.404    1.185
##  Residual             6.346    2.519
## Number of obs: 2625, groups:  subj, 525
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                      7.207515   0.347626  20.734
## firstpos                        -0.037267   0.104413  -0.357
## age                             -0.002957   0.006157  -0.480
```

Comment: Linear effect of screen positioning is stronger for the sum compared to that for difference, which is natural. The estimated linear effect of age is similar in both models, as expected. The $\hat{\sigma}_1^2$ is bigger for the sum model compared to difference model, as intuitively argued in the previous problem.

Problem 9

If we consider the additive random-effects model directly to the $2n$ undifferenced assessments, what do you expect the correlation to be for two assessments made by the same subject? Give a numerical value and explain your reasoning.

Under the additive random-effects model for the Y-values with variance-component parameters $(\sigma_\epsilon^2, \sigma_b^2)$ i.e.

$$Y_i^+(x) = \alpha_i + u^+(x) + \epsilon_i^+(x), \quad Y_i^-(x) = \alpha_i + u^-(x) + \epsilon_i^-(x); \quad var(\epsilon_i^+(x)) = var(\epsilon_i^+(x)) = \sigma_\epsilon^2, \;\; var(\alpha_i) = \sigma_b^2$$

So, correlation is given by $\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\epsilon^2}$. One solution would be to fit a general model on whole of $\mathbf{Y}$, and estimate each of these components. Hence, $\hat{\rho} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2} = 0.26$

```r
joint_Y_data=rbind(joint_data,joint_data)
joint_Y_data=cbind(joint_Y_data,c(joint_data$Yp,joint_data$Y_n))
colnames(joint_Y_data)[12]="Y_val"
fit_Y<-lmer(Y_val~scenario*concept+as.factor(order)+first+age+(1|subj),
            data=joint_Y_data,REML=TRUE)
summary(fit_Y)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Y_val ~ scenario * concept + as.factor(order) + first + age +      (1 | subj)
##    Data: joint_Y_data
##
## REML criterion at convergence: 18734.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4141 -0.5538  0.0552  0.5867  3.1393
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subj     (Intercept) 0.5043   0.7102
##  Residual             1.7927   1.3389
## Number of obs: 5250, groups:  subj, 525
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                     3.637e+00  1.523e-01  23.881
## firstpos                       -1.437e-01  4.021e-02  -3.574
## age                            -1.899e-05  3.111e-03  -0.006
```

Another solution is to use the model summaries from last question (the difference and the sum model): The variance of a within-subject difference $Var(Y_i^+(x) - Y_i^-(x)) = Var(\epsilon_i^+(x) - \epsilon_i^-(x))$ is $2\sigma_\epsilon^2$, the covariance of non-overlapping differences is zero, the variance of a within-subject sum

$$Var(2\alpha_i + \epsilon_i^+(x) + \epsilon_i^-(x) + u^+(x) + u^-(x)) = Var(2\alpha_i) + Var(\epsilon_i^+(x) + \epsilon_i^-(x)) = 2\sigma_\epsilon^2 + 4\sigma_b^2,$$

and the covariance of two non-overlapping within-subject sums is $4\sigma_b^2$. the approximate least-squares solution is $\hat{\sigma_\epsilon}^2 = 1.58$ $\hat{\sigma}_b^2 = 0.57$, implying $\hat{\rho} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2} = 0.26$ for the within-subject correlations.

Problem 10

The pairs of vignettes in this design were coded one pos the other neg, implying that the set of vignettes is a triple Cartesian product.

$$A = \{5 \text{ scenarios}\} \times \{5 \text{ concepts}\} \times \{\text{pos, neg}\}$$

Some of the analyses suggested in previous questions are based on the assumption that the correspondence between the actual levels and the Cartesian product set is natural. In fact, the actual vignette labels for the concept ML are lucky/unlucky, while those for IvD are impulsive/deliberative, and so on, with contrasting pairs of adjectives describing the other concepts. One might guess that lucky has been coded for administrative purposes as pos, and unlucky as neg, but the coding for IvD and other concepts is less obvious, and perhaps even entirely arbitrary. If that is the case, the set of vignettes is a two-fold Cartesian product

$$B = \{5 \text{ scenarios}\} \times \{\{\text{lucky, unlucky}\}, \cdots, \{\text{implusive, deliberative}\}\}$$

where the second factor is a partition of ten adjectives into five pairs, one pair for each concept. In that case, no 1–1 correspondence $A \iff B$ is natural in the sense of preserving Cartesian or other relationships among elements. Show that some of the models that were fitted and tested are not independent of the assignment of labels pos and neg to vignette pairs. Which ones?

Explain the statistical implications of dependence on administrative coding. For the models that are not independent of the label assignment, suggest a modification to remedy the problem.

We saw two linear models, one for the sums and one for the differences, both having the same mean subspace

$$\mathcal{X} = \text{scenario * concept + order + left-panel+age}$$

and the same family of covariance matrices. The pair sums are independent of the coding, so the model for sum is unaffected by the coding if and only if the subspace X and the family of covariance matrices are also independent of the coding.

Unfortunately, the mean subspace includes the term left-panel coded administratively as pos or neg. If all five codes were switched simultaneously, the basis vectors are switched, but the subspace is unaffected. However, if the IvD code alone is switched, which is allowable here, the new mean-value space is not same anymore.

One simple remedy is to omit the offending term left-panel. Another remedy is to replace left-panel with concept:left-panel, so the code-dependent two-dimensional subspace left-panel is expanded to a code-independent ten-dimensional subspace spanned by the indicator functions for the ten adjectives in the second factor in B. The overall increase in the dimension of the mean-value subspace is four.

Problem 11

For the difference vector, are the effects of order, age, sex and race of the assessor appreciable? Start with any reasonable baseline model of your choice for comparison. What are the implications of this for generalizability of the conclusions?

```
joint_data=morality_data[morality_data$mode=="je",]
base_model=lmer(Yp-Yn~scenario+concept+(1|subj),data=joint_data,REML=TRUE)

#age
age_model= lmer(Yp-Yn~scenario+concept+age+(1|subj),data=joint_data,REML=TRUE)
anova(base_model,age_model)
#sex
sex_model= lmer(Yp-Yn~scenario+concept+gender+(1|subj),data=joint_data,REML=TRUE)
anova(base_model,sex_model)
#race
race_model= lmer(Yp-Yn~scenario+concept+race+(1|subj),data=joint_data,REML=TRUE)
anova(base_model,race_model)
#order
order_model= lmer(Yp-Yn~scenario+concept+as.factor(order)+(1|subj),data=joint_data,REML=TRUE)
anova(base_model,order_model)
```

Should we add age?

```
## Data: joint_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
```

```
## age_model: Yp - Yn ~ scenario + concept + age + (1 | subj)
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model   11 9887.1 9951.7 -4932.5   9865.1
## age_model    12 9887.7 9958.2 -4931.9   9863.7 1.3409  1     0.2469
```

Should we add sex?

```
## Data: joint_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
## sex_model: Yp - Yn ~ scenario + concept + gender + (1 | subj)
##             npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model   11 9887.1 9951.7 -4932.5   9865.1
## sex_model    12 9887.5 9958.0 -4931.8   9863.5 1.5811  1     0.2086
```

Should we add race?

```
## Data: joint_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
## race_model: Yp - Yn ~ scenario + concept + race + (1 | subj)
##             npar    AIC     BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model   11 9887.1  9951.7 -4932.5   9865.1
## race_model   24 9899.6 10040.5 -4925.8   9851.6 13.528 13     0.4079
```

Should we add order?

```
## Data: joint_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
## order_model: Yp - Yn ~ scenario + concept + as.factor(order) + (1 | subj)
##              npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model    11 9887.1 9951.7 -4932.5   9865.1
## order_model   15 9891.3 9979.4 -4930.7   9861.3 3.7757  4     0.4372
```

```r
single_data=morality_data[morality_data$mode=="se",]
base_model=lmer(Yp-Yn~scenario+concept+(1|subj),data=single_data,REML=TRUE)

#age
age_model= lmer(Yp-Yn~scenario+concept+age+(1|subj),data=single_data,REML=TRUE)
anova(base_model,age_model)
#sex
sex_model= lmer(Yp-Yn~scenario+concept+gender+(1|subj),data=single_data,REML=TRUE)
anova(base_model,sex_model)
#race
race_model= lmer(Yp-Yn~scenario+concept+race+(1|subj),data=single_data,REML=TRUE)
anova(base_model,race_model)
#order
order_model= lmer(Yp-Yn~scenario+concept+as.factor(order)+(1|subj),data=single_data,REML=TRUE)
anova(base_model,order_model)
```

Should we add age?

```
## Data: single_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
## age_model: Yp - Yn ~ scenario + concept + age + (1 | subj)
##             npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model   11 10023 10088 -5000.3    10001
## age_model    12 10024 10095 -5000.1    10000 0.3954  1     0.5295
```

Should we add sex?

```
## Data: single_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
## sex_model: Yp - Yn ~ scenario + concept + gender + (1 | subj)
##            npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model   11 10023 10088 -5000.3    10001
## sex_model    12 10024 10095 -5000.2    10000 0.1657  1      0.684
```

Should we add race?

```
## Data: single_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
## race_model: Yp - Yn ~ scenario + concept + race + (1 | subj)
##            npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model   11 10023 10088 -5000.3  10000.6
## race_model   26 10041 10194 -4994.3   9988.6 11.988 15     0.6799
```

Should we aff order?

```
## Data: single_data
## Models:
## base_model: Yp - Yn ~ scenario + concept + (1 | subj)
## order_model: Yp - Yn ~ scenario + concept + as.factor(order) + (1 | subj)
##             npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## base_model    11 10023 10088 -5000.3  10000.6
## order_model   15 10024 10113 -4997.2   9994.3 6.2966  4     0.1781
```

For both models, none of the p-values are significant. Hence, there is no statistical evidence of any effect of these features on the assessment for both modes. However, we didn't study the significance of the features in our base model. These are good signs for generalization purpose, in the sense that we can make similar conclusion for the general population, even outside the amazon survey participants.