Design Analysis on the Causality Heuristic Propagation

Stefano Dalla Bona

2024-03-25

1. Aim

The primary goal of this study is to examine whether the illusion of causality can extend its influence beyond the merely perceived evaluations of the strength of a causal link between events. A future series of experiments could assess whether the illusion of causality influences decision-making across the confrontation of different scenarios and among different independent judges. This study serves as an initial exploration toward achieving this broader objective.

Firstly, our aim is to evaluate causality assessment within contexts featuring multiple sets of scenarios. Each scenario will inform participants about the causal link between a cause and an effect.

2. General Experimental Procedure

A straightforward approach to begin exploring the extended impact of the bias is to examine whether the illusion propagates across decisions based on two sets of scenarios. For example, a doctor may be tasked with evaluating the effectiveness of a medicine to treat a disease at one hospital and then asked to perform the same assessment for another medicine at a different hospital. Based on these assessments, the doctor can form opinions regarding the effectiveness of the two medicines and decide whether to adopt one, the other, both, or neither for a third hospital.

From a learning perspective, participants assuming the role of medical personnel will be exposed to two sets of scenarios:

```
effect <- c("a","c"); no_effect <- c("b","d");
table_1 <- c("cause 1", "no_cause 1"); table_2 <- c("cause 2", "no_cause 2");

data <- data.frame(table_1,effect,no_effect); data2 <- data.frame(table_2,effect,no_effect);

knitr::kable(data); knitr::kable(data2)
```

table_1	effect	no_effect
cause 1	a	b
no_cause 1	c	d

table_2	effect	no_effect
cause 2	a	b
no_cause 2	c	d

Where a represent the frequency of cause-effect coincidences, where the medicine is present and the patient heals; b denotes the frequency of cause-no effect coincidences, indicating instances where the medicine is present but the patient does not heal; c stands for the frequency of no cause-effect coincidences, where the medicine is not present but the patient heals; finally, d signifies the frequency of no cause-no effect coincidences, where neither the medicine is present nor the patient heals. Participants are presented with

a specific and randomly ordered configuration of *a,b,c* and *d* frequencies for each scenario. Based on these frequencies, they form an understanding of the effectiveness of the medicine.

We have chosen to maintain separate scenarios to preserve the reference to the ΔP index, which is a mathematical rule commonly utilized by individuals to assess causality and can be applied to a 2x2 matrix.

$\Delta P = \frac{a}{a+b} - \frac{c}{c+d} = P(Cause|Effect) - P(Cause|\neg Effect)$

If we were to extend the ΔP rule to a higher-order matrix, we would first need to conduct a study to assess the appropriateness of adapting the ΔP index rule. Additionally, even though we have two distinct scenarios, cells with the absence of a cause in one scenario (i.e., cells c and d in table 1) can provide evidence of the behavior of the disease without the presence of a medicine in the other hospital (i.e., cells c and d in table 2). This could potentially lead to a reformulation of the ΔP index, as the evidence of one table can be transposed to the other. However, this aspect could be of interest to investigate further. It would be valuable and informative to observe whether people utilize previous evidence to consider new scenarios.

3. Conditions

Key Concepts:

- When a relationship between the cause and the effect is present, the ΔP index is positive. People typically can recognize the presence of a causal link in this scenario (denoted later as E, where E = Effective).

- When a relationship between the cause and the effect is absent, the ΔP index equals 0. People generally can acknowledge the absence of a causal link in this scenario (denoted later as N, where N = No causal link).

- If we manipulate the frequencies of scenarios such that the ΔP index remains 0 but the cause frequency or the effect frequency is increased, individuals may be inclined to overestimate a causal link between the events, resulting in an illusion of causality (denoted later as I, where I = Illusion).

- The strength of the perceived causal link between events can be modulated. For example, a stronger true statistical relationship between cause and effect may lead to a greater evaluation of the causal link between events. Conditions with greater outcome-density or cause-density can also contribute to a greater illusion of causality, and so forth. Although we will utilize categorical conditions, their characteristics will depend on the assigned frequency values that we intend to allocate to each condition.

We have formulated four comparative conditions:

- Participants will compare two scenarios in the following sequence: one scenario will feature an illusory medicine I, while the other will involve a genuinely effective medicine E. The sequence of scenarios will be denoted as S(I,E).

- Participants will compare two scenarios in the following sequence: one scenario will feature an effective medicine E, while the other will involve an illusory medicine I. The sequence of scenarios will be denoted as S(E,I).

- Participants will compare two scenarios in the following sequence: one scenario will feature an illusory medicine I, while the other will involve an ineffective medicine N. The sequence of scenarios will be denoted as S(I,N).

- Participants will compare two scenarios in the following sequence: one scenario will feature an ineffective medicine N, while the other will involve an illusory medicine I. The sequence of scenarios will be denoted as S(N,I).

```
S_IE <- c("I", "E"); S_EI<- rev(S_IE); S_IN <- c("I", "N"); S_NI<- rev(S_IN)

Sequence <- c("First", "Second")
```

```r
data <- data.frame(Sequence,S_IE,S_EI,S_IN,S_NI)

knitr::kable(data)
```

Sequence	S_IE	S_EI	S_IN	S_NI
First	I	E	I	N
Second	E	I	N	I

4. Measures

4.1 Assessing causality connection

For each condition, we have three dependent variables: the perceived causal strength of the first medicine S(A,-) after the first scenario presentation, the perceived causal strength of the first medicine after the second scenario presentation S(A,-), and the perceived causal strength of the second medicine after the second scenario presentation S(-,B). Participants typically assess the strength of the causal link between events using a 101-point Likert scale ranging from 0 to 100. However, this scale may not be sufficiently clear to participants, as it is prone to misleading interpretation. For example, participants may find it unclear whether a value of 50 represents indecisiveness or moderate evidence of the causal/effect link. As shown by Ng *et al.* (2024), results can vary depending on the final assessment question; for instance, a bipolar scale reduces the magnitude of the illusion. However, the bipolar scale proposed by the authors may also be insufficient for our purpose, as its extremes encompass the inhibition of the cause on the effect. Therefore, we will maintain a 101-point scale ranging from 0 to 100, with the midpoint designated as the indecisiveness point between the continuum of effectiveness and ineffectiveness.

4.2 Making a decision

For each condition, after seeing both scenarios, each participant will provide a quantitative decisional vote for each medicine on a 6-point Likert Scale, ranging from -3 to 3, where:

- -3: "I would strongly suggest to the hospital not to adopt medicine X"
- +3: "I would strongly suggest to the hospital to adopt medicine X"

The second measure is introduced to assess whether participants would choose one medicine, the other, both, or none, and to gauge the relative strength of each decision. Unlike the assessing scale, this measure is constructed using an even set of numbers to prompt participants to make a decision.

5. Hypothesis

5.1 Assessing causality connection

Firstly, we hypothesize that the second evaluation of the illusory medicine in condition S(I,E) will be lower than the first evaluation of the illusory medicine in the same condition S(I,E). We believe that comparing the illusory medicine to the effective one will diminish the perceived effectiveness of the illusory medicine, leading participants to update their previous belief about its effectiveness.

- H0: First evaluation of I in S(I,E) = Second evaluation of I in S(I,E)
- H1: First evaluation of I in S(I,E) < Second evaluation of I in S(I,E)

Thus, a paired sample t-test is required to try to support that data are unlikely under the null-point hypothesis. This hypothesis serves as the foundation for our Design Analysis in the next section. It must be noticed that a null result would be very interesting, supporting the idea of the existence of an assessment inertia in the illusion frame. Additionally we are planning to compute Bayes Factor for assessing if data are more likely under the null model (BF_{01}) or the alternative model (BF_{10}).

The S(E,I) conditions will be used to assess whether there is an effect of the presentation sequence, by comparing the second evaluations of the two medicines for conditions S(I,E) and S(E,I). We also want to test if comparing the effective medicine S(E,-) to the illusory one S(E,I) will prompt the perceived effectiveness of the effective medicine, and if there are comparable second evaluations for both conditions S(I,E) and S(E,I).

Conditions S(I,N) and S(N,I) are symmetrical to the previous conditions, postulating an inverse effect: we believe that a no-contingency scenario S(-,N) is able to increase the magnitude of the causality heuristic in condition S(I,N), while the causality heuristic is likely to prompt a reduction in the evaluation of the ineffective medicine in condition S(N,I).

5.2 Decisional pattern

Additionally, we aim to test whether in conditions S(I,E) and S(E,I), the extent of the reduction in perceived evidence for the illusory medicine prompts a negative vote for the illusory medicine and a positive vote for the effective medicine, or if, even though the effective medicine is considerate more plausible, still the illusory medicine will be choose, also with a lower vote compared to the effective medicine.

We also aim to test whether in conditions S(I,N) and S(N,I), the extent of the increase in perceived evidence for the illusory medicine prompts a positive vote for the illusory medicine and a negative vote for the ineffective medicine, or if, the illusory medicine will be not selected.

The decisional matrix for the illusory medicine coming from S(I,E), S(E,I), S(I,N), S(N,I) will show if there is a difference in the decisional pattern for the illusory medicine dependent on the type of comparison.

- H0: Proportion of decided I with E = Proportion of decided I with N
- H1: Proportion of decided I with E < Proportion of decided I with N

6. Power Analysis (using our previous experiment)

We consider data from our prior experiment (Dalla Bona et Vicovaro, 2024). The dataset from this previous experiment is available for download on Open Science Framework repository. In this prior study, we presented an illusory medicine (outcome-density type) with a ΔP value of 0 and an effective medicine with a ΔP value of 0.60. We show the overall causal assessment for each medicine.

```r
#setwd()
data1 <- read.csv("Dataframe1_CA.csv", header = T, sep=";")
data2 <- read.csv("Dataframe2_CA.csv", header = T, sep=";")

data1 <- data1[1:209,]
data2 <- data2[1:102,]

illusory <- c( data1$vd[data1$valido=="si" & data1$gruppo<3],
               data2$vd[data2$valido=="si" & data2$gruppo<2] )

effective <- c( data1$vd[data1$valido=="si" & data1$gruppo>2],
                data2$vd[data2$valido=="si" & data2$gruppo>1] )

library(ggplot2); library(gridExtra)

a <- ggplot(mapping = aes(x=illusory))+
  geom_density(fill="darkred",alpha=.5,col=NA)+theme_bw()+
  geom_vline(xintercept = mean(illusory), lty="dashed")+xlim(c(0,100))+
  geom_boxplot(width=.001,position = "dodge2", fill="darkred")

b<- ggplot(mapping = aes(x=effective))+
  geom_density(fill="darkblue",alpha=.5,col=NA)+theme_bw()+
  geom_vline(xintercept = mean(illusory), lty="dashed")+xlim(c(0,100))+
```
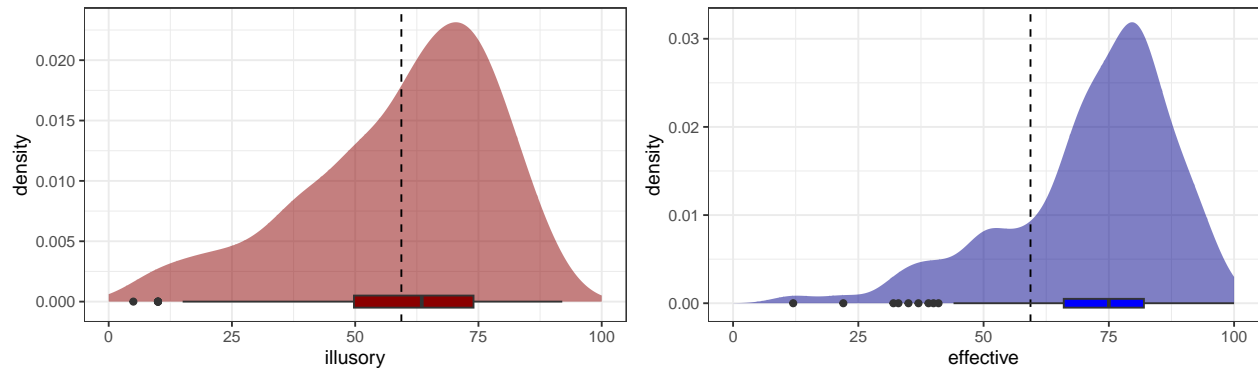
```r
  geom_boxplot(width=.001,position = "dodge2", fill="blue")

grid.arrange(a,b,ncol=2)
```


As evident from the density plots, different medicines are not uniformly evaluated by participants: the effective medicine (M = 59.36, SD = 19.41) is perceived as more effective than the illusory medicine (M = 71.65, SD = 16.58), although there is a 71% overlap in distributions. It's noteworthy that the illusory medicine yields a median evaluation greater than 50, suggesting that participants not only overestimate the causal connection between variables but also tend to provide a positive assessment for the ineffective medicine. We aim to employ the ΔP values established in the previous experiment to investigate whether learning from the effective medicine leads to a reduction in the assessment of the illusory medicine, or if there is an inertia in the assessment of the illusory medicine.

```r
summary(effective); var(effective) ; sd(effective)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.00   66.00   75.00   71.65   82.00  100.00
```

```
## [1] 275.0025
```

```
## [1] 16.5832
```

```r
summary(illusory); var(illusory) ; sd(illusory)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.00   49.75   63.50   59.36   74.00   92.00
```

```
## [1] 376.7465
```

```
## [1] 19.40996
```

```r
overlapping::overlap(list(effective,illusory))
```

```
## $OV
## [1] 0.7186178
```

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. We set the value of *beta* to a standard probability of 0.8 to calculate the necessary sample size for each group. Using the same ΔP value from the previous experiment enables us to make a precise statement regarding the type of reduction we hypothesize. We posit that the comparison with the effective medicine is capable of reducing the causal assessment for the illusory medicine to a point where participants are, on average, moderately indecisive about its efficacy (i.e., 50 points on a 101-point Likert scale). We can calculate Cohen's d for paired sample using the same mean and standard deviation of the illusory medicine from our previous experiment, along with the new hypothesized mean, and assuming homoscedasticity.

```r
delta_M <- mean(illusory) - 50

delta_M/sd(illusory)
```

```
## [1] 0.4824395
```

```r
library(pwr)
pwr.t.test(d= delta_M / sd(illusory), power = 0.8,
           alternative = "greater", type="paired" )
```

```
##
##       Paired t test power calculation
##
##              n = 27.96697
##              d = 0.4824395
##       sig.level = 0.05
##           power = 0.8
##     alternative = greater
##
## NOTE: n is number of *pairs*
```

Thus we aim to collect 40 partecipants for each of the four groups, i.e., S(I,E), S(E,I), S(I,N), S(N,I)., to increase our power above 0.90.

```r
pwr.t.test(d= delta_M / sd(illusory), n=40,
           alternative = "greater", type="paired" )
```

```
##
##       Paired t test power calculation
##
##              n = 40
##              d = 0.4824395
##       sig.level = 0.05
##           power = 0.9119338
##     alternative = greater
##
## NOTE: n is number of *pairs*
```

We can visualize the power curve as a function of the sample size.

```r
x <- c(1:45)
y <- pwr.t.test(d= delta_M / sd(illusory), n = x,
           alternative = "greater", type="paired" )$power
library(ggplot2)
ggplot(mapping=aes(x,y)) + theme_bw()+
  geom_line(color="navy") + xlab("Numerosity") + ylab(expression(1-beta))+
  geom_hline(yintercept = .91, color="darkred", alpha=.3, linetype="dashed")+
  geom_vline(xintercept = 40, color="darkred", alpha=.3,linetype="dashed")+
  geom_point(aes(x=40,y=.91),colour="orange", shape=20, size=3)
```
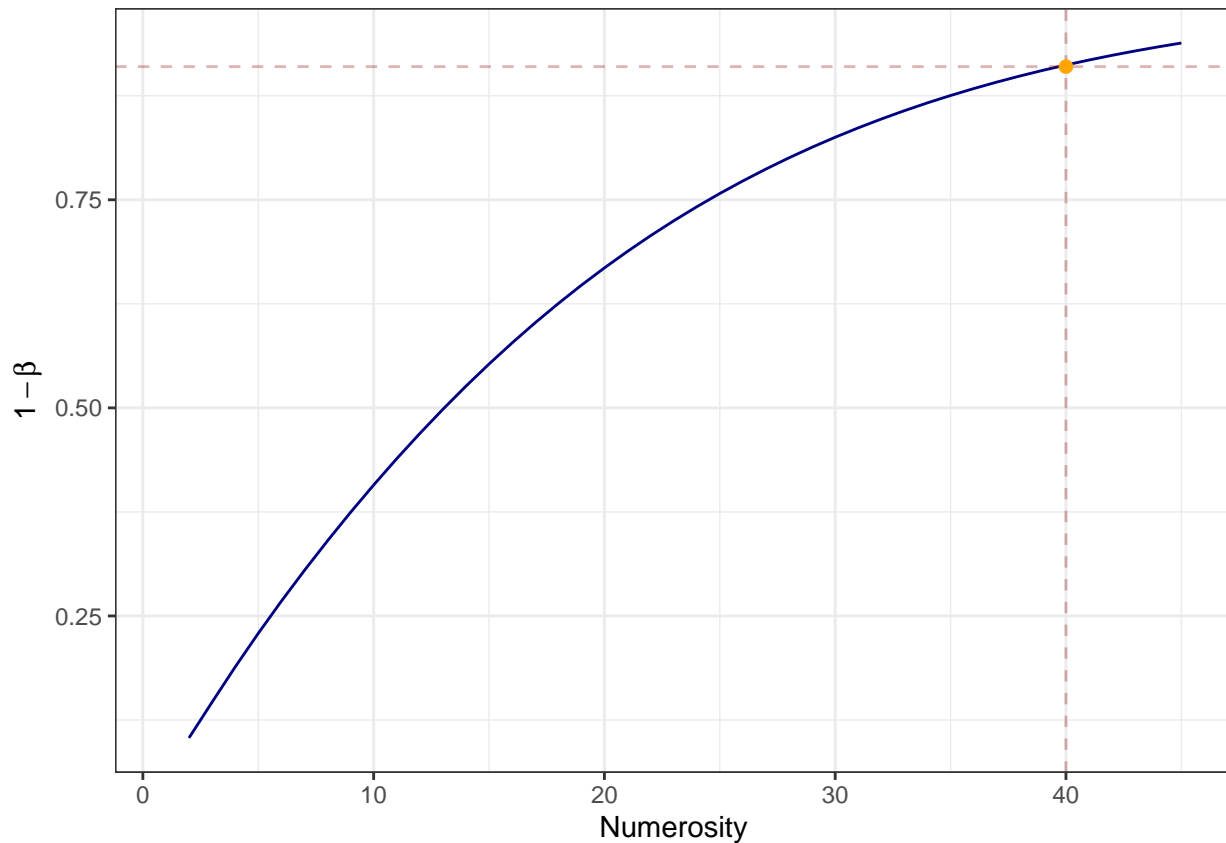
7. Inferential risks

Type S error is the probability of obtaining a statistically significant result in the opposite direction to the plausible one (Callegher *et al.*, 2021). Since we have formulated an unidirectional hypothesis, the Type S error should be equal to 0.

Type M error is the factor by which a statistically significant effect is on average exaggerated. The Type M error is calculated to be 1.06. This indicates that, on average, the effect size of statistically significant results from the simulation exaggerated the inputted effect size by approximately 6%.

```r
set.seed(20240401)
library(PRDA)
retrospective(effect_size = delta_M/sd(illusory), sample_n1 = 40,
              test_method = "paired", sample_n2 = 40,
              alternative = "greater", sig_level = 0.05)
```

```
##
##  Design Analysis
##
## Hypothesized effect:  cohen_d = 0.4824395
##
## Study characteristics:
##    test_method   sample_n1   sample_n2   alternative   sig_level   df
##    paired        40          40          greater       0.05        39
##
## Inferential risks:
##    power   typeM   typeS
##    0.912   1.059   0
```

```
##
## Critical value(s): cohen_d  =  0.266
```

8. References

Callegher, C., Bertoldo, G., Toffalini, E., Vesely, A., Andreella, A., Pastore, M., Altoè, G., et al. (2021). PRDA: An R package for prospective and retrospective design analysis. *Journal of Open Source Software*, *6*(58).

Dalla Bona, S. et Vicovaro, M. (2024). Does perceptual disfluency affect the illusion of causality? *The Quarterly Journal of Experimental Psychology*.

Ng, D. W., Lee, J. C. et Lovibond, P. F. (2024). Unidirectional rating scales overestimate the illusory causation phenomenon. *Quarterly journal of experimental psychology*, *77*(3), 551-562.

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