CREP Turri et al. Report Form

**Descriptive Statistics of the Sample**

Split the data by para\_condition and scale\_condition. Then calculate the percent of Female participants (gender), the race/ethnicity of participants (race\_ethnic), and the average age of participants with the standard deviation. Fill in the table below.

**Note:** Percentages reflect percent across the entire study *N* = 73.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Gettier Case | Knowledge Case | Ignorance Case |
| Binary Original Scale | Sample Size | 19 | 12 | 9 |
| Female | % 8.22 | 5.48 | 6.85 |
| White/European | % 20.55 | 13.70 | 9.59 |
| Latin/Hispanic | % 1.37 | 0.00 | 1.37 |
| Black/African | % 2.74 | 0.00 | 0.00 |
| Asian | % 0.00 | 2.74 | 1.37 |
| Hawaiian/Pacific Islander | % 0.00 | 0.00 | 0.00 |
| Other | % 1.37 | 0.00 | 0.00 |
| Age | M 19.26  (SD) 1.56 | 20.44  4.72 | 20.17  3.64 |
| Visual Analog Scale | Sample Size | 6 | 12 | 15 |
| Female | % 4.11 | 9.59 | 12.33 |
| White/European | % 6.85 | 12.33 | 13.70 |
| Latin/Hispanic | % 0.00 | 0.00 | 5.48 |
| Black/African | % 0.00 | 1.37 | 0.00 |
| Asian | % 0.00 | 1.37 | 0.00 |
| Hawaiian/Pacific Islander | % 0.00 | 1.37 | 0.00 |
| Other | % 1.37 | 0.00 | 1.37 |
| Age | M 18.83  (SD) 0.41 | 19.40  1.12 | 19.67  1.61 |

**Percent Correct of Attention Question**

In the Gettier and Knowledge case, you should find that the answer to squirrel\_prairie is Ground Squirrel and the answer to the Ignorance case is Prairie Dog, regardless of which scale they are using.

Calculate the percent correct for each condition:

Gettier Case: 80.0%

Knowledge Case: 95.83%

Ignorance Case: 79.17%

**Hypothesis Tests:**

**Direct replication:**

Here’s the table from the original study:



Recreate the table on our study by calculating the percent of each answer using para\_condition and knows\_believes:

|  |  |  |  |
| --- | --- | --- | --- |
|  | No threat  Knowledge Case | Threat  Gettier Case | No Detection  Ignorance Case |
| Knows | % 66.67 | 57.89 | 27.28 |
| Reasonable | % 100.0 | 94.74 | 90.91 |

Is that difference statistically significant? Calculated a chi-square test on just knows\_believes:

χ2(*df* = 2, *N* = 39) = 3.72, *p* = .155, Cramer’s *V* = .309

If *p* < .05, then use a chi-square test on each pairwise combination:

* No threat versus threat
* No threat versus no detection
* Threat versus no detection

Here’s their results:

**Overall:** As predicted, assignments to conditions affected the rates of knowledge attribution, χ2(df = 2, N = 135) = 39.63, p <.001, Cramer’s V = .542 (all tests are two-tailed unless otherwise noted).

**Follow up:** Pairwise comparisons detected no difference in knowledge attributions between the No Threat and Threat conditions, Fisher’s p = .164, n.s., and a large difference between Threat and No Detection, Fisher’s p < .001, Cramer’s V = .509.

Are our results the same?

Not statistically significant, but the Cramer’s V is still large, but the smaller sample size could account for the non-significant p-values.

Is that difference statistically significant (part 2)? Calculated a chi-square test on just reason\_unreason:

χ2(*df* = 2, *N* = 39) = 0.84, *p* = .656, Cramer’s *V* = .147

If *p* < .05, then use a chi-square test on each pairwise combination:

* No threat versus threat
* No threat versus no detection
* Threat versus no detection

Here’s their results:

**Overall:** Condition did not affect whether people said that it was reasonable for the protagonist to think he was looking at an object of the relevant sort, χ2(df= 2, N = 135) = 4.49, p = .106, n.s., so the differences in knowledge attribution can not be due to perceived differences in what it was reasonable for the protagonist to believe.

Are our results the same?

Not statistically significant, but are similar.

**Extension:**

As an extension, we added the visual analog scale to see if the categorical results would extend to a continuous measure.

For knows\_believes, calculate a one-way ANOVA to determine if there are overall differences in ratings:

Df Sum Sq Mean Sq F value Pr(>F)

para\_condition 2 66 33.2 0.037 0.964

Residuals 30 26964 898.8

*F*(2,30) = 0.04, *p* = .964, *R2 =* .002

Means:

Gettier Case Ignorance Case Knowledge Case

60.0 64.0 63.2

SDs:

Gettier Case Ignorance Case Knowledge Case

23.09112 33.78542 28.96106

If *p* < .05, then use a t-test on each pairwise combination (you can use a Bonferroni correction option in your program if you want):

* No threat versus threat
* No threat versus no detection
* Threat versus no detection

Do these results match the results we found above?

Not significant, so they do match

For reason\_unreason, calculate a one-way ANOVA to determine if there are overall differences in ratings:

Df Sum Sq Mean Sq F value Pr(>F)

para\_condition 2 1433 716.6 1.532 0.232

Residuals 30 14028 467.6

*F*(2,30) = 1.532, *p* = .232, *R2 =* .09

Means:

Gettier Case Ignorance Case Knowledge Case

79.33333 65.16667 78.66667

SDs:

Gettier Case Ignorance Case Knowledge Case

21.51898 20.54633 22.47115

If *p* < .05, then use a t-test on each pairwise combination (you can use a Bonferroni correction option in your program if you want):

* No threat versus threat
* No threat versus no detection
* Threat versus no detection

Do these results match the results we found above?

No significant difference in ratings – but the effect size here is larger.