**Investigation Online Experiment – April 2020**

**Description of the data files**

**Step 1 – All parts**

**[Read CSV and generate three “raw data” summary CSV files, one per task]**

- The columns for the three files are identical (as in the raw separate files)

- All strings variables are replaced with integers (e.g. the variable "task=part1-guilty" becomes two separate variables, "part" [values 1/2/3] and treatment [values 0=guilty/1=innocent])

**PART 1 – FIXED NUMBER OF ROUNDS**

**Part 1 - Step 2**

**[Clean the variables to facilitate analysis]**

* Participant\_ID [progressive number for each participant]
* treatment [1 for guilty treatment, 0 for innocent treatment]
* part (this will always be 1 within the file "TASK1")
* trial\_no (progressive number for each trial, reset to 1 at the beginning of each participant)
* rounds (number of rounds available, 1-9)
* prior (prior probability of red being guilty, 0.1-0.9) – check if the sign of prob is correct
* true\_guilty\_suspect (0=red, 1=blue)
* accused\_suspect (0=red, 1=blue)
* 9 columns “choice” for the investigations (0=red, 1=blue)
* 9 columns “outcome” for the outcome of the investigation (1= evidence found/0=no-evidence)
* 11 columns “timing\_choice” (milliseconds for each action) – up to 9 investigations + accusation + move to the next screen
* correct\_suspect\_accused (1 if correct, 0 otherwise)
* evidence\_round (X indicates that the first evidence was observed in round X, 0 if not found)

**Part 1 – Step 3**

**[Same as step 2, but we make the “guilty” and the “innocent” treatments compatible]**

The only change is in the nine investigation columns “choice”, and only for the “innocent treatments”. When treatment=0 (innocent treatment), now the code is

* 9 columns “choice” for the investigation (0=blue, 1=red)

No change in any other column.

[right now the change is for the code treatment=1 – we can check it quickly together during next meeting]

**Part 1 – Step 4**

**[Move from a line-for-trial to a line-for-round visualization of the dataset]**

* Participant\_ID - progressive number for each participant
* Treatment - 1 for "guilty" treatment, 0 for "innocent" treatment
* Part (this will always be 1 within the file "TASK1")
* Trial\_no (progressive number for each trial, reset to 1 at the beginning of each participant)
* Rounds (number of rounds available, 1-9)
* Red\_prior\_probability (prior probability of red being guilty, 0.1-0.9)
* Current\_round - Each round is an action (investigate/accuse)
* Action\_type - 0 if investigate, 1 if accuse
* Suspect - target of the action: 0 if red, 1 if blue
* Count\_red\_samples - number of red samples observed before the current action - update for the "accusation round"
* Count\_blue\_samples - number of blue samples observed before the current action - update for the "accusation round"
* Evidence\_found - initialize at 0 (always in round 1). After a "fully revealing evidence" is found, change this to 1 including in the current round and keep it like that the rest of the trial (including when the player accuses)
* Evidence\_suspect - initialize at -1 (always in round 1). After a "fully revealing evidence" is found [and the previous column has 1], change this to 0 (if red is found guilty) or 1 (f blue is found guilty)
* Posterior - Posterior probability of red being guilty
* True\_guilty\_suspect - 0 if red, 1 if blue
* Suspect\_accused - 0 if red, 1 if blue
* Timing - Time used to make the decision

**PART 2 – VARIABLE NUMBER OF ROUNDS – SYMMETRIC COST**

**Part 2 - Step 2**

**[Clean the variables to facilitate analysis]**

* Participant\_ID [progressive number for each participant]
* treatment [1 for guilty treatment, 0 for innocent treatment]
* Part (this will always be 2 within the file "TASK2")
* trial\_no (progressive number for each trial, reset to 1 at the beginning of each participant)
* NO COLUMN FOR rounds
* prior (prior probability of red being guilty, 0.1-0.9)
* NEW cost\_red (cost for collecting a red sample, between 5 and 80)
* NEW cost\_blue (cost for collecting a red sample, between 5 and 80)
* true\_guilty\_suspect (0=red, 1=blue)
* accused\_suspect (0=red, 1=blue)
* 60 columns “choice” for the investigations (0=red, 1=blue)
* 60 columns “outcome” for the outcome of the investigation (1= evidence found/0=no-evidence)
* 62 columns “timing\_choice” (milliseconds for each action) – up to 9 investigations + accusation + move to the next screen
* correct\_suspect\_accused (1 if correct, 0 otherwise)
* evidence\_round (X indicates that the first evidence was observed in round X, 0 if not found)

**Part 2 – Step 3**

**[Same as step 2, but we make the “guilty” and the “innocent” treatments compatible]**

The only change is in the 60 investigation columns “choice”, and only for the “innocent treatments”. When treatment=0 (innocent treatment), now the code is

* 60 columns “choice” for the investigation (0=blue, 1=red)

No change in any other column.

**Part 2 – Step 4**

**[Move from a line-for-trial to a line-for-round visualization of the dataset]**

* Participant\_ID - progressive number for each participant
* Treatment - 1 for "guilty" treatment, 0 for "innocent" treatment
* Part (this will always be 2 within the file "TASK2")
* Trial\_no (progressive number for each trial, reset to 1 at the beginning of each participant)
* NO COLUMN FOR rounds
* Red\_prior\_probability (prior probability of red being guilty, 0.1-0.9)
* NEW cost\_red (cost for collecting a red sample, between 5 and 80)
* NEW cost\_blue (cost for collecting a red sample, between 5 and 80)
* Current\_round - Each round is an action (investigate/accuse)
* Action\_type - 0 if investigate, 1 if accuse
* Suspect - target of the action: 0 if red, 1 if blue
* Count\_red\_samples - number of red samples observed before the current action - update for the "accusation round"
* Count\_blue\_samples - number of blue samples observed before the current action - update for the "accusation round"
* Evidence\_found - initialize at 0 (always in round 1). After a "fully revealing evidence" is found, change this to 1 from the following round and for the rest of the trial (including when the player accuses) – [now the change occurs in the same round instead of the following one]
* Evidence\_suspect - initialize at -1 (always in round 1). After a "fully revealing evidence" is found [and the previous column has 1], change this to 0 (if red is found guilty) or 1 (f blue is found guilty)
* Posterior - Posterior probability of red being guilty
* True\_guilty\_suspect - 0 if red, 1 if blue
* Suspect\_accused - 0 if red, 1 if blue
* Timing - Time used to make the decision

**PART 3 – VARIABLE NUMBER OF ROUNDS – ASYMMETRIC COST**

**Part 3 - Step 2**

**[Clean the variables to facilitate analysis]**

The only difference here is that the sampling cost is asymmetric (the raw file contains two costs, one for red and one for blue).

**OVERVIEW OF THE COMMENTS:**

Things to add

* Changes for task 2 (compared to task 1)

Things to check

* Probabilities – check what probability is displayed (red or blue being guilty)
* Treatment – check which treatment is coded as 0 or 1

Things to fix

* Evidence\_found – in the round in which it is found, it should not appear as “found”