CF_MPT

Stefano Dalla Bona

2024-04-04

1. Aim

This study aims to unravel the cognitive processes underlying the conjunction fallacy (CF). Numerous theories have been proposed to explain the CF, some of which may yield redundant predictions, suggesting functional similarity. What I find intriguing is the possibility that the various processes posited by different theories may each hold validity. For example, individuals may reason about the task in diverse ways. To disentangle these processes we will employ a Multinomial Processing Tree (MPT) model.

2. Experimental Procedure

The conjunction fallacy (CF) task highlights a common challenge people face in correctly assessing that the probability of the conjunction of two events is greater than the probability of a single event, as expressed by $P(A \wedge B) < P(A)$. The task can be presented in various forms, with one of the most common examples involving Linda, as described below:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Participants are then asked to determine which of the following statements is more probable:

- *Linda is a bank teller* (Event A)
- *Linda is a bank teller and is active in the feminist movement* (Events A and B)

The correct response is the first alternative. It's important to note that the task is brief, typically requiring only 1 to 2 minutes to complete.

In some versions of the CF task, the task is modified to include both Event A and Event B, with participants asked to rank the probabilities of each statement:

- 1. *Linda is a bank teller* (Event A, the most probable between the two)
- 2. *Linda is active in the feminist movement* (Event B, the least probable between the two)
- 3. *Linda is a bank teller and is active in the feminist movement* (Events A and B)

Tversky et Kahneman (1983) have shown that participants often rank the conjunction statement above the less probable Event B statement, a phenomenon observed in approximately 85% of cases. For our study, we will utilize a version of the task that includes Event B and employs a ranking system.

3. Development

3.1 Categories

To fit our data in a MPT model, responses have to be intended as categorical data. As our task ask people to rank alternatives, the total number of categories will be equal to the permutations of the three categories $(3! = 6)$. For instance:

- $Ranking = (1, 2, 3)$ denoted as $Correct_{123}$, where the conjunction of Event A and Event B occupies the correct placement.

- $Ranking = (2, 1, 3)$ denoted as $Correct_{213}$, where the conjunction of Event A and Event B occupies the correct placement.
- $Ranking = (1, 3, 2)$ denoted as $SingleError_{132}$, where the conjunction of Event A and Event B occupies the placement adjacent to the correct one.
- $Ranking = (2, 3, 1)$ denoted as $SingleError_{231}$, where the conjunction of Event A and Event B occupies the placement adjacent to the correct one.
- $Ranking = (3, 1, 2)$ denoted as $DoubleError_{312}$, where the conjunction of Event A and Event B occupies two placements away from the correct one.
- $Ranking = (3, 2, 1)$ denoted as $DoubleError_{321}$, where the conjunction of Event A and Event B occupies two placements away from the correct one.

3.2 Predictions for the model

- The representative heuristic suggests that individuals tend to perceive the most representative case as the most probable when engaging in intuitive reasoning. In this context, Event A is considered the most representative, and the conjunction of Events A and B (which includes Event A) is perceived as more representative than Event B alone. This tendency leads to the prediction that people will commit a $SingleError_{132}$, when employing this heuristic reasoning. This prediction is operationalized with the h probability parameter in the model.
- Alternatively, if individuals do not employ a heuristic system, they may utilize a more analytical approach with a probability of $(1 - h)$. In this case, they may attempt to derive a response based on the application of mathematical rules, resulting in the correct response $Correct_{123}$.
- Individuals may also rely on other mathematical rules, which are formally incorrect. This prediction is operationalized with the s probability parameter in the model. For example, individuals may confuse the conjunction rule with the disjunction rule, operationalized with the p probability parameter in the model, resulting in a $DoubleError_{312}$. Alternatively, they may use the signed summation rule $(1 - p)$, where the A event is assigned a positive number, the B event a negative number, and the conjunction rule is perceived as the summation of the two numbers, resulting in a $SingleError_{132}$.
- Furthermore, individuals may not fully understand the question or may lack motivation during the study, leading to random guessing among the six rankings. This constitutes the error component of the model, operationalized with the $(1 - t)$ probability in the model, where t represents the probability parameter of "thinking" during the experiment.

3.3 Conditions and hypothesis

We propose that in a second condition where the application of a mathematical rule is more apparent, individuals are less likely to rely on heuristic processes and are more inclined to apply a mathematical rule.

```r
Conditions <- c("Heuristic", "Mathematical")

Parameter_Difference <- c("h", "h-d")

data <- data.frame(Conditions,Parameter_Difference)

knitr::kable(data)
```

Conditions	Parameter_Difference
Heuristic	h
Mathematical	h-d

Therefore, the parameter h for the heuristic process is expected to be significantly reduced $(h - d)$ in the "Mathematical" condition.

- *H0: h probability in the heuristic condition = h probability in the mathematical condition*
- *H1: h probability in the heuristic condition > h probability in the mathematical condition*

4. MPT model, equations, parameters recovery

We developed a flowchart to depict the MPT model. Circles represent categories, while rectangles denote processes.

```
library('DiagrammeR')
grViz("digraph flowchart {
      graph[layout = dot, rankdir = LR]
      node [fontname = Helvetica, shape = rectangle]
      tab1 [label = '@@1']
      tab2 [label = '@@2']
      tab3 [label = '@@3']
      tab4 [label = '@@4']
      tab5 [label = '@@5']
      tab6 [label = '@@6']
      tab7 [label = '@@7']
      tab8 [label = '@@8']
      tab9 [label = '@@9']
      node [fontname = Helvetica, shape = circle]
      tab10 [label = '@@10']
      tab11 [label = '@@11']
      tab12 [label = '@@12']
      tab13 [label = '@@13']
      tab14 [label = '@@14']
      tab15 [label = '@@15']
      tab1 -> tab2 [label = 't'];
      tab1 -> tab3 [label = '(1-t)'];
      tab2 -> tab4 [label = 'h'];
      tab2 -> tab5 [label = '(1-h)'];
      tab5 -> tab6 [label = 's'];
      tab5 -> tab7 [label = '(1-s)'];
      tab6 -> tab8 [label = 'p'];
      tab7 -> tab10[label = '1'];
      tab9 -> tab13[label = '1'];
      tab4 -> tab13[label = '1'];
      tab8 -> tab14[label = '1'];
      tab6 -> tab9 [label = '(1-p)'];
      tab3 -> tab10[label = '(1/6)'];
      tab3 -> tab11[label = '(1/6)'];
      tab3 -> tab12[label = '(1/6)'];
      tab3 -> tab13[label = '(1/6)'];
      tab3 -> tab14[label = '(1/6)'];
      tab3 -> tab15[label = '(1/6)']
      }
      [1]: 'Thinking or Guessing?'
      [2]: 'Heuristic reasoning'
      [3]: 'Guessing'
      [4]: 'Rappresentative Heuristic'
      [5]: 'Analytical reasoning'
```

```
      [6]: 'Mathematical rules'
      [7]: 'Correct Assment of conjunction'
      [8]: 'OR (disjunction'
      [9]: 'Signed Summation'
      [10]: 'Correct(123)'
      [11]: 'Correct(213)'
      [12]: 'Single Error (231)'
      [13]: 'Single Error (132)'
      [14]: 'Double Error (312)'
      [15]: 'Double Error (321)'

      ")
```

The equations for each category can be formulated by summing the equation paths that lead to the same category. These equations represent the MPT model. It's important to note that while we derived these equations from the graph we built, it's possible for the same equations to result in different trees. The model parameters are probability values denoted by t, h, s and p.

$Correct_{123} = (1-t) \times \frac{1}{6} + t \times (1-h) \times (1-s)$

$Correct_{213} = (1-t) \times \frac{1}{6}$

$SingleError_{132} = (1-t) \times \frac{1}{6} + t \times h + t \times (1-h) \times s \times (1-p)$

$SingleError_{231} = (1-t) \times \frac{1}{6}$

$DoubleError_{312} = (1-t) \times \frac{1}{6} + t \times (1-h) \times s \times p$

$DoubleError_{321} = (1-t) \times \frac{1}{6}$

The model equations can be specified in R using the MPT package. We'll generate two trees, one for each condition anticipated in our experiment. This duplication of our model equations results in double number of parameters and categories. To simplify the model and reduce the number of parameters, we assume that for the two conditions, there are no differences in the models except for the h parameter. Therefore, the t, p, and s parameters are specified to be equal for these models.

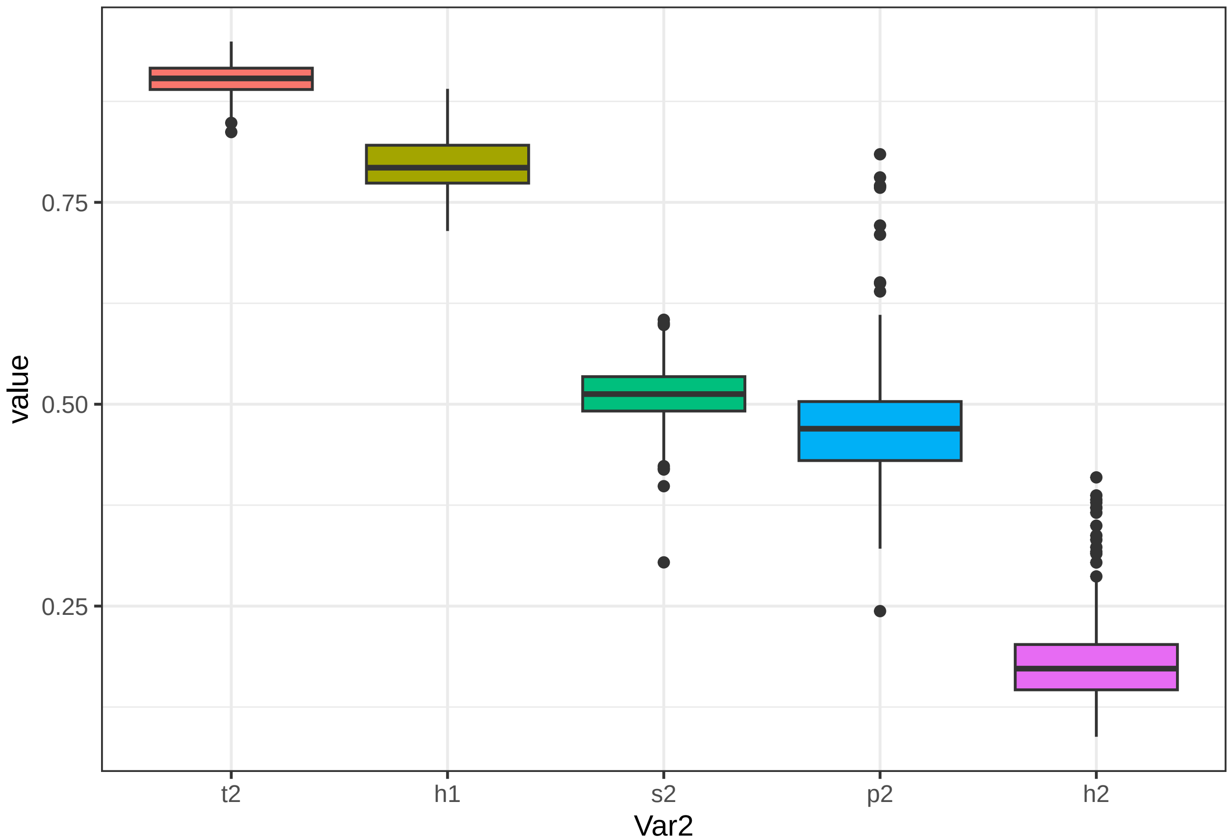
```r
library(mpt)

s <- mptspec(
  E.123cor = (1-t) * 1/6 + t * (1-h) * (1-s),
  E.132single = (1-t) * 1/6 + t * h + t * (1-h) * s * (1-p),
  E.231single = (1-t) * 1/6,
  E.321double = (1-t) * 1/6,
  E.312double = (1-t) * 1/6 + t * (1-h) * s* p,
  E.321cor = (1-t) * 1/6,
  .replicates= 2,
  .restr = list(t1=t2, s1=s2, p1=p2)
  )
```

To ensure that the model's parameters are identifiable, we conducted parameter recovery analysis. This involved generating a synthetic dataset from the model using a set of hypothesized parameters, and then fitting the model to the synthetic data. We assigned arbitrary values to parameters for both the recovery and the power analysis: t was set to 0.9 (indicating that few people would guess a response), s and p were both set to 0.5, and h for the first condition was set to 0.5 plus a $d/2$ quantity, where d equals 0.6. This corresponds to a greater probability of accessing heuristic thinking in the first condition. Similarly, h for the second condition was set to 0.5 minus a $d/2$ quantity, reflecting the reduced probability of accessing heuristic thinking in that condition. The generated parameters closely matched the hypothesized parameters, indicating satisfactory parameter recovery.

```
library("ggplot2")

dataGen <- function(nn, d) {
  structure(list(
    treeid = s$treeid,
    n = setNames((nn/2 * c(2, 1, 2, 1)/3)[s$treeid], s$treeid),
    pcat = s$par2prob(c(t2=.9, h1=0.8,s2=0.5,p2=0.5,h2=0.2))
  ), class = "mpt") |>
    simulate()
}

replicate(200, dataGen(960, 0.1) |> mpt(s, data = _) |> coef()) |>
  t() -> x
require(reshape2)
```

```
## Loading required package: reshape2
```

```
ggplot(data = melt(x), aes(x=Var2, y=value)) +
  geom_boxplot(aes(fill=Var2))+theme_bw()+theme(legend.position = "None")
```

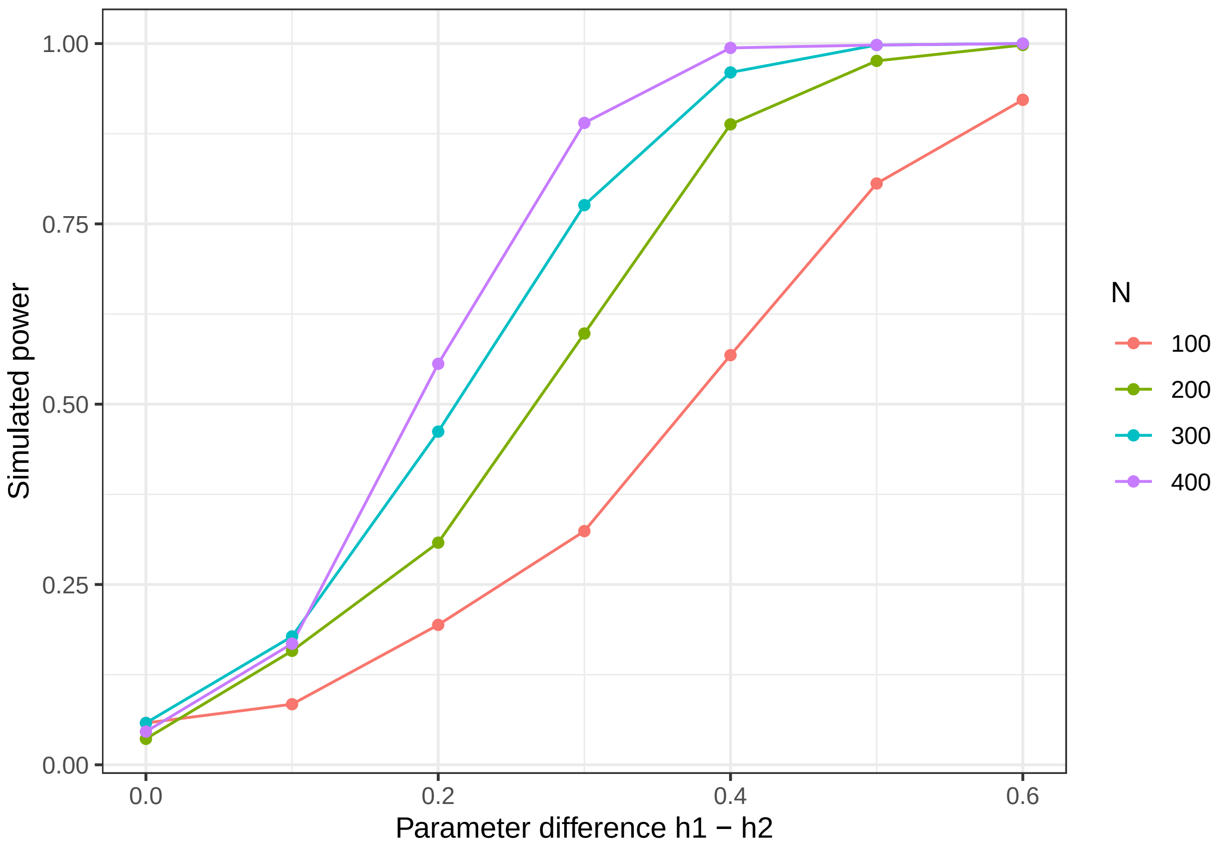


5. Power Analysis

Power is viewed as the complement of *beta*, the false negative rate. The power of the test is the chance to reject the null hypothesis, given the null hypothesis is false. Given a hypothesized parameter difference of 0.6 between h_1 and h_2, we aim to recruit 100 participants for each condition to achieve a statistical power above 0.8.

```
dataGen <- function(nn, d) {
  structure(list(
    treeid = s$treeid,
    n = setNames((nn/2 * c(2, 1, 2, 1)/3)[s$treeid], s$treeid),
    pcat = s$par2prob(c(t2=.9, h1=0.5+d/2,s2=0.3,p2=0.5,h2=0.5-d/2))
  ), class = "mpt") |>
    simulate()
}

testFun <- function(nn, d) {
  y <- dataGen(nn, d)
  m1 <- mpt(s, y)
  m2 <- mpt(update(s, .restr = list(h1 = h2)), y)
  anova(m2, m1)$"Pr(>Chi)"[2]  # return p value
}

conddf <- expand.grid(d = seq(0, 0.6, 0.1), n = 100 * (0:4))

#THIS FUNCTION TAKES 30 MINUTES TO COMPLETE
#system.time(
```

```r
#  conddf$pval <-
#    mapply(function(nn, d) replicate(500, testFun(nn, d), simplify = FALSE),
#           nn = conddf$n, d = conddf$d, SIMPLIFY = FALSE)
#)
#conddf$pwr <- sapply(conddf$pval, function(p) mean(p < .05))
#conddf <- conddf[-c(1:7),]
#save.image(file="conddf")


#INSTEAD LOAD THE ALREADY PROVIDED DATABASE
load(file="conddf")

ggplot(data=conddf, aes(x=d, y=pwr, col=as.factor(n)))+
  geom_line()+geom_point()+theme_bw()+
  labs(x="Parameter difference h1 - h2", y="Simulated power",
       color="N")
```



6. References

R Core Team. (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. https://www.R-project.org

Tversky, A. et Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, *90*(4), 293.