**Last updated by Marjan on 23rd of July 2024**

This document explains what the columns in the excel file “Surprise\_task\_pilots\_development.xlsx” contain. To create our surprise task, we have run several pilots either about the different components of the task (e.g. to choose the best images for the virtual players) or find the best version of the task to be used in the final experiment (which turned out to be surprise pilot 21).

There were some additional analyses done that are not included here as they were not about the task design. For example, analysing the acoustic features for pilot 22, which is also placed on [GitHub](https://github.com/transatlantic-comppsych/aim_lab_1/tree/main/surprise_pilot_22_acoustic_feature_analysis) but not mentioned in this excel document; or the two attention pilots that we decided to pause, which can be found [here](https://github.com/transatlantic-comppsych/aim_lab_1/tree/main/De_identified_data/att_pilot_1/SUP_PRF_pilot_no_fdbk_attention_vid_v3) and [here](https://github.com/transatlantic-comppsych/aim_lab_1/tree/main/De_identified_data/att_pilot_2/SUP_PRF_pilot_no_fdbk_attention_vid_cbal_v5).

**Pilot\_name:**

This column shares the name of the corresponding pilots on Github and Gorilla with the format “surprise\_pilot\_#number”.

**Study\_description:**

Describing the study design, and how it is different from previous pilot and the reason for change. It is quite similar to the last column (Pilot\_outcome). However, Pilot\_outcome also discusses the findings of the pilot and the decisions we made.

**Changes\_summary:**

Very brief description of the changes made in the corresponding pilot. This column contains NA if a pilot was the very first pilot of its kind (so no changes).

**Changes\_details:**

A more detailed description of the changes.

**Task\_link\_Gorilla:**

The link to the Gorilla task, Elena is the owner of the experiments on this date (23rd of July 2024).

**GitHub\_link\_raw\_data:**

The link in this column takes you to the GitHub folder containing the de-identified raw data.

**GitHub\_link\_script\_&\_results:**

The link in this column takes you to the GitHub folder containing the scripts used to analyse the corresponding data, and also the PDF’s containing the output of those scripts. The scripts are usually in “.Rmd” format.

**Size\_of\_PE:**

This corresponds to the size of the feedback that the participants received from the virtual players, in order to create different sizes for the PE’s and to test their impact on our response variables mood and anxiety. For pilots where no feedback was provided either due to the design or if they were about other task components this column contains “NA” for Not Applicable.

**Judges\_feedback:**

This column explains how the feedback was calculated per judge and per pilot.

**Modality:**

This corresponds to the modality of the responses (e.g. written format (text), video/audio).

**Task\_instructions\_per\_pilot:**

The exact instructions per pilot are provided in this column. These are not the complete instructions as they would be too long to fit here, but rather the instructions about the narrative of the task and appeared always in the first few slides of each pilot in the instructions node before the actual task started.

**How\_FEEDBACK\_was\_provided:**

This column shows the exact phrase that was used per pilot before proving the percentage feedback.

**Prediction\_Question:**

This column shows the exact phrase that was used to ask participants to provide their predictions.

**anxiety\_Question:**

This column shows the exact phrase that was used to ask about momentary/trial-by-trial anxiety ratings on a Visual Analogue Scale ranging from 0 to 100.

**Mood\_Question:**

This column shows the exact phrase that was used to ask about momentary/trial-by-trial mood ratings on a Visual Analogue Scale ranging from 0 to 100.

**Certainty\_Question:**

This column shows the exact phrase that was used to ask about momentary/trial-by-trial certainty ratings on a Visual Analogue Scale ranging from 0 to 100. This question was dropped however as we did not find any correlations with PE or mood/anxiety ratings for most pilots.

**Other\_Questions:**

Indicating any other questions that were asked that were not described in other columns. This could be either about the questions asked at the end of the task, or questions that were asked in the initial pilots that did not have the main surprise task design to collect feedback statements or choose the picture of the judges.

**Feedback\_YPAG:**

This column provides feedback we received from the YPAG members that were taken into account to modify our task.

**Feedback\_participants:**

This column provides feedback we received from our participants that were taken into account to modify our task.

**Feedback\_professional(Eleanor/Georgina):**

This column provides feedback we received from Eleanor/Georgina/ that were taken into account to modify our task.

**Beta\_mood:**

This column provides the beta values, or correlation coefficients for the relationship between subjective PE’s and trial-by-trial mood ratings. For some earlier pilots we had looked at the average of all correlations to inform our decisions, and instead of beta’s the average correlation coefficient is reported.

**Beta\_anxiety:**

This column provides the beta values, or correlation coefficients for the relationship between subjective PE’s and trial-by-trial anxiety ratings. For some earlier pilots we had looked at the average of all correlations to inform our decisions, and instead of beta’s the average correlation coefficient is reported.

**model\_mood/model\_anxiety:**

Since the best model was sometimes different across pilots, in this column we report which model was used to arrive at the beta’s that were reported in previous columns. For some earlier pilots we used average correlations instead of the LME’s and we mention “average correlation” instead of the model.

**ICC\_mood:**

This column provides the ICC values for mood ratings within subjects. A very high ICC indicates low variability within subjects, a very low ICC indicates too much variability. A moderate ICC may be more ideal as we do expect some variability in response to different PEs within subjects, but also not too much.

**ICC\_anxiety:**

This column provides the ICC values for anxiety ratings across subjects.

**Changes\_Task\_instructions/narrative:**

This column indicates whether there were any changes in the task instructions or narrative for the corresponding pilot.

**Changes\_judge\_&\_hist\_presentation:**

This column indicates whether there were any changes in the judge pictures or histogram presentation in the corresponding pilot. This is about the screen before each social interaction where the judge and their histogram are presented.

**Changes\_in\_asking\_for\_prediction:**

This column indicates whether there were any changes in the task module/section where we ask for prediction.

**Changes\_in\_social\_interaction:**

This column indicates whether there were any changes in the task module/section where the social interaction takes place, i.e. where participants describe the images to the judges.

**Changes\_in \_Receiving\_feedback:**

This column indicates whether there were any changes in the screen where feedback is provided.

**Changes\_in \_Mood\_rating:**

This column indicates whether there were any changes in the screen where we asked about momentary mood.

**Changes\_in \_Anxiety\_rating:**

This column indicates whether there were any changes in the screen where we asked about momentary anxiety.

**Pilot\_outcome:**

This column indicates the main results of the pilot and what decision was made based on these findings.