Analysis script for AdvisorChoice (Agreement advisors)

Matt Jaquiery ([matt.jaquiery@psy.ox.ac.uk](mailto:matt.jaquiery@psy.ox.ac.uk))

Table of Contents

September 2018  
[Script run 2018-11-15 17:01:26]

# 0) Load utility files

## Loading required package: jsonlite

## Loading required package: BayesFactor

## Loading required package: coda

## Loading required package: Matrix

## \*\*\*\*\*\*\*\*\*\*\*\*  
## Welcome to BayesFactor 0.9.12-4.2. If you have questions, please contact Richard Morey (richarddmorey@gmail.com).  
##   
## Type BFManual() to open the manual.  
## \*\*\*\*\*\*\*\*\*\*\*\*

## Loading required package: tidyverse

## -- Attaching packages --------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts ------------------------------------------------------------------------------------ tidyverse\_conflicts() --  
## x tidyr::expand() masks Matrix::expand()  
## x dplyr::filter() masks stats::filter()  
## x purrr::flatten() masks jsonlite::flatten()  
## x dplyr::lag() masks stats::lag()

## Loading required package: reshape2

##   
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':  
##   
## smiths

## Loading required package: lme4

## Loading required package: lsr

## Loading required package: ez

## Loading required package: knitr

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

## The following objects are masked from 'package:base':  
##   
## format.pval, units

## Loading required package: scoring

## Loading required package: prettyMD

# 1) Load data

# 2) Demographics

Demographic data are not collected and therefore not analysed

Responses were collected between 2018-09-19 13:28:08 and 2018-09-20 14:40:54.

# 3) Manipulation checks

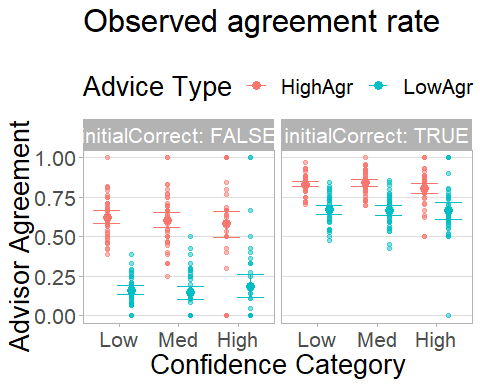
## 3.i) Overall agreement by contingency

These advisors don’t show contingent agreement, so there shouldn’t be much here. This is retained mostly for comparison purposes

## Warning: You have removed one or more Ss from the analysis. Refactoring  
## "pid" for ANOVA.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effect | DFn | DFd | F | p | p<.05 | ges |
| 2 | confidenceCategory | 2 | 82 | 0.31 | .731 |  | .0010 |
| 3 | adviceType | 1 | 41 | 529.87 | .000 | \* | .4908 |
| 4 | initialCorrect | 1 | 41 | 824.99 | .000 | \* | .5873 |
| 5 | confidenceCategory:adviceType | 2 | 82 | 0.85 | .431 |  | .0044 |
| 6 | confidenceCategory:initialCorrect | 2 | 82 | 0.39 | .679 |  | .0014 |
| 7 | adviceType:initialCorrect | 1 | 41 | 71.47 | .000 | \* | .1741 |
| 8 | confidenceCategory:adviceType:initialCorrect | 2 | 82 | 0.20 | .816 |  | 9e-04 |

## 3.ii) Graph: overall agreement by contingency

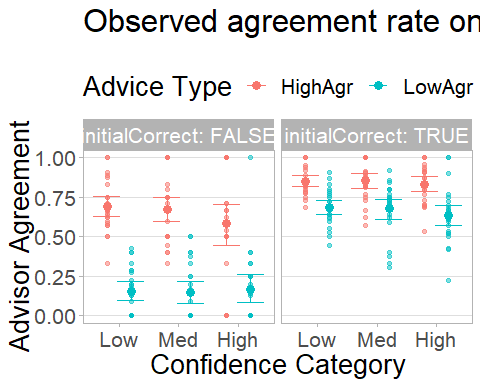


## 3.iii) Initial block agreement by contingency

## Warning: You have removed one or more Ss from the analysis. Refactoring  
## "pid" for ANOVA.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effect | DFn | DFd | F | p | p<.05 | ges |
| 2 | confidenceCategory | 2 | 46 | 1.17 | .318 |  | .0102 |
| 3 | adviceType | 1 | 23 | 675.93 | .000 | \* | .4748 |
| 4 | initialCorrect | 1 | 23 | 314.86 | .000 | \* | .4998 |
| 5 | confidenceCategory:adviceType | 2 | 46 | 0.45 | .638 |  | .0038 |
| 6 | confidenceCategory:initialCorrect | 2 | 46 | 0.05 | .955 |  | 3e-04 |
| 7 | adviceType:initialCorrect | 1 | 23 | 93.52 | .000 | \* | .1639 |
| 8 | confidenceCategory:adviceType:initialCorrect | 2 | 46 | 0.97 | .386 |  | .0093 |

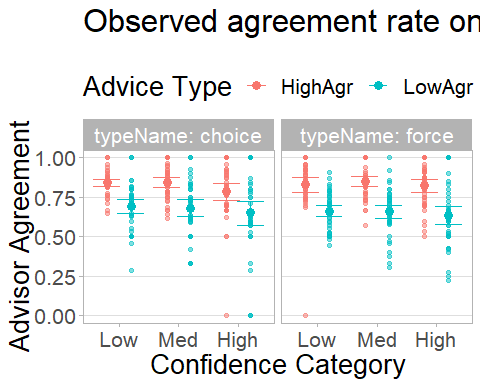
## 3.iv) Graph: initial block agreement by contingency



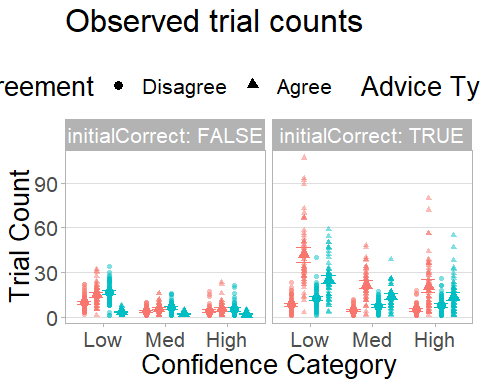
### 3.iv.i) Graph of agreement by block on initialCorrect trials

## Warning: You have removed one or more Ss from the analysis. Refactoring  
## "pid" for ANOVA.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effect | DFn | DFd | F | p | p<.05 | ges |
| 2 | confidenceCategory | 2 | 84 | 2.87 | .062 |  | .0105 |
| 3 | adviceType | 1 | 42 | 109.09 | .000 | \* | .2209 |
| 4 | typeName | 1 | 42 | 0.16 | .691 |  | 3e-04 |
| 5 | confidenceCategory:adviceType | 2 | 84 | 0.15 | .858 |  | 5e-04 |
| 6 | confidenceCategory:typeName | 2 | 84 | 0.33 | .719 |  | .0015 |
| 7 | adviceType:typeName | 1 | 42 | 1.59 | .214 |  | .0027 |
| 8 | confidenceCategory:adviceType:typeName | 2 | 84 | 0.10 | .905 |  | 4e-04 |



## 3.v) Trial count by contingency



# 4) Exclusions

Exclusion rules:

* Proportion of correct initial judgements must be (.60 < cor1/n < .90)
  + NB:practice trials are INCLUDED in this since they are used in part for determining confidence calibration
* Having fewer than 3 confidence categories
* Having fewer than 5% of trials in each confidence category
* There being more data collected than specified in pre-registration

|  |  |
| --- | --- |
| exclusionReason | count |
| Confidence | 7 |
| Confidence.cat | 11 |
| FALSE | 50 |

# 5) Descriptives

## 5.i) Proportion correct

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| decision | adviceType | target | meanCorrect | cl95Min | cl95Max | rangeMin | rangeMax |
| initial | HighAgr | 0.71 | 0.72 | 0.71 | 0.73 | 0.64 | 0.79 |
| initial | LowAgr | 0.71 | 0.71 | 0.70 | 0.72 | 0.64 | 0.80 |
| initial | Total | 0.71 | 0.71 | 0.71 | 0.72 | 0.68 | 0.75 |
| final | HighAgr | NA | 0.73 | 0.72 | 0.74 | 0.64 | 0.80 |
| final | LowAgr | NA | 0.73 | 0.71 | 0.74 | 0.62 | 0.85 |
| final | Total | NA | 0.73 | 0.72 | 0.73 | 0.68 | 0.79 |

## 5.ii) Agreement rate

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| adviceType | name | probAgree | cl95Min | cl95Max | rangeMin | rangeMax |
| HighAgr | low | 0.84 | 0.82 | 0.86 | 0.71 | 0.97 |
| HighAgr | medium | 0.84 | 0.82 | 0.86 | 0.70 | 1.00 |
| HighAgr | high | 0.80 | 0.77 | 0.83 | 0.50 | 1.00 |
| HighAgr | allCorrect | 0.83 | 0.82 | 0.84 | 0.75 | 0.90 |
| HighAgr | allWrong | 0.62 | 0.60 | 0.64 | 0.45 | 0.85 |
| HighAgr | All | 0.77 | 0.76 | 0.78 | 0.66 | 0.85 |
| LowAgr | low | 0.67 | 0.64 | 0.69 | 0.47 | 0.85 |
| LowAgr | medium | 0.67 | 0.64 | 0.69 | 0.43 | 0.86 |
| LowAgr | high | 0.65 | 0.60 | 0.70 | 0.00 | 1.00 |
| LowAgr | allCorrect | 0.66 | 0.64 | 0.67 | 0.55 | 0.77 |
| LowAgr | allWrong | 0.16 | 0.15 | 0.18 | 0.05 | 0.30 |
| LowAgr | All | 0.52 | 0.50 | 0.53 | 0.42 | 0.63 |

## 5.iii) Mean confidence

### 5.iii.i Mean confidence by correctness

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| decision | correct | meanConfidence | cl95Min | cl95Max | rangeMin | rangeMax |
| initial | TRUE | 27.87 | 25.30 | 30.44 | 6.72 | 44.39 |
| initial | FALSE | 23.46 | 20.72 | 26.19 | 5.07 | 44.10 |
| initial | Both | 26.61 | 24.02 | 29.21 | 6.23 | 44.30 |
| final | TRUE | 30.88 | 28.23 | 33.54 | 10.61 | 48.37 |
| final | FALSE | 24.70 | 21.90 | 27.51 | 8.41 | 44.91 |
| final | Both | 29.20 | 26.55 | 31.85 | 10.03 | 47.40 |

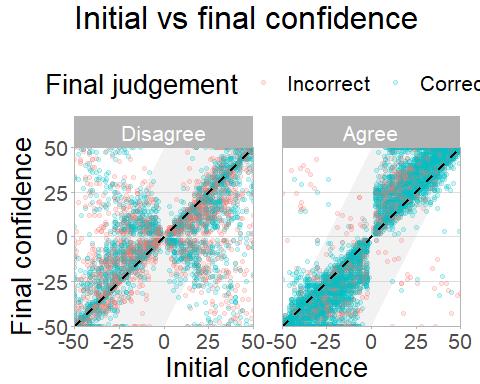
### 5.iii.ii Mean confidence by advisor

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| decision | adviceType | meanConfidence | cl95Min | cl95Max | rangeMin | rangeMax |
| initial | HighAgr | 26.62 | 24.05 | 29.19 | 6.35 | 44.41 |
| initial | LowAgr | 26.63 | 24.01 | 29.26 | 5.93 | 44.00 |
| initial | Both | 26.61 | 24.02 | 29.21 | 6.23 | 44.30 |
| final | HighAgr | 30.28 | 27.66 | 32.90 | 10.38 | 47.39 |
| final | LowAgr | 27.82 | 25.08 | 30.55 | 9.61 | 47.42 |
| final | Both | 29.20 | 26.55 | 31.85 | 10.03 | 47.40 |

### 5.iii.iii Mean confidence by dis/agreement and advice type

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| agree | advisor | meanConfidence | cl95Min | cl95Max | rangeMin | rangeMax |
| TRUE | HighAgr | 32.89 | 30.17 | 35.61 | 11.33 | 49.19 |
| TRUE | LowAgr | 32.25 | 29.47 | 35.04 | 11.91 | 49.98 |
| TRUE | Both | 32.69 | 29.97 | 35.40 | 11.54 | 49.50 |
| FALSE | HighAgr | 21.16 | 18.23 | 24.08 | 3.17 | 45.46 |
| FALSE | LowAgr | 22.92 | 19.74 | 26.10 | 3.18 | 44.57 |
| FALSE | Both | 22.25 | 19.21 | 25.29 | 3.18 | 45.01 |

## 5.iv) Graph: Initial vs Final confidence

Influence of the advisors is evident in the deviation from the dashed y=x line. Points lying below the line indicate a more leftward response from initial to final judgement. Points above the line indicate a more rightward response in the final judgement. The further away from the y=x line, the greater the change from initial to final judgement. Separate plots show agreement vs disagreement trials (between the advisor and judge), and separate colours indicate whether the judge’s final decision was correct or incorrect. The shaded area indicates the boundary for the symmetrical influence measure. Points outside this area are truncated by moving them vertically until they meet the grey area. 

### Response contingencies

|  |  |  |  |
| --- | --- | --- | --- |
| advisorAgrees | increaseConfPerC | noChangePerC | decreaseConfPerC |
| TRUE | 0.6078406 | 0.2452423 | 0.1469170 |
| FALSE | 0.1745993 | 0.1894123 | 0.6359883 |

## 5.v) Questionnaire responses

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| timepoint | question | adviceType | mean | cl95Min | cl95Max | rangeMin | rangeMax |
| 1 | likeability | HighAgr | 70.65 | 65.66 | 75.64 | 31.5 | 100 |
| 1 | ability | HighAgr | 67.83 | 62.25 | 73.41 | 36.0 | 100 |
| 1 | benevolence | HighAgr | 71.64 | 66.76 | 76.52 | 39.0 | 100 |
| 2 | likeability | HighAgr | 78.41 | 73.68 | 83.14 | 27.0 | 100 |
| 2 | ability | HighAgr | 74.22 | 69.25 | 79.19 | 6.5 | 100 |
| 2 | benevolence | HighAgr | 77.73 | 72.24 | 83.22 | 5.5 | 100 |
| 1 | likeability | LowAgr | 68.38 | 62.93 | 73.83 | 19.5 | 100 |
| 1 | ability | LowAgr | 66.09 | 60.13 | 72.05 | 11.0 | 100 |
| 1 | benevolence | LowAgr | 67.52 | 61.94 | 73.10 | 15.0 | 100 |
| 2 | likeability | LowAgr | 65.21 | 58.57 | 71.85 | 17.5 | 100 |
| 2 | ability | LowAgr | 57.02 | 50.24 | 63.80 | 6.5 | 100 |
| 2 | benevolence | LowAgr | 59.25 | 51.77 | 66.73 | 0.0 | 100 |

## 5.vi) Advisor accuracy

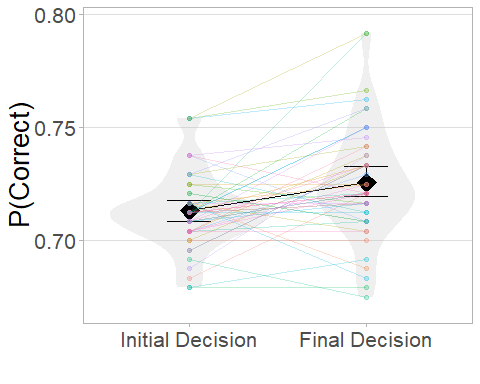
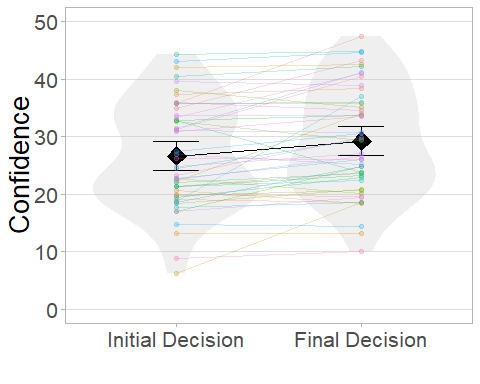
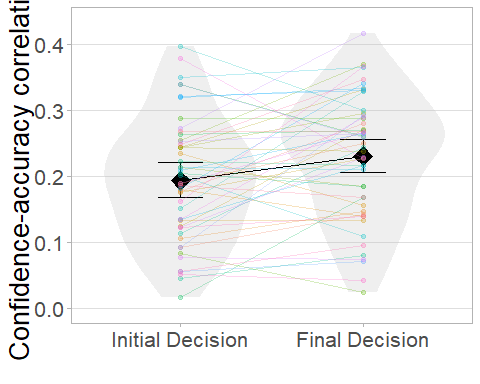
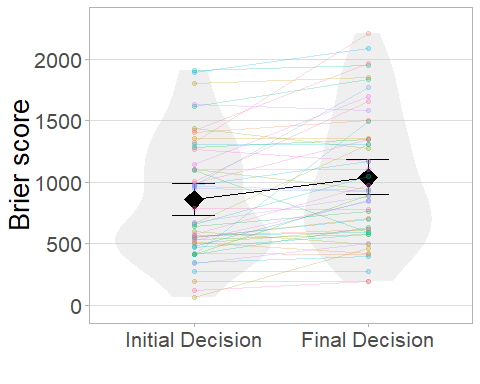
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| agree | advisor | meanAccuracy | cl95Min | cl95Max | rangeMin | rangeMax |
| TRUE | HighAgr | 0.77 | 0.76 | 0.78 | 0.69 | 0.85 |
| TRUE | LowAgr | 0.91 | 0.90 | 0.92 | 0.82 | 0.98 |
| TRUE | Both | 0.82 | 0.81 | 0.83 | 0.78 | 0.87 |
| FALSE | HighAgr | 0.47 | 0.44 | 0.50 | 0.17 | 0.61 |
| FALSE | LowAgr | 0.50 | 0.48 | 0.52 | 0.33 | 0.60 |
| FALSE | Both | 0.49 | 0.48 | 0.50 | 0.37 | 0.58 |
| Both | HighAgr | 0.70 | 0.69 | 0.71 | 0.62 | 0.80 |
| Both | LowAgr | 0.71 | 0.70 | 0.72 | 0.61 | 0.81 |
| Both | Both | 0.71 | 0.70 | 0.71 | 0.64 | 0.75 |

## 5.vii) Contingency Ns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| confidenceCategory | meanProp | cl95L | cl95H | rangeL | rangeH |
| .00 | .43 | .39 | .47 | .15 | .72 |
| 1.00 | .30 | .27 | .33 | .10 | .51 |
| 2.00 | .27 | .22 | .31 | .05 | .72 |
| NaN | .29 | .28 | .29 | .25 | .32 |

## 5.viii) Performance summary table

Data sanity checks - did participants demonstrate good use of confidence, did participants benefit from advice was their calibration better after advice, was their accuracy better after advice?

[1] “Decision accuracy:” [1] “Initial=0.71 [95.00%CI: 0.71, 0.72] [Range: 0.68, 0.75]” [1] “Final=0.73 [95.00%CI: 0.72, 0.73] [Range: 0.68, 0.79]” [1] “t(49) = -3.87 [95%CI: -0.02, -0.01], p = .000 , d = 0.59, BF = 79.171” [1] “Decision confidence:” [1] “Initial=26.61 [95.00%CI: 24.02, 29.21] [Range: 6.23, 44.30]” [1] “Final=29.20 [95.00%CI: 26.55, 31.85] [Range: 10.03, 47.40]” [1] “t(49) = -3.82 [95%CI: -3.95, -1.22], p = .000 , d = 0.28, BF = 68.054” [1] “Confidence-accuracy correlation:” [1] “Initial=0.19 [95.00%CI: 0.17, 0.22] [Range: 0.02, 0.40]” [1] “Final=0.23 [95.00%CI: 0.20, 0.26] [Range: 0.02, 0.42]” [1] “t(49) = -3.18 [95%CI: -0.06, -0.01], p = .003 , d = 0.39, BF = 12.257” [1] “Brier score:” [1] “Initial=856.96 [95.00%CI: 720.87, 993.06] [Range: 67.49, 1909.88]” [1] “Final=1036.67 [95.00%CI: 884.53, 1188.81] [Range: 190.89, 2208.93]” [1] “t(49) = -4.35 [95%CI: -262.67, -96.74], p = .000 , d = 0.35, BF = 325.648” 

# 6) Is the HighAgr advisor selected more often?

## 6.i) Overall

We want to know whether the advisor who agrees more is selected more often by the participant when a choice is offered.

We will find this out by taking the number of times each participant selected the high agreement advisor and dividing by the total number of choice trials for that participant (should be the same for all participants). We can then take the mean of this proportion across participants and test it for significant versus the null hypothesis of random picking (0.5).

tmp <- aggregate(adviceType ~ pid,   
 data = trials[trials$type==trialTypes$choice, ],  
 FUN = function(x)sum(x==adviceTypes$HighAgr)/length(x))  
md.ttest(tmp$adviceType, mu=0.5)

*t*(49) = 5.43, *p* < .001, *d* = 0.77, BF = 9818.09; *M* = 0.61 [0.57, 0.65], mu = 0.5

## 6.ii) Medium-confidence trials

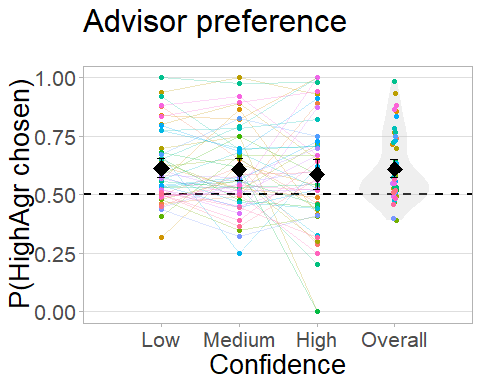
And the same with mid-confidence trials only:

tmp <- aggregate(adviceType ~ pid,   
 data = trials[trials$type==trialTypes$choice   
 & trials$confidenceCategory==confidenceCategories$medium, ],  
 FUN = function(x)sum(x==adviceTypes$HighAgr)/length(x))  
md.ttest(tmp$adviceType, mu=0.5)

*t*(49) = 4.39, *p* < .001, *d* = 0.62, BF = 368.14; *M* = 0.61 [0.56, 0.66], mu = 0.5

## 6.iii) Graph: Advisor preference by confidence category

Proportion of the time each participant picked the high agreement advisor. Connected points of a colour indicate data from a single participant, while the diamond indicates the mean proportion across all participants. The dashed reference line indicates picking both advisors equally, as would be expected by chance. Error bars give 95% bootstrapped confidence intervals.

This graph is likely to change in the write-up because confidence categories aren’t very useful for these advisors 

# 7) ANOVA investigating influence

## 7.i) Adjusted influence, all trials

Influence is defined as the extent to which the judge’s (participant’s) final decision has moved from their initial decision in the direction of the advice received. We begin by calculating influence for all trials and saving that information since it will come in handy for looking at influence on subsets of trials later. Below, we run an ANOVA using the influence data.

2x2x2 ANOVA investigating effects of advisor type (High/Low accuracy), choice (un/forced), and agreement (dis/agree) on influence. These are all within-subjects manipulations.

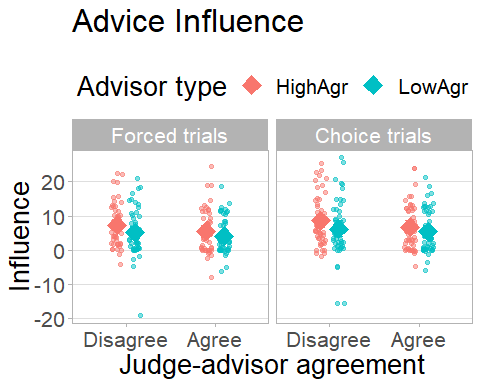
## [1] "2x2x2 Mixed ANOVA of advisor type x choice x agreement"

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effect | DFn | DFd | F | p | p<.05 | ges |
| 2 | adviceType | 1 | 49 | 7.77 | .008 | \* | .0181 |
| 3 | hasChoice | 1 | 49 | 13.64 | .001 | \* | .0078 |
| 4 | advisorAgrees | 1 | 49 | 2.20 | .144 |  | .0106 |
| 5 | adviceType:hasChoice | 1 | 49 | 0.00 | .969 |  | .0000 |
| 6 | adviceType:advisorAgrees | 1 | 49 | 2.72 | .106 |  | .0014 |
| 7 | hasChoice:advisorAgrees | 1 | 49 | 0.01 | .923 |  | .0000 |
| 8 | adviceType:hasChoice:advisorAgrees | 1 | 49 | 0.06 | .813 |  | .0000 |

## [1] "Mean|HighAgr=7.04 [95.00%CI: 6.11, 7.98] [Range: -7.87, 25.52]"  
## [1] "Mean|LowAgr=5.22 [95.00%CI: 4.26, 6.18] [Range: -19.10, 27.00]"  
## [1] "Mean|Choice=6.73 [95.00%CI: 5.71, 7.74] [Range: -15.56, 27.00]"  
## [1] "Mean|Forced=5.54 [95.00%CI: 4.65, 6.42] [Range: -19.10, 24.41]"  
## [1] "Mean|Agree=5.44 [95.00%CI: 4.57, 6.30] [Range: -7.87, 24.41]"  
## [1] "Mean|Disagree=6.83 [95.00%CI: 5.80, 7.86] [Range: -19.10, 27.00]"

## NULL

## 7.ii) Graph: Adjusted Advice influence, all trials

Influence of advice under varied conditions. Points indicate mean values for a participant, while diamonds indicate the mean of participant means, with error bars specifying 95% confidence intervals. 

## 7.iii) Adjusted influence, medium-confidence trials

The bias-sharing advisor and anti-bias advisors differ in their frequency with which they agree with the participant as a function of participant confidence (by design). To control for background effects where people are influenced different amounts depending on their own initial confidence, we also look at only those trials where participant confidence was in the mid-range (i.e. where both advisors agree 70% of the time, and thus where agreement and confidence balance out). This subset only includes trials on which the participant was CORRECT. Where incorrect, advisors also agree equally often (30% of the time), so these trials could be included.

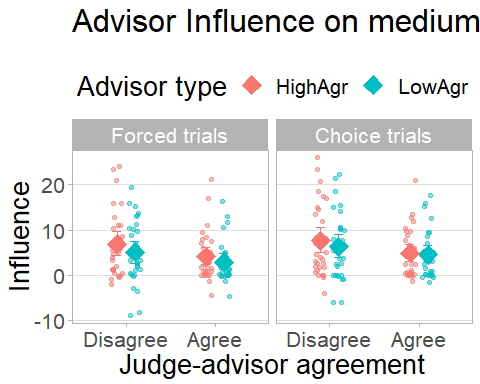
## Warning: You have removed one or more Ss from the analysis. Refactoring  
## "pid" for ANOVA.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effect | DFn | DFd | F | p | p<.05 | ges |
| 2 | adviceType | 1 | 29 | 3.66 | .066 |  | .0082 |
| 3 | hasChoice | 1 | 29 | 5.43 | .027 | \* | .0075 |
| 4 | advisorAgrees | 1 | 29 | 4.69 | .039 | \* | .0347 |
| 5 | adviceType:hasChoice | 1 | 29 | 0.37 | .545 |  | 5e-04 |
| 6 | adviceType:advisorAgrees | 1 | 29 | 1.09 | .304 |  | .0010 |
| 7 | hasChoice:advisorAgrees | 1 | 29 | 0.04 | .838 |  | 1e-04 |
| 8 | adviceType:hasChoice:advisorAgrees | 1 | 29 | 0.03 | .855 |  | .0000 |

## [1] "Mean|HighAgr=5.83 [95.00%CI: 4.61, 7.06] [Range: -4.57, 26.00]"  
## [1] "Mean|LowAgr=4.68 [95.00%CI: 3.55, 5.82] [Range: -9.00, 22.25]"  
## [1] "Mean|Choice=5.81 [95.00%CI: 4.59, 7.02] [Range: -6.00, 26.00]"  
## [1] "Mean|Forced=4.71 [95.00%CI: 3.57, 5.85] [Range: -9.00, 24.00]"  
## [1] "Mean|Agree=4.06 [95.00%CI: 3.12, 5.00] [Range: -4.67, 21.25]"  
## [1] "Mean|Disagree=6.46 [95.00%CI: 5.11, 7.81] [Range: -9.00, 26.00]"

## NULL

And the graph for medium-confidence trials only



## 7.iv) Raw influence, all trials

N.B. This is not a core analysis!

2x2x2 ANOVA investigating effects of advisor type (agree-in-confidence/uncertainty), choice (un/forced), and agreement (dis/agree) on influence. These are all within-subjects manipulations.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effect | DFn | DFd | F | p | p<.05 | ges |
| 2 | adviceType | 1 | 49 | 6.00 | .018 | \* | .0130 |
| 3 | hasChoice | 1 | 49 | 10.80 | .002 | \* | .0061 |
| 4 | advisorAgrees | 1 | 49 | 18.27 | .000 | \* | .0967 |
| 5 | adviceType:hasChoice | 1 | 49 | 3.55 | .065 |  | 9e-04 |
| 6 | adviceType:advisorAgrees | 1 | 49 | 4.31 | .043 | \* | .0026 |
| 7 | hasChoice:advisorAgrees | 1 | 49 | 0.28 | .597 |  | 1e-04 |
| 8 | adviceType:hasChoice:advisorAgrees | 1 | 49 | 1.34 | .253 |  | 3e-04 |

## [1] "Mean|HighAgr=9.82 [95.00%CI: 8.26, 11.39] [Range: -23.31, 67.25]"  
## [1] "Mean|LowAgr=7.36 [95.00%CI: 5.76, 8.96] [Range: -19.10, 78.37]"  
## [1] "Mean|Choice=9.43 [95.00%CI: 7.78, 11.09] [Range: -15.56, 75.93]"  
## [1] "Mean|Forced=7.75 [95.00%CI: 6.23, 9.27] [Range: -23.31, 78.37]"  
## [1] "Mean|Agree=5.08 [95.00%CI: 4.15, 6.02] [Range: -23.31, 24.41]"  
## [1] "Mean|Disagree=12.10 [95.00%CI: 10.17, 14.03] [Range: -19.10, 78.37]"

## NULL

# 7.iv.i) Raw influence adjustments

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P.influenceChanges. | clLow | clHigh | rangeLow | rangeHigh |
| 0.0815833 | 0.0585631 | 0.1046035 | 0 | 0.3583333 |

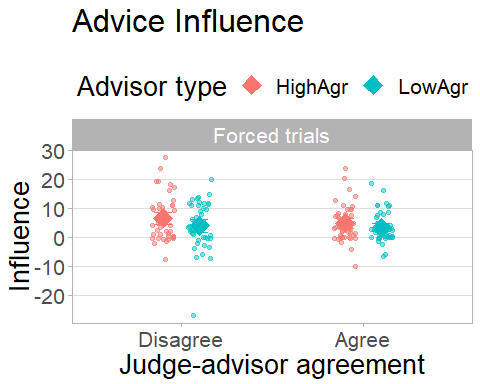
# 7.v) Forced trials, Med confidence

## Warning: You have removed one or more Ss from the analysis. Refactoring  
## "pid" for ANOVA.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effect | DFn | DFd | F | p | p<.05 | ges |
| 2 | adviceType | 1 | 44 | 3.48 | .069 |  | .0198 |
| 3 | advisorAgrees | 1 | 44 | 1.20 | .279 |  | .0079 |
| 4 | adviceType:advisorAgrees | 1 | 44 | 1.10 | .299 |  | .0014 |

## [1] "Mean|HighAgr=5.80 [95.00%CI: 4.34, 7.26] [Range: -9.88, 27.50]"  
## [1] "Mean|LowAgr=3.87 [95.00%CI: 2.47, 5.28] [Range: -26.80, 20.25]"  
## [1] "Mean|Agree=4.23 [95.00%CI: 3.01, 5.45] [Range: -9.88, 23.86]"  
## [1] "Mean|Disagree=5.44 [95.00%CI: 3.81, 7.08] [Range: -26.80, 27.50]"

## NULL



# 8) Trust questionnaire answers

## 8.i) Bayesian no-difference tests for advisor properties

We want to show that the randomly assigned advisor race/age/portrait/name had no effect. We will do this by showing that they did not differ between timepoints.

### 8.i.i) Race

|  |  |  |  |
| --- | --- | --- | --- |
| variable | BF | mu1 | mu2 |
| likeability | 0.2744751 | 68.56164 | 71.87008 |
| ability | 0.1993088 | 64.58904 | 67.26772 |
| benevolence | 0.3157981 | 66.58904 | 70.44094 |

### 8.i.ii) Age

|  |  |
| --- | --- |
| variable | corellationBF |
| likeability | 0.1272621 |
| ability | 0.2705326 |
| benevolence | 0.1562775 |

### 8.i.iii) Portrait

|  |  |
| --- | --- |
| variable | BF |
| likeability | 0.1611990 |
| ability | 0.0579154 |
| benevolence | 0.0453045 |

### 8.i.iv) Name

|  |  |
| --- | --- |
| variable | BF |
| likeability | 1.4603288 |
| ability | 0.3959631 |
| benevolence | 1.1639185 |

If any of the above do show significant differences then we’ll have to show that the factors which differ are not systematically linked to the advice type in order to demonstrate that they’re not responsible for any advice type effects we observe

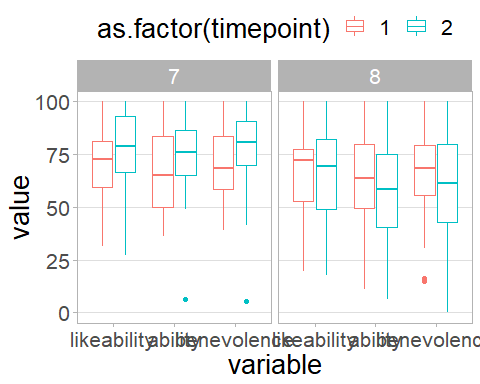
## 8.ii) 2x2 AdviceType x Timepoint MANOVA

**TODO** Check this MANOVA actually accounts for multiple observations per participant

## Df Pillai approx F num Df den Df Pr(>F)   
## adviceType 1 0.052814 7.3229 3 394 8.673e-05 \*\*\*  
## timepoint 1 0.017213 2.3002 3 394 0.076854 .   
## adviceType:timepoint 1 0.028890 3.9071 3 394 0.009032 \*\*   
## Residuals 396   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | domain | adviceType | timepoint | mu | ci95low | ci95hi |
| 1 | likeability | HighAgr | 1 | 70.65 | 65.66 | 75.64 |
| 3 | likeability | LowAgr | 1 | 68.38 | 62.93 | 73.83 |
| 5 | ability | HighAgr | 1 | 67.83 | 62.25 | 73.41 |
| 7 | ability | LowAgr | 1 | 66.09 | 60.13 | 72.05 |
| 9 | benevolence | HighAgr | 1 | 71.64 | 66.76 | 76.52 |
| 11 | benevolence | LowAgr | 1 | 67.52 | 61.94 | 73.10 |
| 2 | likeability | HighAgr | 2 | 78.41 | 73.68 | 83.14 |
| 4 | likeability | LowAgr | 2 | 65.21 | 58.57 | 71.85 |
| 6 | ability | HighAgr | 2 | 74.22 | 69.25 | 79.19 |
| 8 | ability | LowAgr | 2 | 57.02 | 50.24 | 63.80 |
| 10 | benevolence | HighAgr | 2 | 77.73 | 72.24 | 83.22 |
| 12 | benevolence | LowAgr | 2 | 59.25 | 51.77 | 66.73 |

## 8.iii) Graph: Pro/retrospective assessments by advice type and dimension



# 9) Do participants simply prefer agreement?

If so, we should see that participants preferentially pick the HighAgr advisor when their initial confidence is high, and LowAgr when their initial confidence is low. We can t-test HighAgr pick proportion in high-confidence vs HighAgr pick proportion in low-confidence.

## 9.i) Pick rate in low- vs high-confidence trials

## [1] "Choice proportion HighAgr in low- vs high-confidence trials"

## [1] "t(49) = 0.8 [95%CI: -0.04, 0.09], p = .425 , d = 0.13"

## [1] "t(49) = 0.8 [95%CI: -0.04, 0.09], p = .425 , d = 0.13"

## [1] "low=0.61 [95.00%CI: 0.57, 0.66] [Range: 0.32, 1.00]"

## [1] "low=0.61 [95.00%CI: 0.57, 0.66] [Range: 0.32, 1.00]"

## [1] "high=0.59 [95.00%CI: 0.52, 0.65] [Range: 0.00, 1.00]"

## [1] "high=0.59 [95.00%CI: 0.52, 0.65] [Range: 0.00, 1.00]"

## [1] "Bayesian examination of above (prior = mean diff of 0, sd as empirically observed)"

## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.707 : 0.2088416 ±0%  
##   
## Against denominator:  
## Null, mu = 0   
## ---  
## Bayes factor type: BFoneSample, JZS

## [1] "Evidence strength for differential HighAgr picking strategy by confidence: BF=0.209"

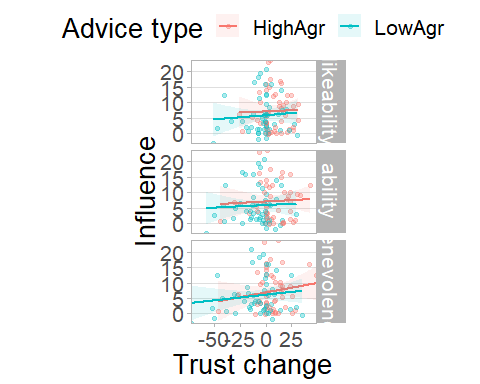
# 10) Subjective-objective correlation

## 10.i) Questionnaire-influence correlation

Participants rate advisors on three factors: ability, benevolence, and likeability. We can investigate these ratings for correlations with the objective influence measure. We would expect ability to show the strongest correlation because it relates to expertise in the literature and should be the primary dimension of variation in appraisal (indeed our theoretical model only considers ability assessments). Benevolence is unlikely to change much, though some participants may use benevolence to compensate for advisors they believe to be high in ability and yet still not very influential (because they’re deliberately giving bad advice). Likeability, reasoning from the results of our previous study, is likely to be largely orthogonal to advice.

##   
## Call:  
## lm(formula = influence ~ ability + benevolence + likeability,   
## data = tmp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.727 -4.864 -1.750 3.767 17.270   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.067835 0.599460 10.122 <2e-16 \*\*\*  
## ability 0.009468 0.042121 0.225 0.8224   
## benevolence 0.062678 0.036432 1.720 0.0869 .   
## likeability -0.056173 0.039785 -1.412 0.1596   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.042 on 196 degrees of freedom  
## Multiple R-squared: 0.03322, Adjusted R-squared: 0.01843   
## F-statistic: 2.245 on 3 and 196 DF, p-value: 0.08433

# 10.ii) Graph: Questionnaire-influence correlation



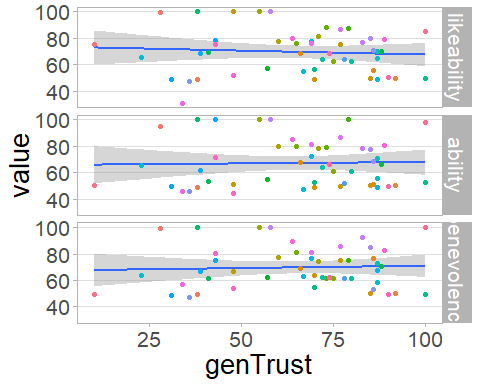
# 11) Generalised Trust

# 11.i) Generalised Trust and subjective assessments

Generalised Trust is a measure of the propensity to trust, so we expect it to correlate with the initial scores for the advisor questionnaires

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | variable | corellation | p.value | method |
| cor | likeability | -0.0800409 | 0.5805734 | Pearson’s product-moment correlation |
| cor1 | ability | 0.0266435 | 0.8542762 | Pearson’s product-moment correlation |
| cor2 | benevolence | 0.0495799 | 0.7324065 | Pearson’s product-moment correlation |

### 11.ii.i) Graph of Generalised Trust and subjective assessments

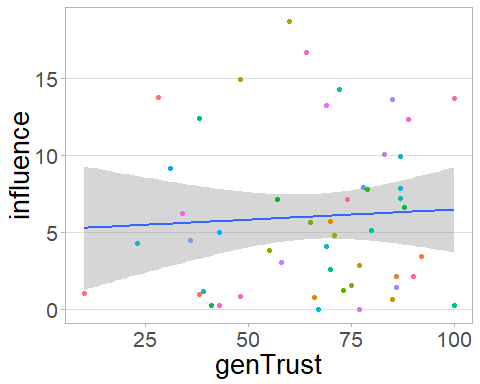


## 11.ii) Generalised Trust and influence

Generalised trust should also correlate with influence given that influence is supposedly a manifestation of trust

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | variable | corellation | p.value | method |
| cor | rawInfluence | -0.0261773 | 0.8567987 | Pearson’s product-moment correlation |
| cor1 | influence | 0.0567883 | 0.6952699 | Pearson’s product-moment correlation |

### 11.ii.i) Graph of Generalised Trust and influence



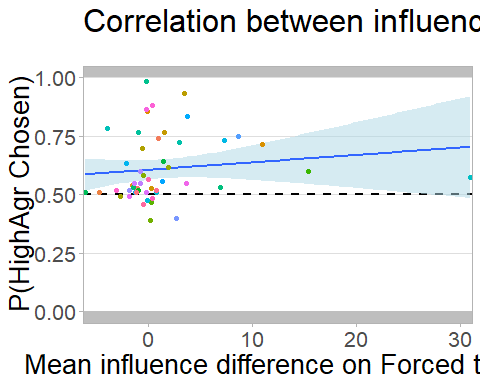
# 12) Participant effort

## 12.i) Preference strength and influence difference

### 12.i.i) By advice profile

##   
## Pearson's product-moment correlation  
##   
## data: df.xii.i.i$HighAgrPref and df.xii.i.i$InfluenceDiff  
## t = 0.88211, df = 48, p-value = 0.3821  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1575852 0.3909076  
## sample estimates:  
## cor   
## 0.1263026

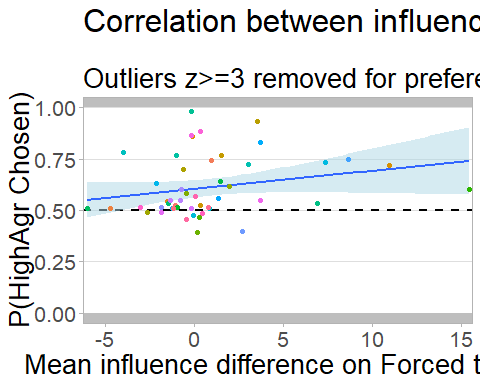
## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.333 : 0.4493166 ±0%  
##   
## Against denominator:  
## Null, rho = 0   
## ---  
## Bayes factor type: BFcorrelation, Jeffreys-beta\*



Looking at the above but removing outliers beyond 3sd.

##   
## Pearson's product-moment correlation  
##   
## data: df.xii.i.i.2$HighAgrPref and df.xii.i.i.2$InfluenceDiff  
## t = 1.631, df = 47, p-value = 0.1096  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.05320671 0.48132305  
## sample estimates:  
## cor   
## 0.2314528

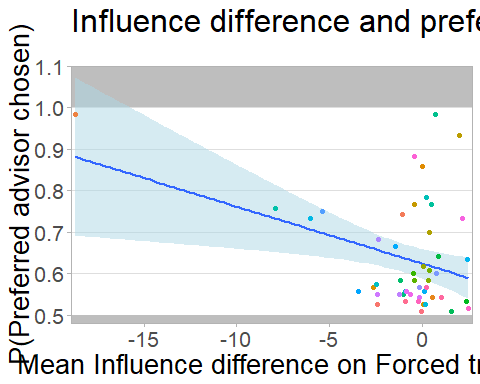
## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.333 : 1.017425 ±0%  
##   
## Against denominator:  
## Null, rho = 0   
## ---  
## Bayes factor type: BFcorrelation, Jeffreys-beta\*



### 12.i.ii) By advisor

##   
## Pearson's product-moment correlation  
##   
## data: df.xii.i.ii$PrefStrength and df.xii.i.ii$InfluenceDiff  
## t = -2.6246, df = 48, p-value = 0.0116  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.5758303 -0.0842218  
## sample estimates:  
## cor   
## -0.3542645

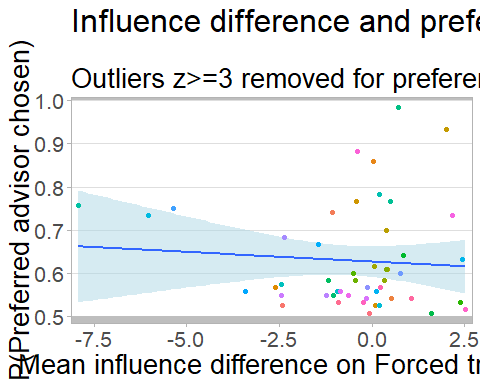
## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.333 : 5.625745 ±0%  
##   
## Against denominator:  
## Null, rho = 0   
## ---  
## Bayes factor type: BFcorrelation, Jeffreys-beta\*



Looking at the above but removing outliers beyond 3sd.

##   
## Pearson's product-moment correlation  
##   
## data: df.xii.i.ii.2$PrefStrength and df.xii.i.ii.2$InfluenceDiff  
## t = -0.53261, df = 47, p-value = 0.5968  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.3510078 0.2082766  
## sample estimates:  
## cor   
## -0.07745635

## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.333 : 0.3644273 ±0%  
##   
## Against denominator:  
## Null, rho = 0   
## ---  
## Bayes factor type: BFcorrelation, Jeffreys-beta\*



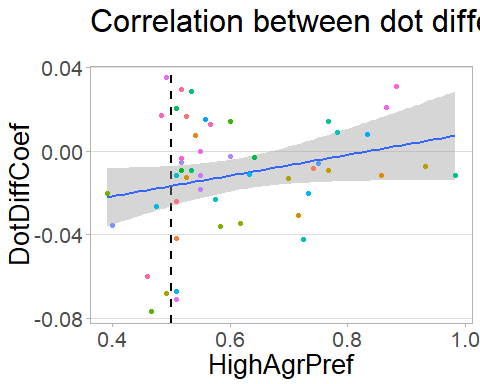
## 12.ii) Task performance and preference strength

### 12.ii.i) Dot difference stability

The slope of a participant’s dot difference change over time is indicative of their consistency. For participants concentrating and applying themselves, their performance after the practice period should be relatively stable, resulting in a very flat slope. For participants whose performance markedly deteriorates, the slope will be larger. Participants who improve will show a negative relationship between dot difference and trial id.  
Participants with a non-linear dot difference pattern appear to be rare enough to ignore. Insofar as the slope is indicative of application to the task, it should correlate with identifying the better advisor.

##   
## Pearson's product-moment correlation  
##   
## data: df.xii.ii.i$HighAgrPref and df.xii.ii.i$DotDiffCoef  
## t = 1.874, df = 48, p-value = 0.06702  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.01859394 0.50290288  
## sample estimates:  
## cor   
## 0.2611052

## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.333 : 1.445146 ±0%  
##   
## Against denominator:  
## Null, rho = 0   
## ---  
## Bayes factor type: BFcorrelation, Jeffreys-beta\*



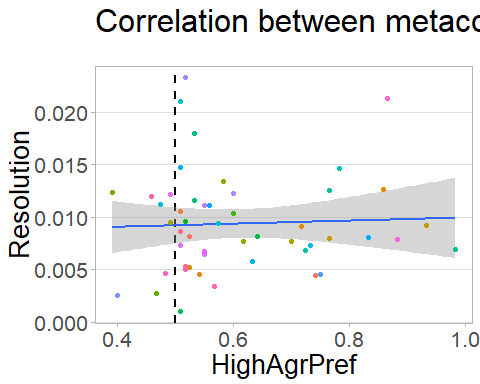
### 12.ii.ii) Metacognitive resolution

A high metacognitive resolution means that the probability of being correct increases as confidence increases. Participants not attending to the task are likely to have low resolution, and thus it may function as a method of identifying application to the task. If this is true, we expect a high resolution to correlate with a marked preference for the more accurate advisor.

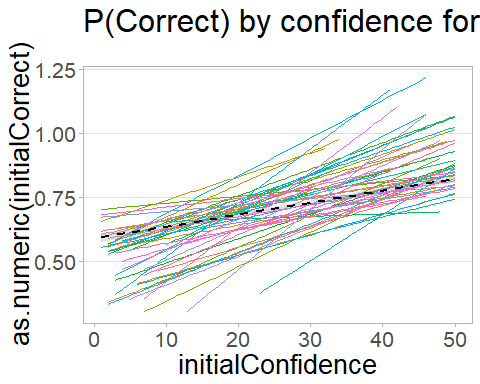
**Note**: High resolution is more truly an indication of good metacognitive performance, which may independantly predict advisor preference.

##   
## Pearson's product-moment correlation  
##   
## data: df.xii.ii.ii$HighAgrPref and df.xii.ii.ii$Resolution  
## t = 0.3112, df = 48, p-value = 0.757  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.2364281 0.3192331  
## sample estimates:  
## cor   
## 0.04487268

## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.333 : 0.3323679 ±0%  
##   
## Against denominator:  
## Null, rho = 0   
## ---  
## Bayes factor type: BFcorrelation, Jeffreys-beta\*



We may as well look at participants’ resolution graphs, too.



# 13) Preference and questionnaire responses

Participants who prefer an advisor may rate that advisor more highly. If participants have made up their mind about their favourite advisor prior to any advice, initial questionnaire responses should be predictive of preferential picking. If not, picking should be more closely associated with positive change in questionnaire responses.

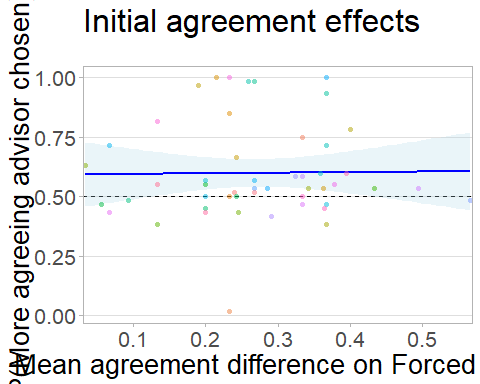
We examine for each questionnaire dimension the association between the difference in answers for the advisors and the preference strength for the most frequently picked advisor. We also examine the difference between advisors of the improvement in questionnaire response from before to after advice.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | v | Correlation | p | BF | DiffCorrelation | pDiff | DiffBF |
| cor | Likeability | 0.2890807 | 0.0417396 | 2.0644280 | -0.1307327 | 0.3655022 | 0.4606056 |
| cor1 | Ability | 0.3005799 | 0.0339234 | 2.4201839 | -0.2319678 | 0.1050243 | 1.0413482 |
| cor2 | Benevolence | 0.2084807 | 0.1462476 | 0.8249612 | -0.0749907 | 0.6047500 | 0.3592164 |

# 14) Does agreement in Forced blocks predict picking in Choice blocks?

##   
## Pearson's product-moment correlation  
##   
## data: df.xiv$agreementDiff and df.xiv$pickRate  
## t = 0.093408, df = 48, p-value = 0.926  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.2658644 0.2907377  
## sample estimates:  
## cor   
## 0.013481

## Bayes factor analysis  
## --------------  
## [1] Alt., r=0.333 : 0.3195553 ±0%  
##   
## Against denominator:  
## Null, rho = 0   
## ---  
## Bayes factor type: BFcorrelation, Jeffreys-beta\*

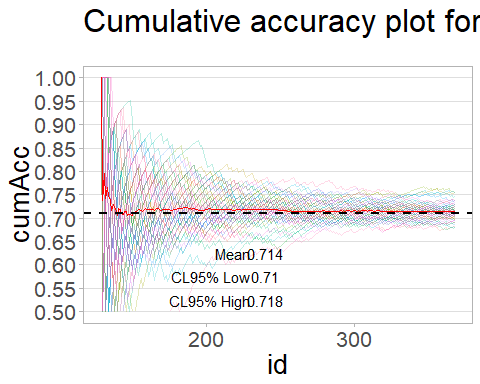


# 15) Cumulative accuracy plot

The staircase procedure should control participant accuracy to be approximately 71% over the course of the experiment. This graph shows cumulative accuracy for all non-practice trials.

## Warning: Removed 148 rows containing non-finite values (stat\_summary).  
  
## Warning: Removed 148 rows containing non-finite values (stat\_summary).

## Warning: Removed 91 rows containing missing values (geom\_path).



# Additional/In progress

# 12) Predicting advisor from confidence ####  
print('12 Predicting advisor from confidence')  
  
df.xii <- NULL  
for(pid in unique(trials$pid)) {  
 tmp <- trials[trials$pid==pid, ]  
 tmp$adviceType <- as.factor(tmp$adviceType)  
 ans <- glm(adviceType ~ initialConfidence, tmp, family = binomial(link = "logit"))  
 s <- summary(ans)  
 df.xii <- rbind(df.xii, data.frame(pid,  
 coef = s$coefficients[2,1],  
 p = s$coefficients[2,4]))  
}  
tmp <- df.xii  
names(tmp)[2] <- 'y'  
tmp$x <- 0  
gg.xii <- ggplot(tmp, aes(x = p)) + geom\_freqpoly(binwidth = 0.1)   
gg.xii  
  
  
# 13) Influence strength by confidence category ####  
print('13 Influence strength by confidence category')  
  
tmp <- aggregate(influence ~ confidenceCategory + pid, data = trials, FUN = mean)  
gg.xiii <- ggplot(tmp, aes(x = confidenceCategory, y = influence)) +   
 geom\_point(aes(colour = as.factor(pid)), alpha = 0.5) +   
 geom\_smooth(method = 'lm',   
 aes(group = as.factor(pid), colour = as.factor(pid)),  
 se = F,  
 alpha = 0.2) +  
 stat\_summary(geom = 'errorbar', fun.data = mean\_cl\_boot) +  
 style.long  
gg.xiii  
  
summary(aov(influence ~ confidenceCategory, data = tmp))  
  
# 14) Confidence autocorrelation plots by participant ####  
for(pid in unique(trials$pid)) {  
 tmp <- trials[trials$pid == pid, ]  
 ggplot(tmp, aes(x = initialConfSpan, y = finalConfSpan)) +  
 geom\_polygon(data = df.poly1, aes(x,y), fill = 'grey', alpha = 0.2) +  
 geom\_polygon(data = df.poly2, aes(x,y), fill = 'grey', alpha = 0.2) +  
 geom\_point(alpha = 0.2, aes(color = factor(finalCorrect))) +  
 geom\_abline(slope = 1, intercept = 0, linetype = 'dashed', size = 1, color = 'black') +  
 scale\_color\_discrete(name = 'Final judgement', labels = c('Incorrect', 'Correct')) +  
 scale\_x\_continuous(limits = c(-50,50), expand = c(0,0)) +  
 scale\_y\_continuous(limits = c(-50,50), expand = c(0,0)) +  
 facet\_grid(~advisorAgrees, labeller = as\_labeller(c('FALSE'='Disagree', 'TRUE'='Agree'))) +  
 labs(title = paste('PID:', pid, 'Initial vs final confidence'),  
 legend = NULL,  
 x = 'Initial confidence',  
 y = "Final confidence") +  
 coord\_fixed() +  
 style +   
 theme(panel.spacing = unit(1.5, 'lines'),  
 plot.margin = unit(c(0,1,0,0.5), 'lines'))  
 #ggsave(paste0('explore/autocorrelations/pid', pid, '.png'), width = 8.96, height = 5.97, units = 'in')  
}  
  
# 15) Examining dot difference ####  
library(gganimate)  
g <- ggplot(trials, aes(x = id, y = dotDifference)) +  
 geom\_line() +  
 geom\_smooth(method = 'lm') +  
 scale\_x\_continuous(limits = c(0,246)) +  
 style.long +  
 labs(title = 'Participant {frame\_time}') +  
 transition\_time(pid)  
  
animate(g, fps = 1)   
  
  
ggplot(all.trials, aes(x = id, y = dotDifference, colour = as.factor(pid))) +  
 geom\_line() +  
 geom\_smooth(method = 'lm', se = F, data = all.trials[all.trials$block == max(all.trials$block), ]) +  
 scale\_x\_continuous(limits = c(0,249)) +  
 style.long +  
 labs(title = 'Difficulty (and trend for the final block) for Advisor Choice (accuracy)')