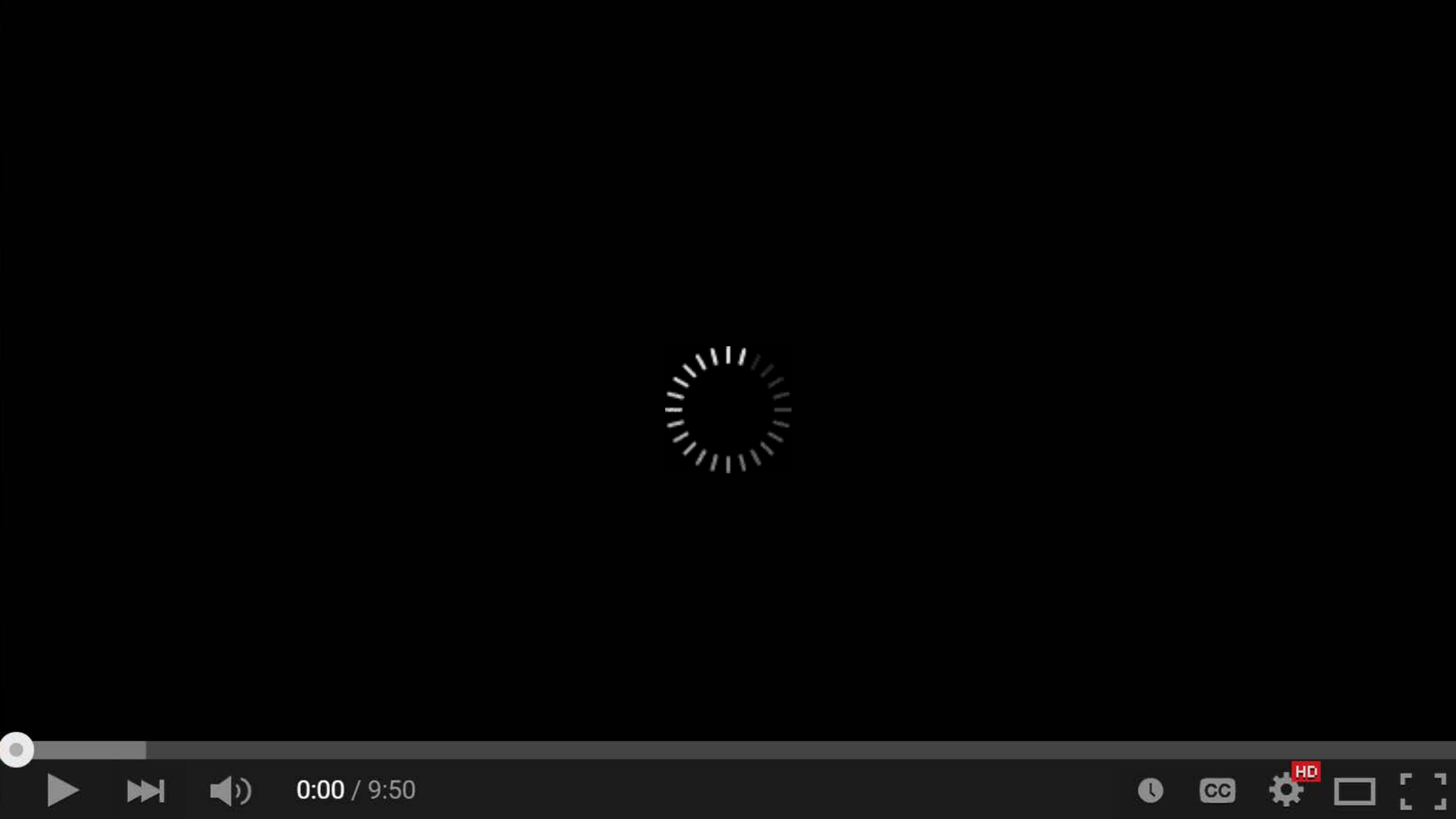
BEHAVIOR-ORIENTED QUALITY OF EXPERIENCE ASSESSMENT IN THE LAB AND AT HOME

Werner Robitza[1], Bernhard Feiten[1], Alexander Raake[2]

[1]Deutsche Telekom AG

[2]Audiovisual Technology Group, TU Ilmenau

VQEG Meeting — May 8–12, 2017 — Los Gatos, USA


0:00 / 9:50
HD

THE PROBLEM WITH CURRENT QUALITY TESTS

- Current audiovisual quality test methods:
  - Passive viewing/listening only
  - Stimulus + Rating, Stimulus + Rating, ...

- Is this real-life usage?
  - Users cannot select stimulus
  - No simulated session
    (e.g., player loading, website access)
  - No way to deal with extreme quality degradations
    other than giving “MOS 1”

Source: www.ecpnorthern.co.uk

MANY WAYS TO DO QUALITY TESTING

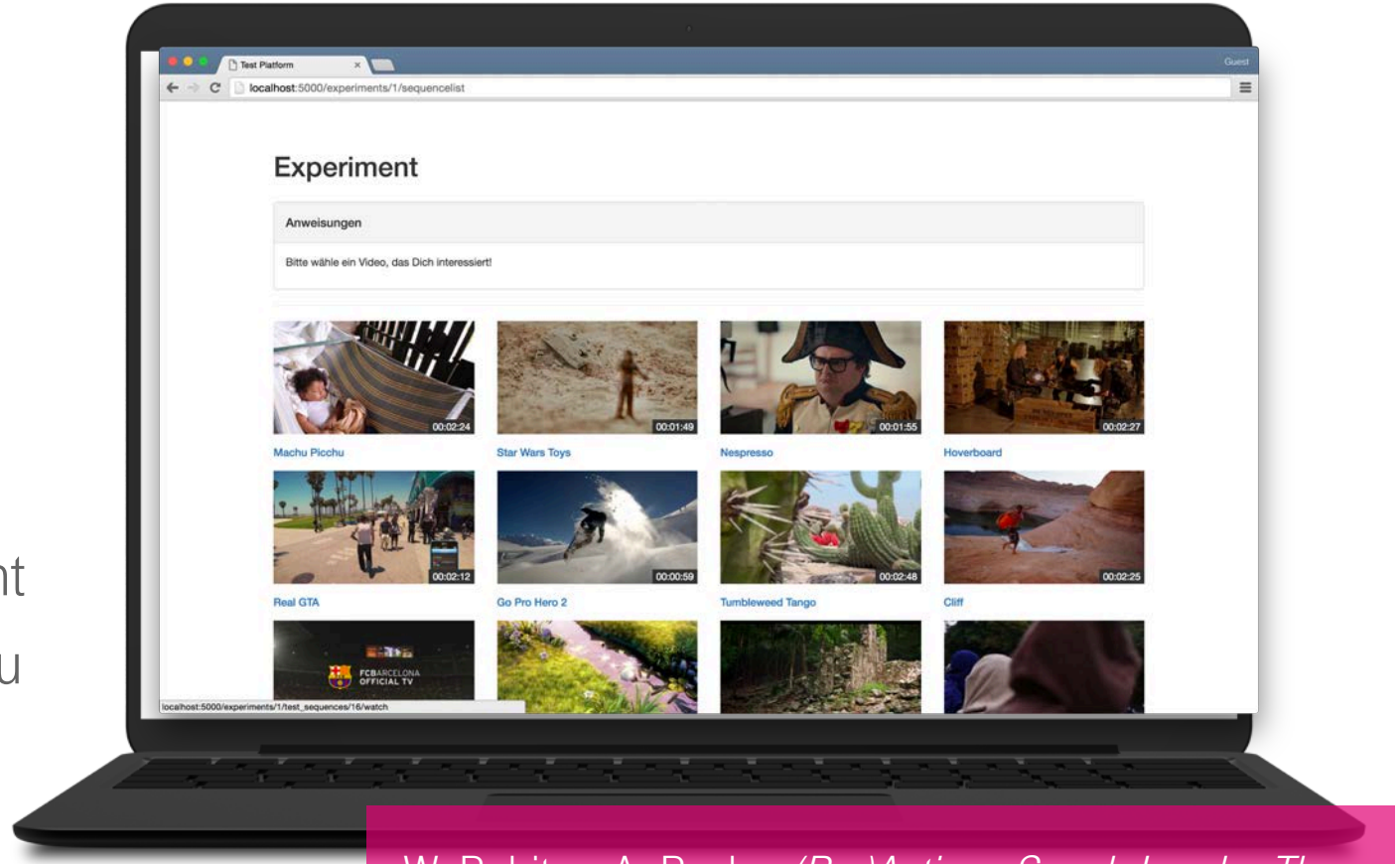
- Lab
- Home-like lab environment
- Friendly user studies
- In-service measurements

BEHAVIORAL QOE TESTS IN THE LAB

OUR TEST

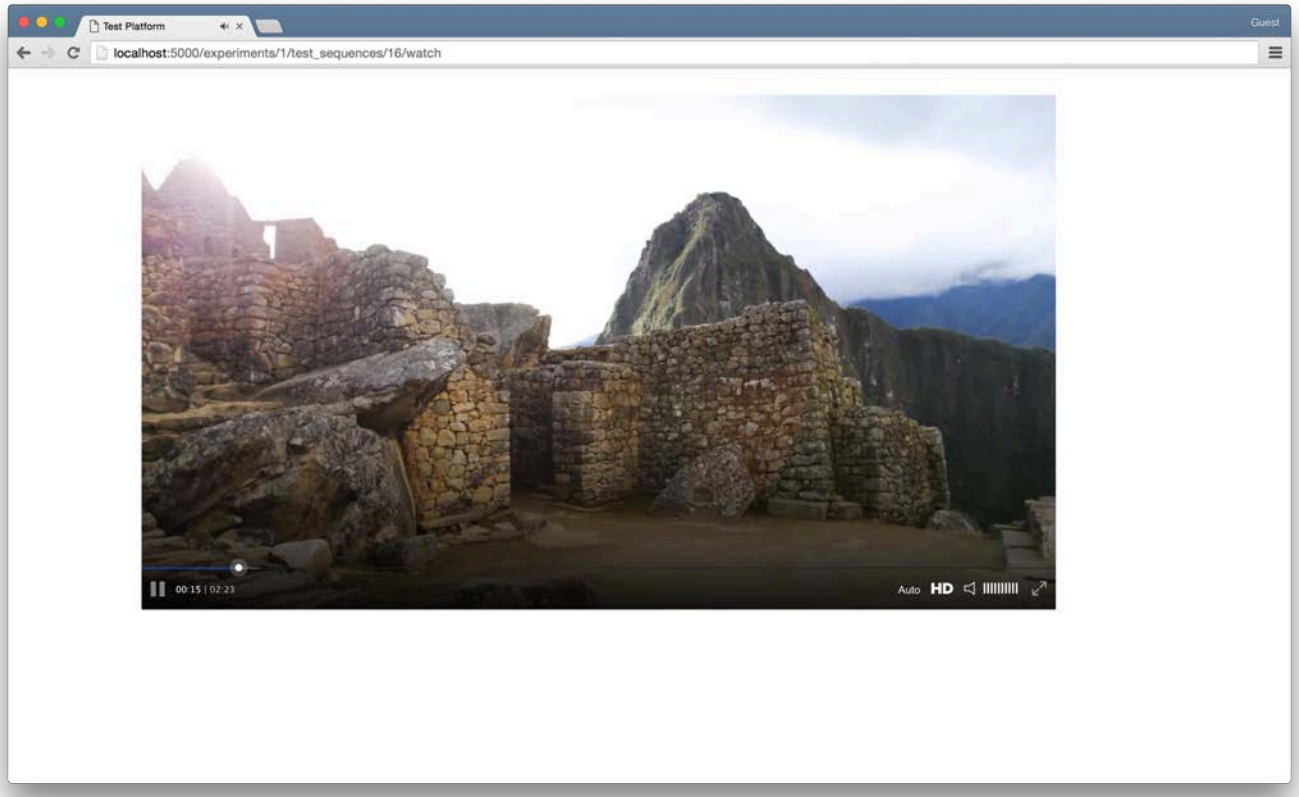
User task:

- Select a video
- Watch it entirely
- Describe the content
- Answer a question about the content
- Rate on a 5-star scale how much you liked the video
- Repeat until finished



W. Robitza, A. Raake, *(Re-)Actions Speak Louder Than Words? A Novel Test Method for Tracking User Behavior in Web Video Services*, QoMEX 2016.

PLAYER



- Dedicated video page for playback
- Player can be manipulated to insert stalling events / quality changes
- Users can seek, change volume, select quality level, enable fullscreen

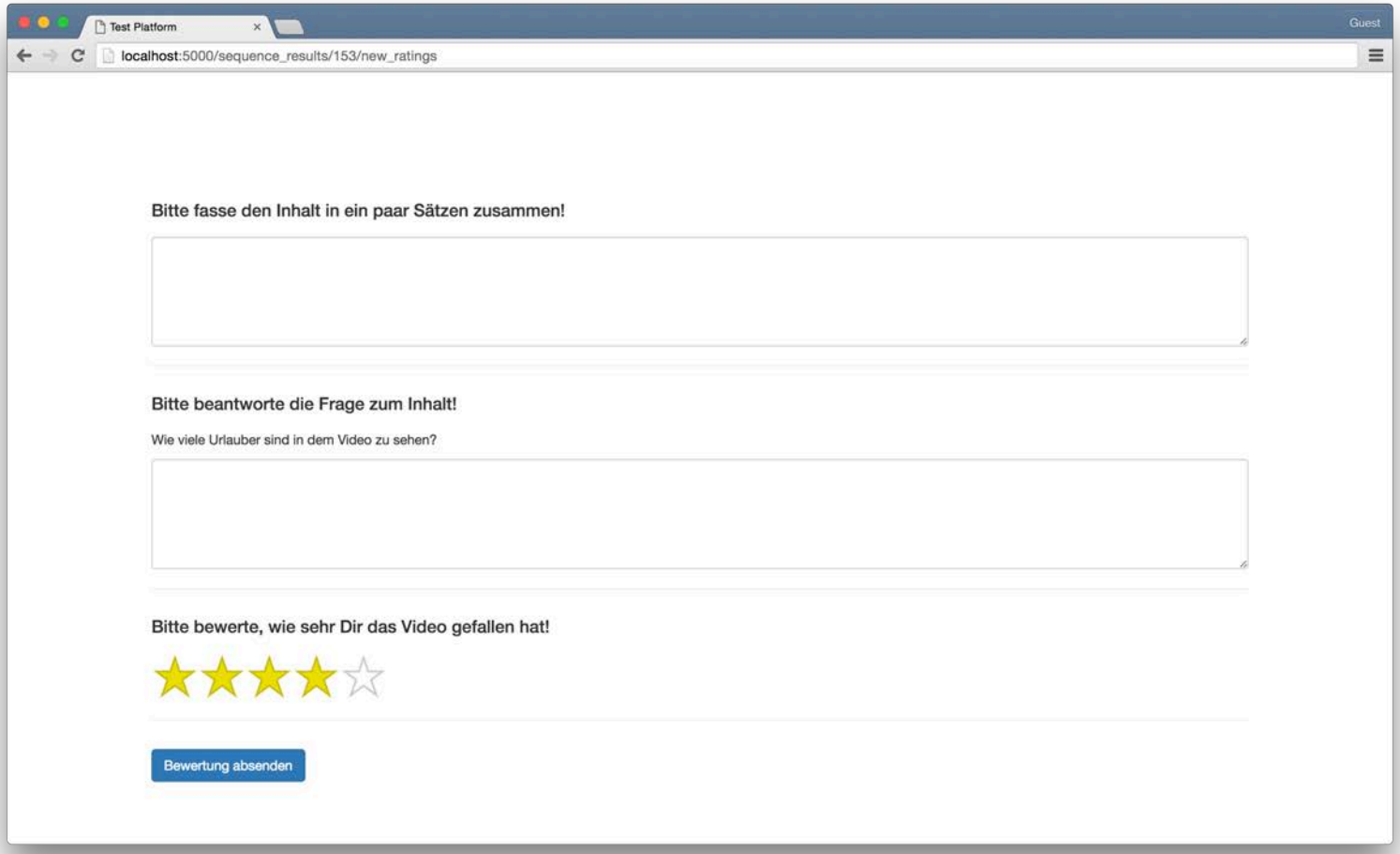
“QUALITY PROBLEM” CONDITIONS

Conditions

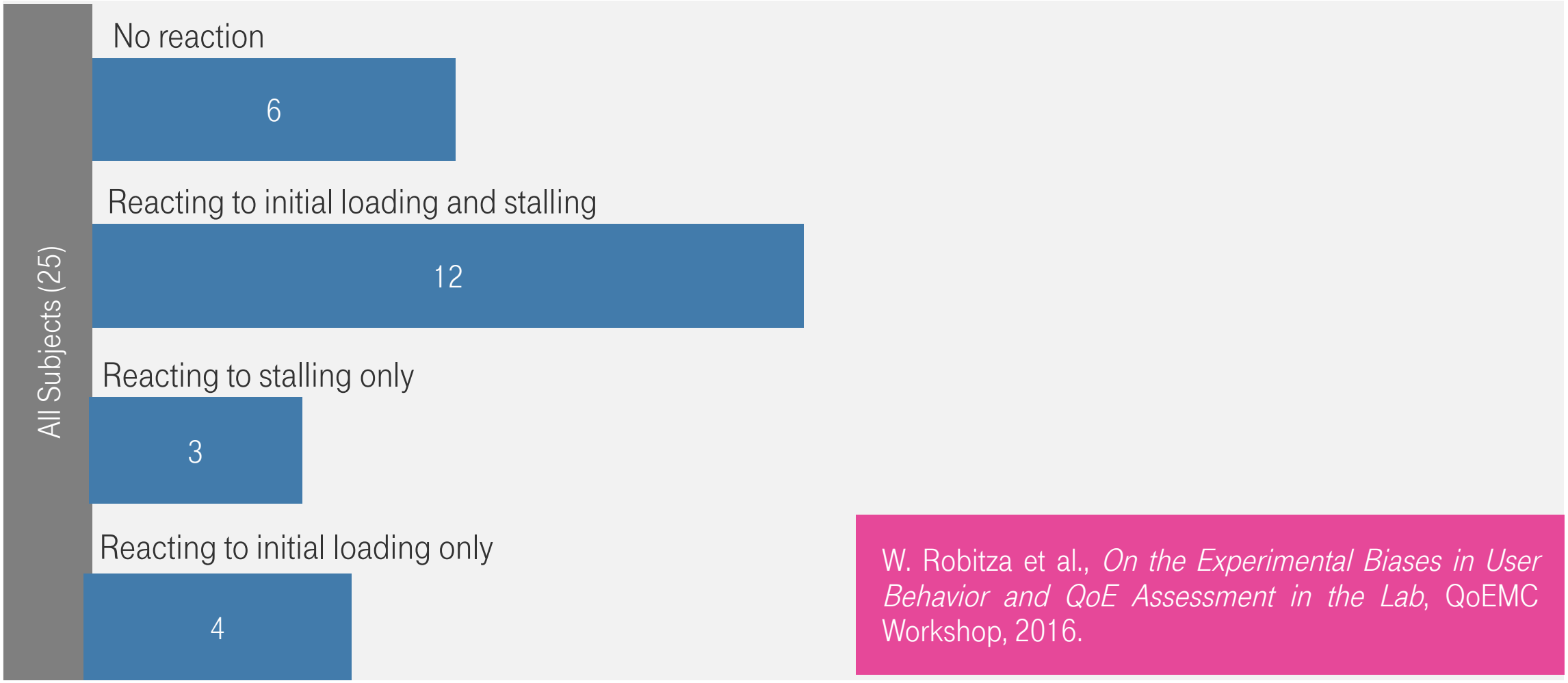
1. Reference
2. Initial loading (30 s)
3. Long stalling (30 s)
4. Quality drop (from highest to lowest)
5. Constant medium quality
6. Constant low quality

TEST INTERFACE – RATING

- Content summary
- Content-specific question
- 5-star rating of how much user likes the video



PRELIMINARY TEST RESULTS



W. Robitza et al., *On the Experimental Biases in User Behavior and QoE Assessment in the Lab*, QoEMC Workshop, 2016.

WHY DID SOME USERS NOT REACT IN THE LAB TEST?

IMPERCEPTIBILITY
Events were not perceived at all

LOW ANNOYANCE
Events not annoying enough

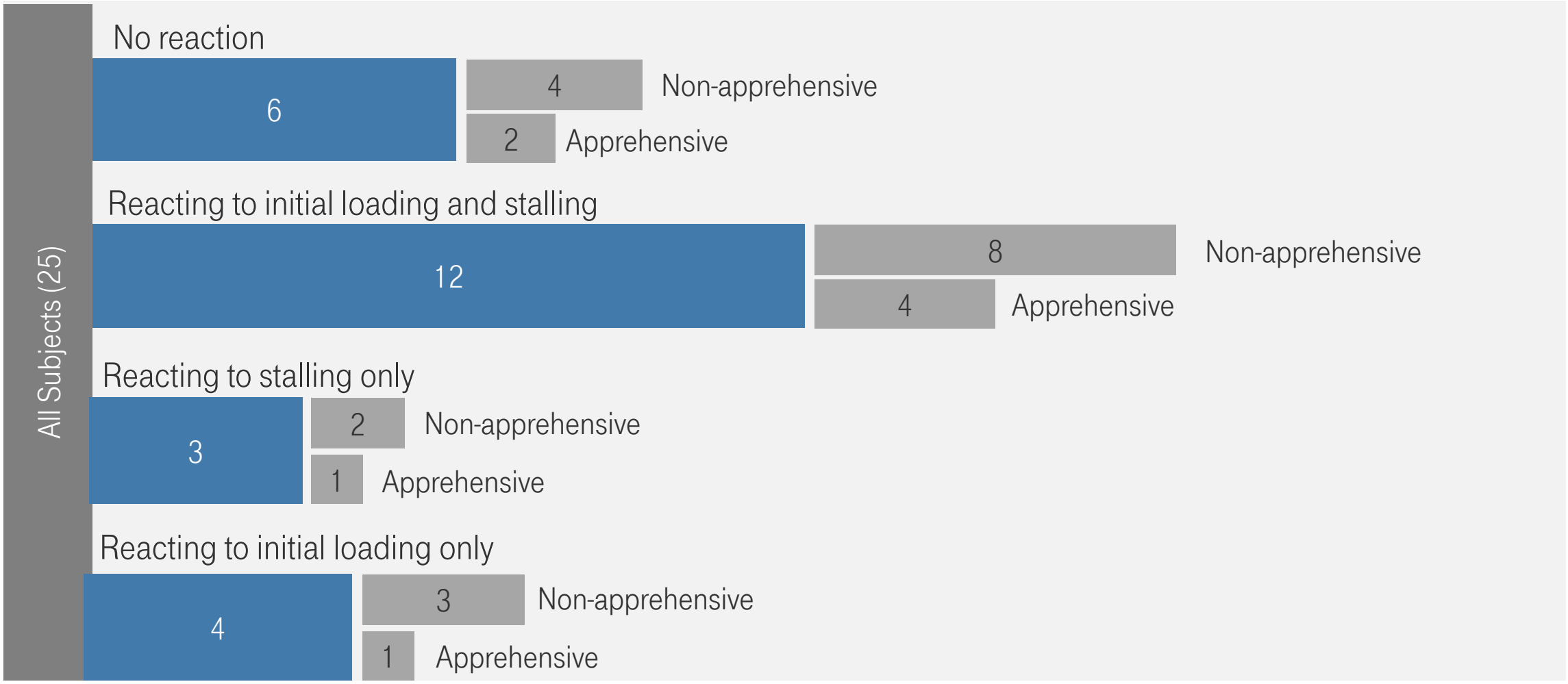
TASK DEPENDENCY
Events did not influence the actual task

TECHNICAL IGNORANCE
Users did not know what to do as a response

PROBLEM ATTRIBUTION
Thinking that the video is “already” bad, not the network

APPREHENSIVENESS
Not wanting to influence the test process

TEST RESULTS



APPREHENSIVENESS AND OTHER BIASES

"[...] the **subject is not a passive responder to stimuli and experimental conditions**. Instead, he is an active participant in a special form of socially defined interaction which we call 'taking part in an experiment.'"

Martin Orne, 1969

W. Robitza et al, *A Theoretical Approach to the Formation of Quality of Experience and User Behavior in Multimedia Services*, PQS Workshop, 2016.

A BEHAVIORAL QOE DILEMMA

“Classic” engineering approach:

- Tell people the experiment purpose
- Give a list of possible and valid answers or interactions

- ✓ Many conditions can be tested
- ✓ Controlled outcome
- ! High amount of experimental biases
- ! Low (ecological) validity of results

Open behavioral approach:

- Give users a fake task
- Leave reactions open

- ✓ Lower influence of expectations
- ✓ Higher ecological validity
- ! Low number of data points
- ! Subjects generate their own hypotheses

FUTURE WORK

- Try variations of experimental paradigm
  - Different tasks (e.g., “find your favorite music video”)
  - Different platform (e.g., use YouTube itself)
- Standardization efforts for methods that include user behavior
- User behavior-based quality prediction models
  (that still translate to MOS, or abandonment ratio)
- Draft new recommendation P.QUIT in ITU-T Study Group 12

BEHAVIORAL QOE WITH CROWDSOURCING

LIFE IS FOR SHARING.

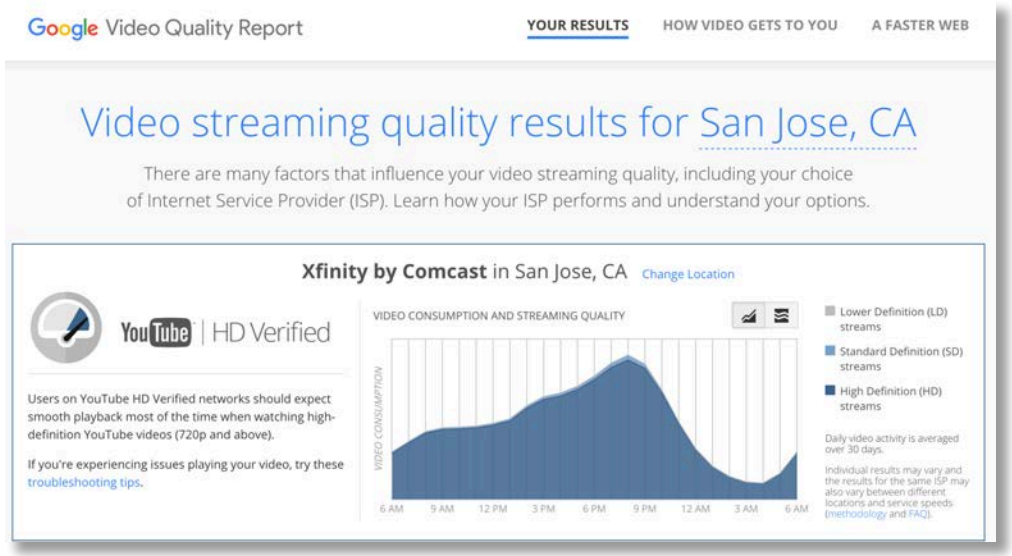
YT CrowdMon

https://ytcrowdmon.de

- Browser extension for Google Chrome (Firefox support coming soon)
- Measures
  - video streaming KPIs
  - background network statistics
  - interactions with video player
  - interactions with web page

MEANINGFUL QOE STATISTICS

- Goal: Implement meaningful statistics for users
  - Typical loading times
  - QoE over time
  - Compare specific locations, ISPs, ...
- Methods:
  - Use new **ITU-T P.1203** standard for estimating QoE of a video streaming session



Google Video Quality Report

Netflix ISP Leaderboard

FUTURE WORK

- Use YTCrowdMon as crowdsourcing platform, giving users specific tasks
- Conduct friendly-user trials over longer periods of time
- Compare results obtained from real streaming sessions with lab results on similar sessions

THANK YOU!

This presentation is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 643072, Network "QoE-Net".

QoENet
MSCA-ITN Network

European Commission