P3 analysis

Mark Whiting

2/26/2020

We first build several core data frames:

- **Tasks** lists the 11 tasks we studied and their IDs
- **Questions** lists the 27 questions, their IDs and the answer type they use
- **Responses** lists 407 total responses, broken out by task-respondant pairs, with questions as columns. These encompass responses from 37 unique individuals.

To evaluate the inter rater reliability, we compute Fleiss' Kappa

-0.03, NaN, -0.03

and Krippendorf's Alpha

-0.03, 1, -0.03,

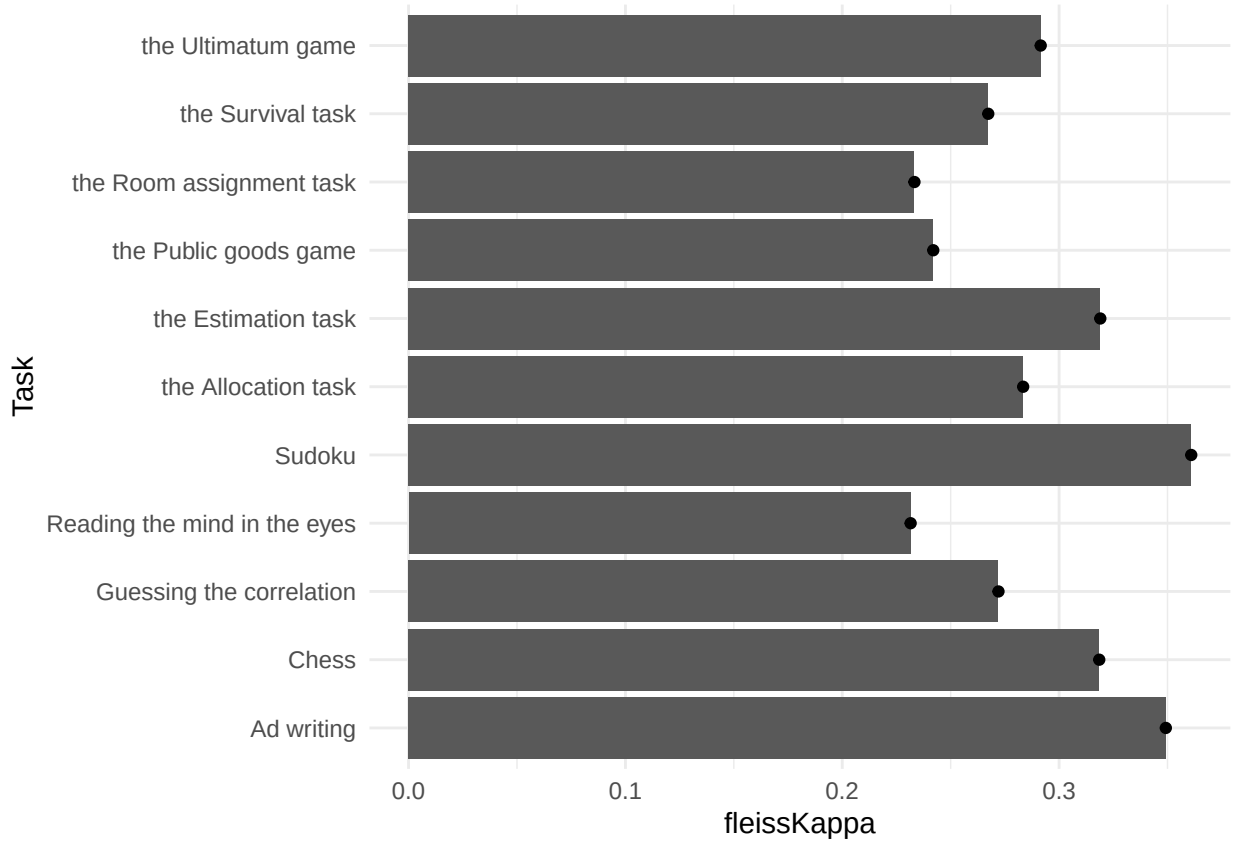
-0.03, -0.03

for each task-question pair (a 26, 11 matrix, because one question was dropped due to being a free text response). However, this does not give us what we want as each position in these results is calculated from a 1 dimensional vector, which is not how these statistics are intended to be used. All this tells us is that

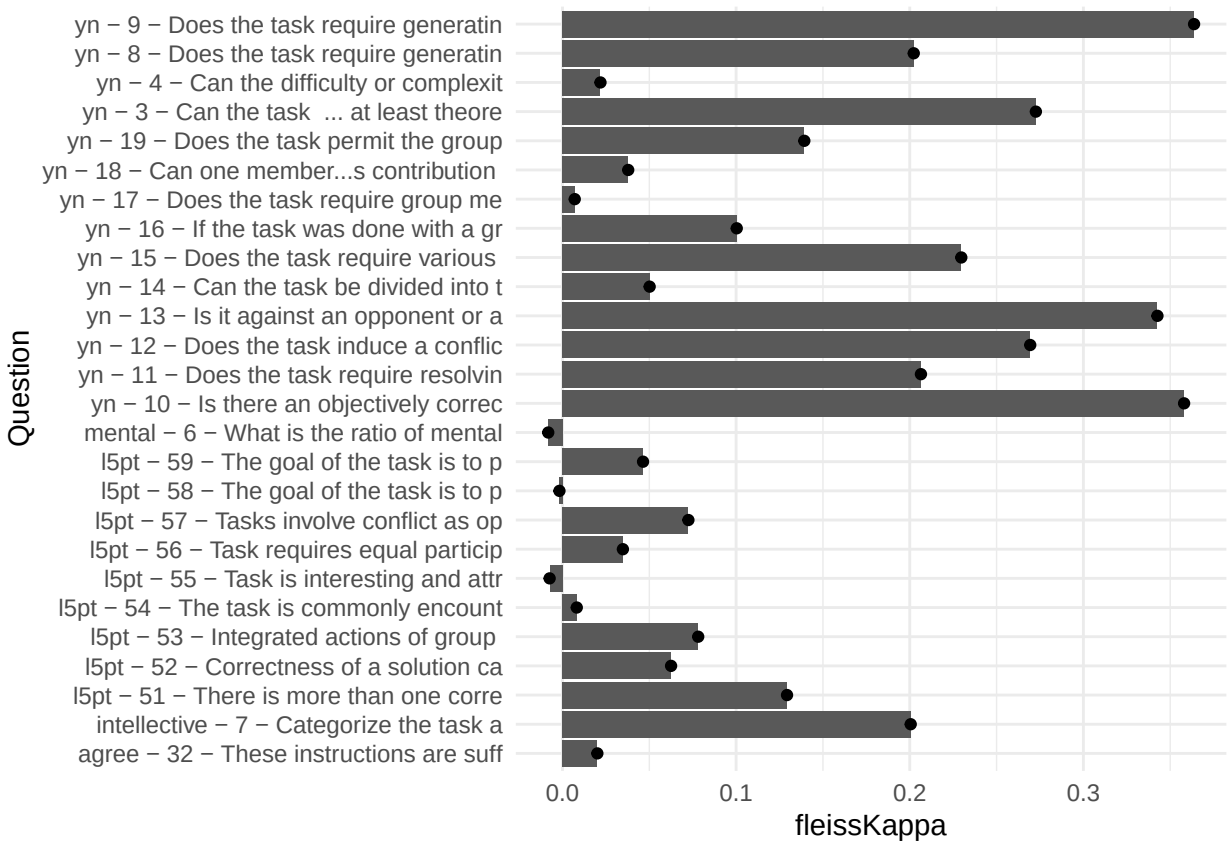
Note that in Fleiss' Kappa, `NA` *for a 1 dimensional vector is similar to 1, as in no variance across respondants. Kripendorf's Alpha does this correctly and shows 1 in those positions.*

So, we'd rather do this for subsets of the data, e.g. for each task, find the agreemt across raters and questions, or for each question, find the agreement across raters and tasks, or even, for each rater, find the agremeent across tasks and questions. In the next section, these assesments are performed.

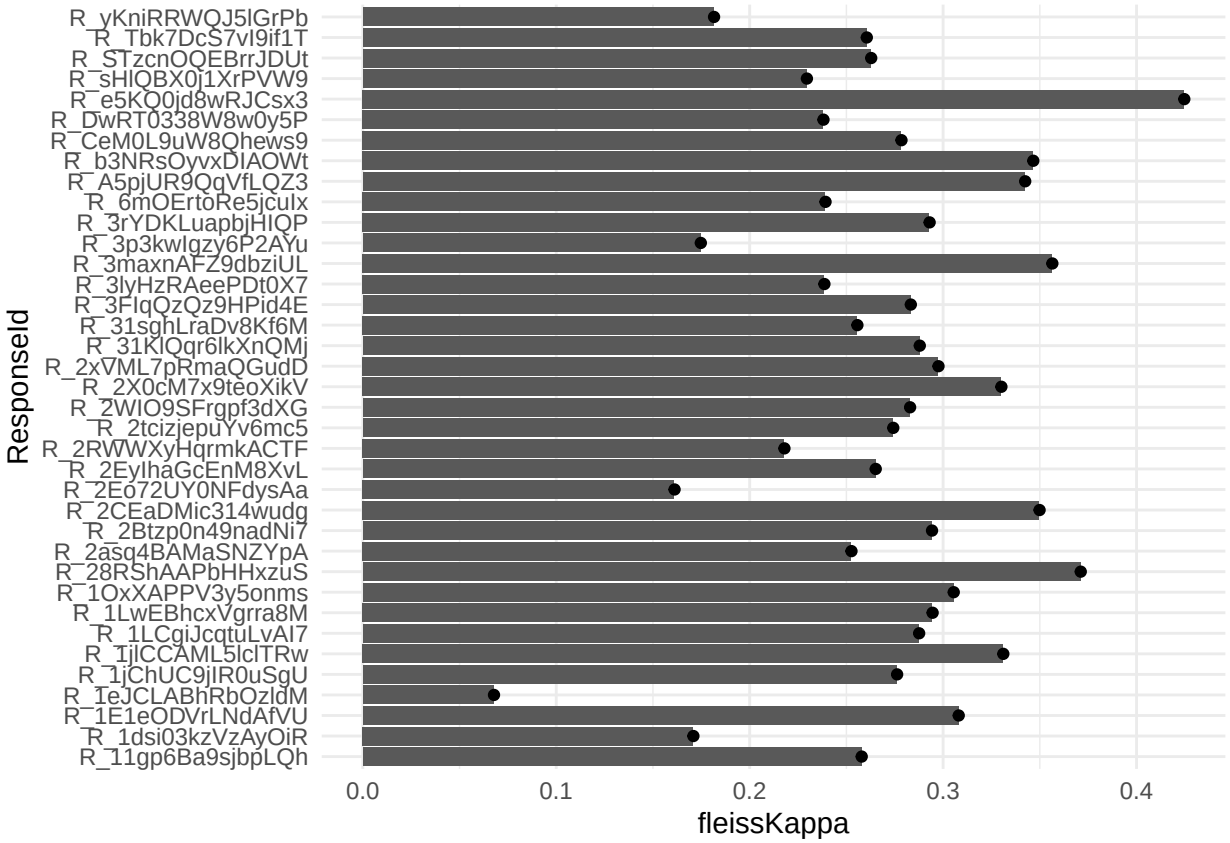
```
ggplot(perTaskAgreement,aes(Task,fleissKappa)) +
  geom_col() +
  geom_point(aes(x = perTaskAgreement$Task, y = perTaskAgreement$kripAlpha)) +
  coord_flip() +
  theme_minimal()
```




```
ggplot(perQuestionAgreement,aes(Question,fleissKappa)) +
  geom_col() +
  geom_point(aes(x = perQuestionAgreement$Question, y = perQuestionAgreement$kripAlpha)) +
  scale_x_discrete(labels=substring(perQuestionAgreement$Question,0,40)) +
  coord_flip() +
  theme_minimal()
```




```
ggplot(perRaterAgreement,aes(ResponseId,fleissKappa)) +
  geom_col() +
  geom_point(aes(x = perRaterAgreement$ResponseId, y = perRaterAgreement$kripAlpha)) +
  coord_flip() +
  theme_minimal()
```

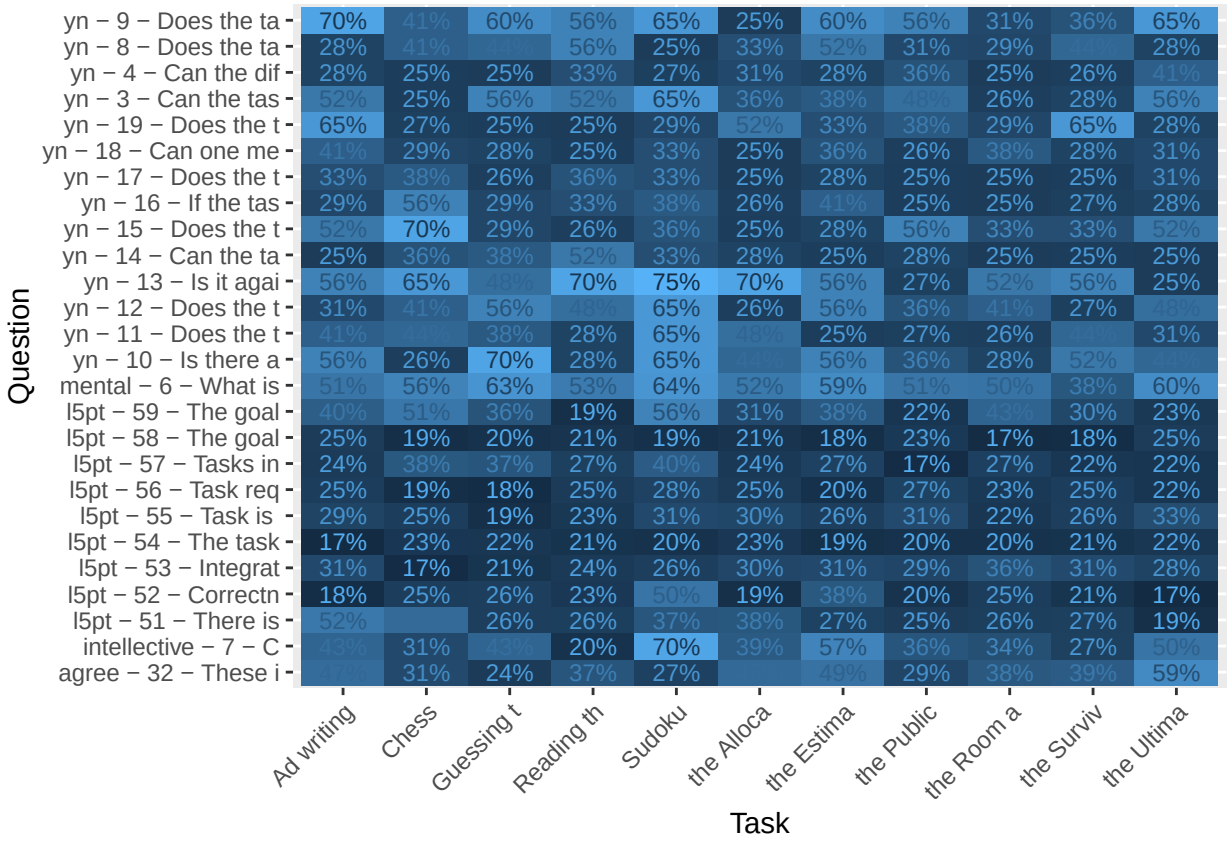



```
# Expanding the 'per question' design
mine = kappam.fleiss(
  (responses %>% select(questionNames[26],Task,ResponseId) %>% spread(ResponseId,questionNames[26])),
  detail = TRUE)$detail

as.data.frame(mine) %>%
  filter(Var2 == "Kappa") %>%
  mutate(
    Task = Var1,
    fleissKappa = Freq
  )
```

```
##                           Var1  Var2   Freq                       Task
## 1                   Ad writing Kappa -0.002                   Ad writing
## 2                        Chess Kappa -0.002                        Chess
## 3     Guessing the correlation Kappa -0.002    Guessing the correlation
## 4                           No Kappa  0.376                          No
## 5  Reading the mind in the eyes Kappa -0.002 Reading the mind in the eyes
## 6                       Sudoku Kappa -0.002                       Sudoku
## 7          the Allocation task Kappa -0.002         the Allocation task
## 8          the Estimation task Kappa -0.002         the Estimation task
## 9        the Public goods game Kappa -0.002       the Public goods game
## 10    the Room assignment task Kappa -0.002    the Room assignment task
## 11           the Survival task Kappa -0.002           the Survival task
## 12         the Ultimatum game Kappa -0.002         the Ultimatum game
## 13                         Yes Kappa  0.397                         Yes
##    fleissKappa
```

```
## 1       -0.002
## 2       -0.002
## 3       -0.002
## 4        0.376
## 5       -0.002
## 6       -0.002
## 7       -0.002
## 8       -0.002
## 9       -0.002
## 10      -0.002
## 11      -0.002
## 12      -0.002
## 13       0.397
```

```r
# percentage of agreements for each qestion for each task - agreement at a random guess
ourMeasure <- responses %>%
  select(-ResponseId) %>%
  group_by(Task) %>%
  summarize_each(function(col) {
    agreement = mean((outer(col,col,function(x, y) as.integer(x==y)) %*% replicate(n(), 1)) / n())
    random = 1/nlevels(col) ** 2
    return(agreement - random)
  })  %>% gather(Question,Agreement,-Task)

ggplot(ourMeasure, aes(Task,Question,fill = Agreement)) +
  geom_tile() +
  geom_text(aes(label=paste(round(Agreement*100,0),"%",sep = ""),color=-Agreement),size=3) +
  scale_y_discrete(labels=substring(unique(ourMeasure$Question),0,20)) +
  scale_x_discrete(labels=substring(unique(ourMeasure$Task),0,10)) +
  theme(legend.position="none",axis.text.x = element_text(angle = 45, hjust = 1))
```


Question
yn − 9 − Does the ta
yn − 8 − Does the ta
yn − 4 − Can the dif
yn − 3 − Can the tas
yn − 19 − Does the t
yn − 18 − Can one me
yn − 17 − Does the t
yn − 16 − If the tas
yn − 15 − Does the t
yn − 14 − Can the ta
yn − 13 − Is it agai
yn − 12 − Does the t
yn − 11 − Does the t
yn − 10 − Is there a
mental − 6 − What is
l5pt − 59 − The goal
l5pt − 58 − The goal
l5pt − 57 − Tasks in
l5pt − 56 − Task req
l5pt − 55 − Task is
l5pt − 54 − The task
l5pt − 53 − Integrat
l5pt − 52 − Correctn
l5pt − 51 − There is
intellective − 7 − C
agree − 32 − These i
70% 41% 60% 56% 65% 25% 60% 56% 31% 36% 65%
28% 41% 44% 56% 25% 33% 52% 31% 29% 44% 28%
28% 25% 25% 33% 27% 31% 28% 36% 25% 26% 41%
52% 25% 56% 52% 65% 36% 38% 48% 26% 28% 56%
65% 27% 25% 25% 29% 52% 33% 38% 29% 65% 28%
41% 29% 28% 25% 33% 25% 36% 26% 38% 28% 31%
33% 38% 26% 36% 33% 25% 28% 25% 25% 25% 31%
29% 56% 29% 33% 38% 26% 41% 25% 25% 27% 28%
52% 70% 29% 26% 36% 25% 28% 56% 33% 33% 52%
25% 36% 38% 52% 33% 28% 25% 28% 25% 25% 25%
56% 65% 48% 70% 75% 70% 56% 27% 52% 56% 25%
31% 41% 56% 48% 65% 26% 56% 36% 41% 27% 48%
41% 44% 38% 28% 65% 48% 25% 27% 26% 44% 31%
56% 26% 70% 28% 65% 44% 56% 36% 28% 52% 44%
51% 56% 63% 53% 64% 52% 59% 51% 50% 38% 60%
40% 51% 36% 19% 56% 31% 38% 22% 43% 30% 23%
25% 19% 20% 21% 19% 21% 18% 23% 17% 18% 25%
24% 38% 37% 27% 40% 24% 27% 17% 27% 22% 22%
25% 19% 18% 25% 28% 25% 20% 27% 23% 25% 22%
29% 25% 19% 23% 31% 30% 26% 31% 22% 26% 33%
17% 23% 22% 21% 20% 23% 19% 20% 20% 21% 22%
31% 17% 21% 24% 26% 30% 31% 29% 36% 31% 28%
18% 25% 26% 23% 50% 19% 38% 20% 25% 21% 17%
52% 26% 26% 37% 38% 27% 25% 26% 27% 19%
43% 31% 43% 20% 70% 39% 57% 36% 34% 27% 50%
47% 31% 24% 37% 27% 49% 29% 38% 39% 59%
Ad writing
Chess
Guessing t
Reading th
Sudoku
the Alloca
the Estima
the Public
the Room a
the Surviv
the Ultima
Task