README

DICE: Advancing Social Media Experiments Through Digital In-Context Experiments

November 16, 2024

Short-cut instructions

The reproduction package is prepared to allow a "push-button reproduction" of the two case studies presented in the paper. Few adjustments (if any at all) should be needed. Simply open the project file `_DICE.Rproj` to open R Studio and initiate R. From within this R project, open `case-study-1/analyses/02-analyses.qmd` and render the file (e.g. by pressing `command` + `shift` + `k`) to reproduce the results presented in the paper as well as `case-study-1/reports/02-analyses.pdf` using the pre-processed data. If you intend to reproduce the results from the raw data, run `case-study-1/analyses/01-pre-processing.qmd` first. That's it. Repeat the procedure for Case Study 2.

Open Science Framework (OSF)

We host our replication package on OSF. Please consider the following links.

- Review: https://osf.io/2xs5c/?view_only=4bf95d2a2c8449218b5fa7cd288f626a
- Permanent, not anonymized & private during review: https://osf.io/2xs5c/

Comprehensive instructions

At the end of each report, we print our session info, that is, the software environment (i.e., R version 4.4.1 using macOS Sonoma 14.4.1) and package versions we used. To increase reproducibilty, we used `{groundhog}` (version 3.2.1), a package version control system for R that allows use to download the packages versions we used. We used a MacBook Pro (2019) with a 2.3 GHz 8-Core Intel Core i9 processor and 16 GB 2400 MHz DDR4 memory.

Throughout the project, we followed a *literate programming* approach (Knuth, 1984) in which we blend our text with executable (quarto) code. This allows the reader of the source code to unambiguously identify how we computed almost any statistic, table or figure presented in the report.

All paths in the scripts are relative, i.e., no adjustments are needed and no working directory needs to be defined. It is important, however, that you access the project via the `_DICE.Rpoj` file at the root of this project.

The reproduction package involves two directories, one for each case study. Both directories contain the same sub-directories such that one can easily apply the steps required to reproduce Case Study 1 to Case Study 2 and vice versa. We describe these sub-directories in what follows

- The folder `analyses/` contains at least two files: `01-pre-processing.qmd`, which pre-processes the raw data and writes `data/processed/meme-feed-data.csv` or `data/processed/brand-safety-short.csv` as well as `data/processed/brand-safety-long.csv`. Note that the files differ slightly between the two Case Studies as we used our R Shiny app for pre-processing the DICE data in Cae Study 2. We did not do so in Case Study 1, to offer a comprehensive analysis script. `02-analyses.qmd` reads the processed data and produces the data visualizations, tables and statistics reported in the paper. In Case Study 2, we also provide `03-robustness-checks.qmd`, which uses an alternative specification to reproduce Tables and Figures. All scripts are Quarto (`*.qmd`) files that use R code. Quarto is *"the next generation version of R Markdown from Posit"*. Instead of rendering the whole document, you can also run its code chunks sequentially.
- The folder `data/` contains both raw and processed data. Please note that we edited the raw datasets slightly as we anonymized personal IDs collected. DICE's raw data files follow oTree's naming convention and are named `all-apps-wide.*.csv`. Also `PageTimes.*.csv` are DICE outputs that simply document time stamps of participants submitting the experiment's pages. The other files (`DICE_Brand_Safety_Aug_2024_August+15,+2024_23` or `DICE_Meme_Feed_October+15,+2024_18.35.csv`) stem from Qualtrics and document our post-feed questionnaire responses. Whereas, Case Study 1 only uses these data, in Case Study 2 we also read `DICE-processed-2024-10-15.csv`. This file is the output of our R Shiny app. The processed data (i.e., `meme-feed-data.csv` or `brand-safety-short.csv`) are used for our analyses. We describe the processed data in our documentation.
- The folder `img/` simply contains two images `design.png` and `stimuli.png` that are presented in the paper and are thus included in the reproduction report.
- The folder `stimuli/` contains the configuration CSV file (i.e., `brazil.csv` or `9gag.csv`) used to render the social media feeds (i.e., our stimuli) used in the respective study. See our documentation for further information. Here, we also provide the Qualtrics surveys as `*.qsf` and `*.pdf` files.

- The folder `report/` simply contains the result of rendering `analyses/02-analyses.qmd`, namely `02-analyses.pdf`.
- We also provide the `_DICE.Rpoj` file to open an R Session as well as a `references.bib` file required to render the citations in our final reports. In addtion, we also provide this README file, all of these files are stored at the root of the directory.

Merging data

As we work with Qualtrics and DICE data (i.e., the stimuli in combination with both raw and pre-processed outputs using our Shiny app), we merge these datasets in our `analyses/01-pre-processing.qmd` files. We do this based on two key identifiers: to merge the output data, we rely on the `participant.label` (which is a participant's Prolific ID, if applicable). To merge these outputs with the stimuli (arguably, the inputs) we use the unique `doc_id` a researcher assigns to each social media post specified in the `stimuli/.*.csv` files.

Estimated runtime

- Pre-processing scripts: < 5 minutes per case study
- Analysis scripts: < 5 minutes per case study
- Robustness checks (Case Study 2): < 5 minutes

Note on licensed data

This research does not rely on licensed databases. All data was collected through primary research using DICE/oTree, and Qualtrics platforms.

File tree

```
.
  README.pdf
  README (best for reading).html
  _references.bib
  case-study-1
      analyses
          01-pre-processing.qmd
          02-analyses.qmd
      data
```

```
        processed
            brand-safety-long.csv
            brand-safety-short.csv
        raw
            DICE_Brand_Safety_Aug_2024_August+15,+2024_23.12.csv
            PageTimes-2024-08-15.csv
            all_apps_wide_2024-08-15.csv
    img
        design.png
        stimuli.png
    report
        02-analyses.pdf
    stimuli
        brazil.csv
        DICE_Brand_Safety_Aug_2024.pdf
        DICE_Brand_Safety_Aug_2024.qsf
case-study-2
    analyses
        01-pre-processing.qmd
        02-analyses.qmd
        03-robustness-check.qmd
    data
        processed
            meme-feed-data.csv
            merged_data.csv
        raw
            DICE-processed-2024-10-15.csv
            DICE_Meme_Feed_October+15,+2024_18.35.csv
            PageTimes-2024-10-15.csv
            all_apps_wide_2024-10-15.csv
    img
        design.png
        stimuli.png
    report
        02-analyses.pdf
        03-robustness-check.pdf
    stimuli
        9gag.csv
        DICE_Meme_Feed.pdf
        DICE_Meme_Feed.qsf
```