

AGH

REPRODUCED



#103 Describing Subjective Experiment Consistency by *p*-Value P–P Plot

Jakub Nawata <jnawata@agh.edu.pl>, Lucjan Janowski, Bogdan Ćmiel, Krzysztof Rusek,
Marc A. Kastner, Jan Zahálka

VQEG F2F Meeting, June 7–11 2021, London, England

Background: Reproducibility

- » [The Reproducibility track](#)@ACM MM'21
- » Authors of papers from ACM MM'20 & '19 only
- » **Interactive & open review process**
- » The *Results Reproduced* badge

“The main results of the paper have been obtained in a subsequent study by a person or team other than the authors, using, in part, artifacts provided by the author.”

Background: Original Paper

Nawała, J., Janowski, L., Ćmiel, B., & Rusek, K. (2020).

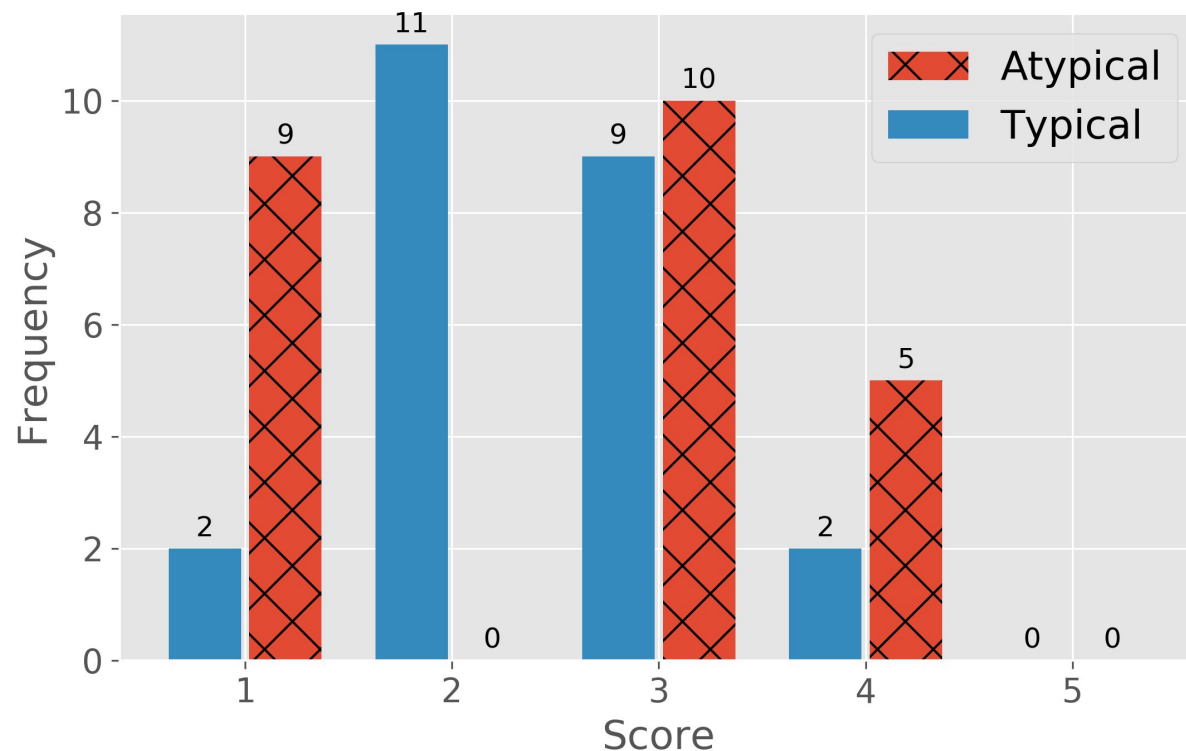
Describing Subjective Experiment Consistency by p -Value P-P Plot.

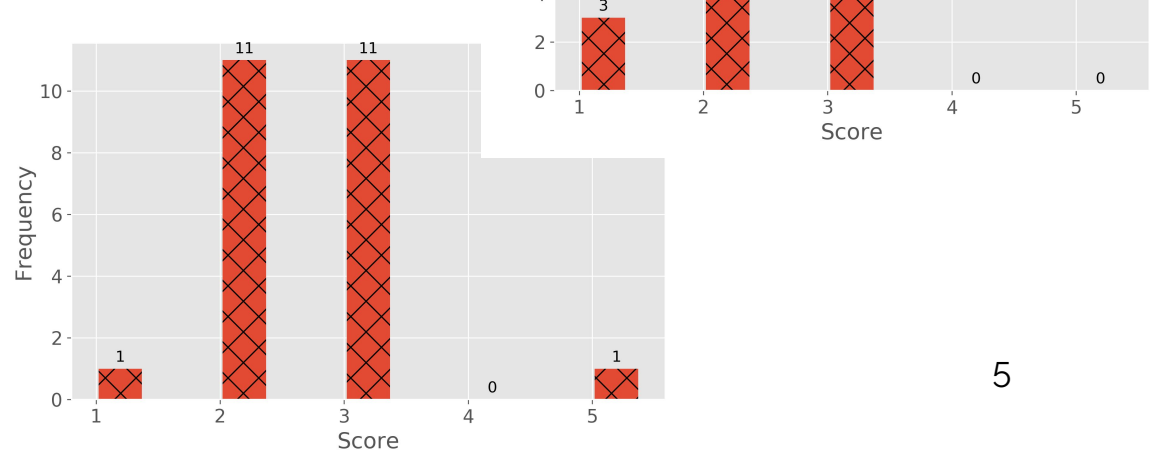
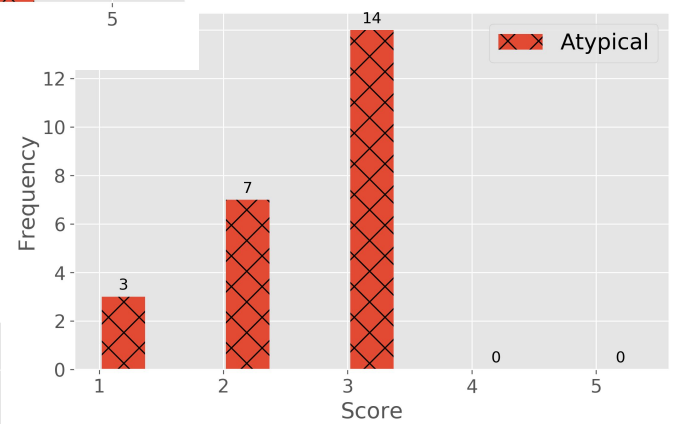
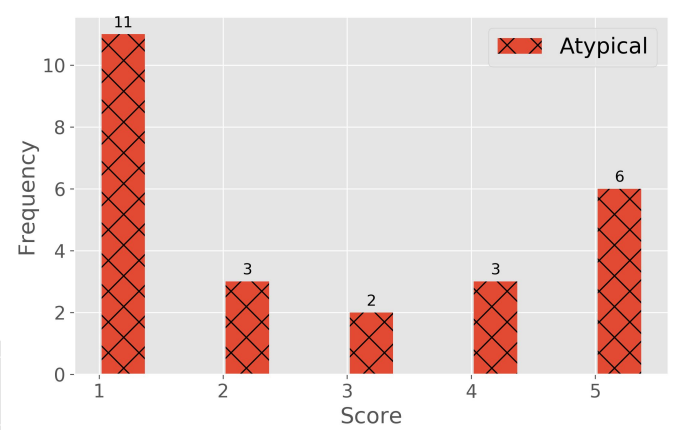
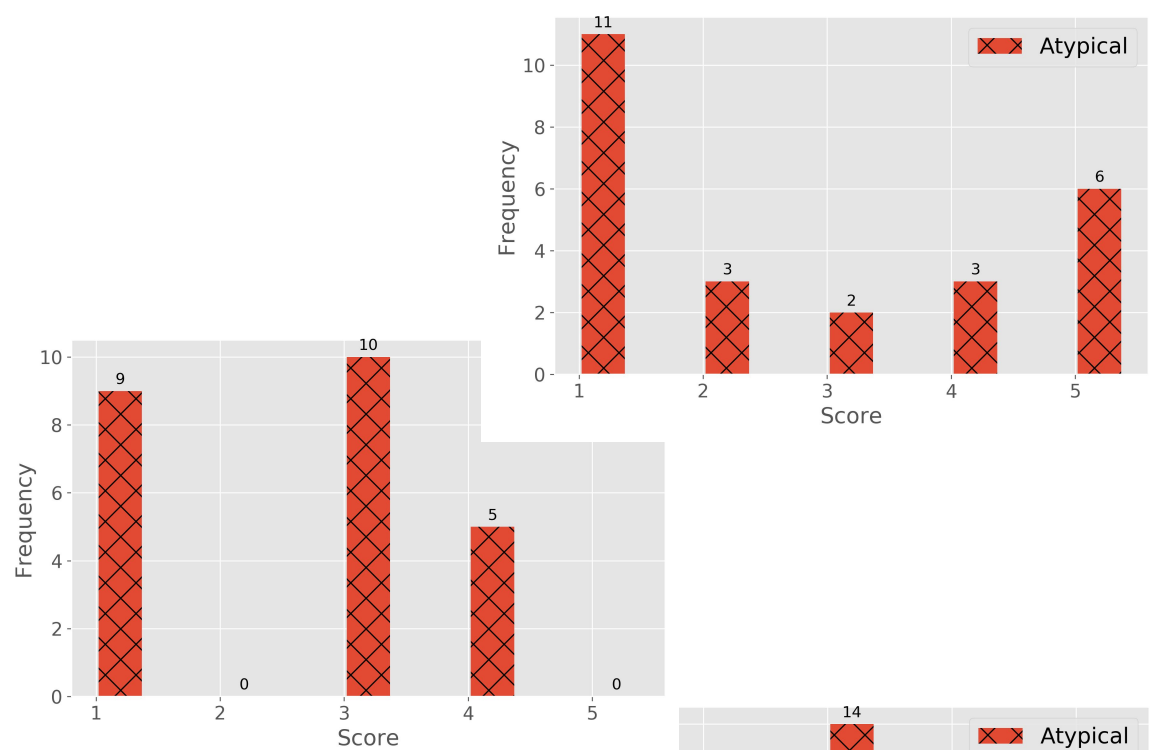
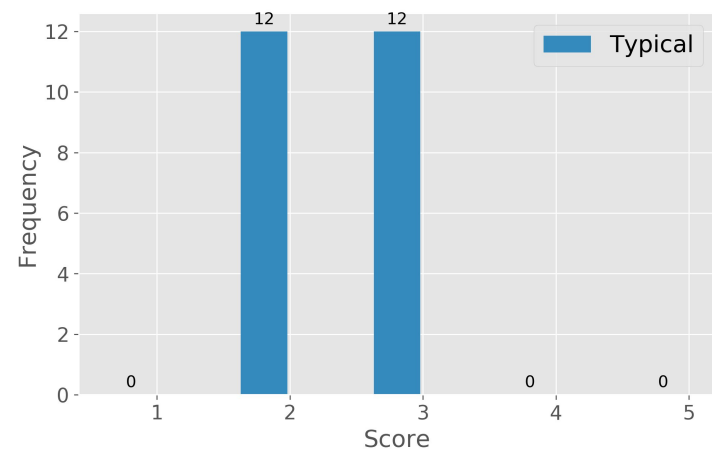
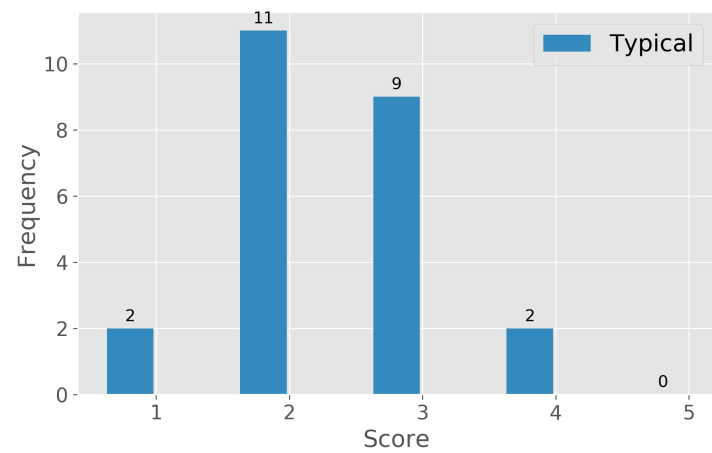
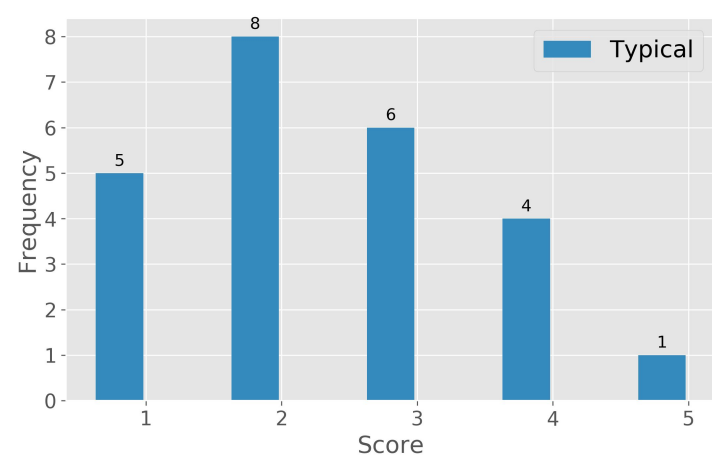
Proceedings of the 28th ACM International Conference on Multimedia, 852–861. <https://doi.org/10.1145/3394171.3413749>.
<https://arxiv.org/abs/2009.13372>.

- » A software tool assessing subjective experiment consistency
- » Pinpoints potentially problematic stimuli
- » Data & code openly available

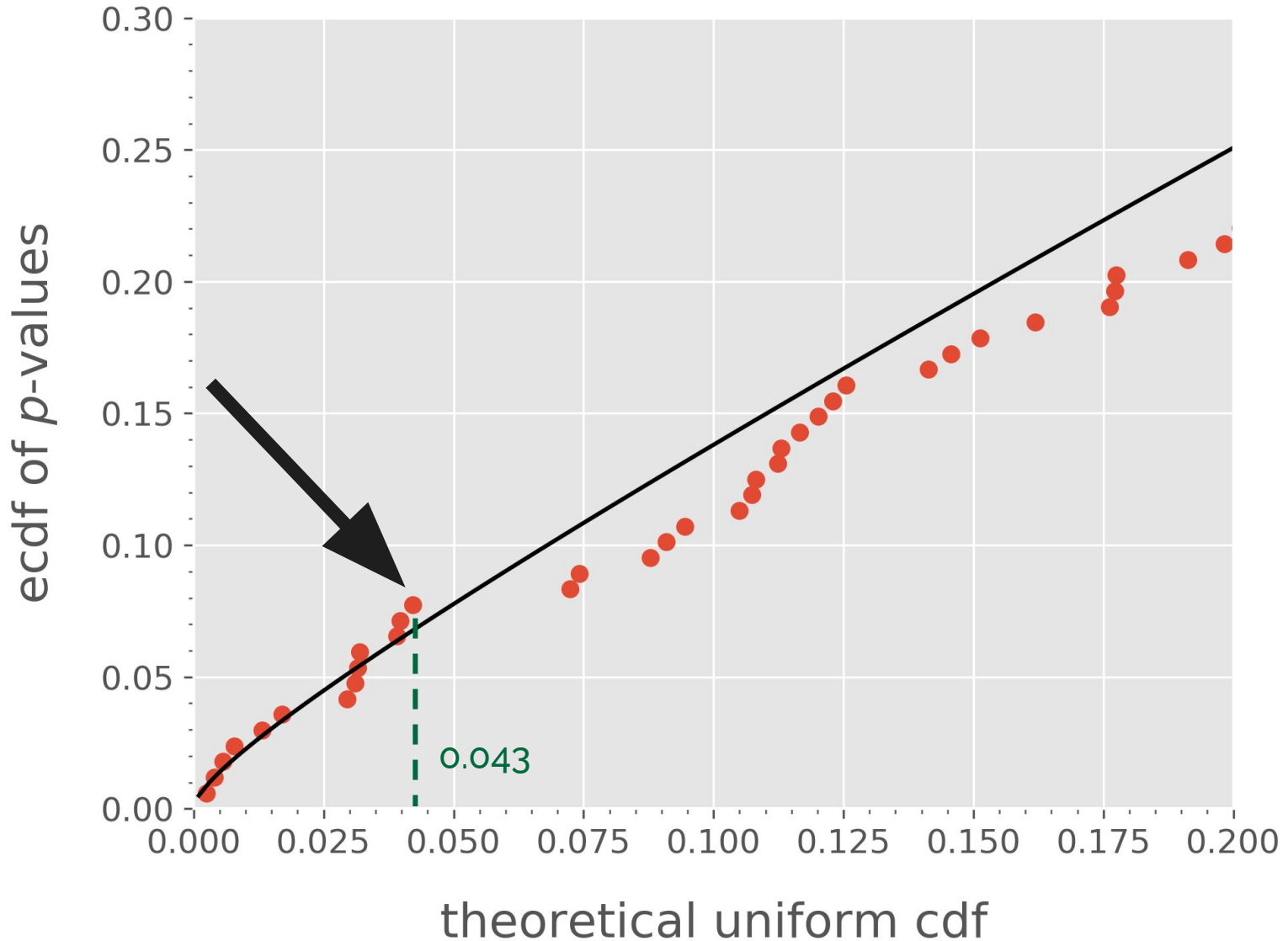
Background: Original Paper

- » Consistent = consisting of mostly typical stimuli.
- » Typical vs atypical
- » $GSD(\psi, \rho)$





Background: Original Paper



Background: Original Paper

- » Source code openly available
 - <https://github.com/Qub3k/subjective-exp-consistency-check>
- » Subjective responses also available (in the *tidy data* format)
 - six studies
 - 21 experiments
 - almost 100,000 responses
- » <https://grouplens.org/datasets/movielens/1m/>

Results: Successful Reproduction

- » What our software is for?
 - assessing consistency of a subjective experiment
 - pinpointing potentially problematic stimuli
 - complementing indications of consistency yielded by other methods*

Results: How to Run the Code?

```
$ python3 friendly_gsd.py  
  hdtv1_exp1_scores_pp_plot_ready.csv
```

- » Assumptions:
 - tidy data as input,
 - required Python packages and Python itself installed.
- » In case of doubts, refer to [the README.md file](#) on GitHub (or to [the new paper](#)).
- » Caveat: computations take a lot of time*

Extending Our Work

- » Franz Hahn (VQA Group at Universität Konstanz) has already contributed ❤️
- » You can create an issue on GitHub.
 - Propose new functionality.
 - Report bugs.
- » You can test the framework using your own model.

Thank You



- » Zhi Li.
- » Netflix, Inc.
- » PL-Grid Infrastructure.
- » Polish Ministry of Science and Higher Education.
- » The Norwegian Financial Mechanism 2014-2021 under project 2019/34/H/ST6/00599 (TUFIQoE).

Jakub Nawata

jnawata@agh.edu.pl

AGH University of Science and Technology
Department of Telecommunications