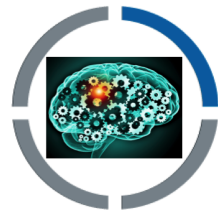


SpeedPerception: Phase-1 Update Results Overview

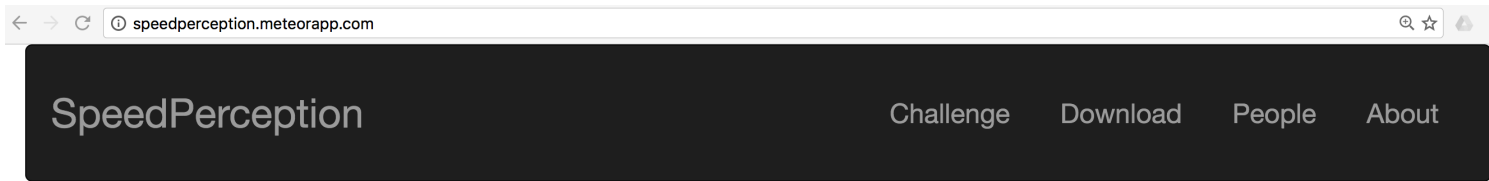
Parvez Ahammad
Head of Data Science and Machine Learning Group
Instart Logic Inc.

October 2016

What is SpeedPerception?



“SpeedPerception is a large-scale web performance crowdsourcing study focused on the perceived loading performance of above-the-fold content.”

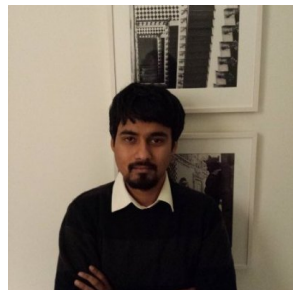


What is SpeedPerception?

Clearly, no one likes slow loading webpages. SpeedPerception is a study trying to understand what “slow” and “fast” mean to the human end-user. You can help improve webpage speed by taking part in the SpeedPerception Challenge. **Click on the [Challenge](#) to play!**

SpeedPerception: Team

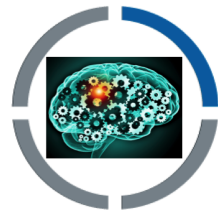
- Clark Gao
- Parvez Ahammad (@perceptPA)
- Prasenjit Dey



Collaborators:

- Estelle Weyl (@estellewv)
- Pat Meenan (@patmeenana)



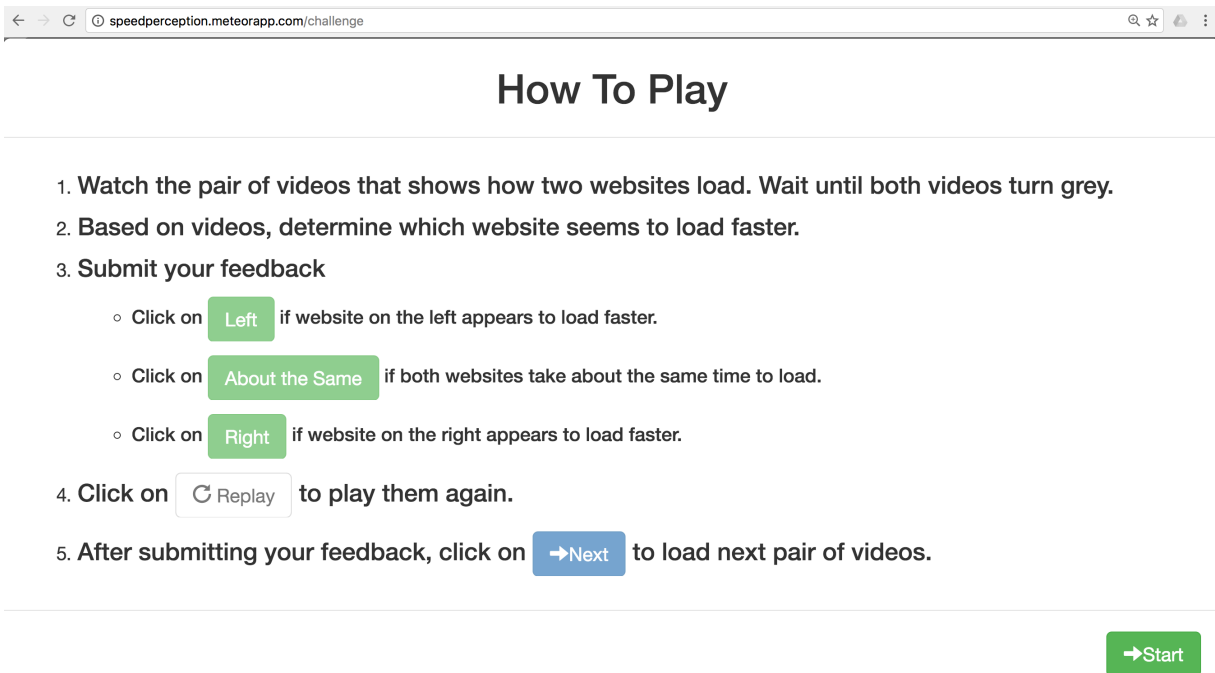


Why did we do SpeedPerception?

- Fill a critical gap for A/B comparisons of human end-user web UX
- Facilitate shift from “*anecData*™” to real human Ground Truth on webperf UX
- Provide a large-scale quantitative benchmark
 - Following MNIST / ImageNet / ActivityNet tradition
- Facilitate modeling / machine learning efforts in future
- Consistent dataset for algorithmic comparisons
- Open source framework / benchmark → reproducible results

SpeedPerception Challenge: Design (1/2)

- Premise: perception of above-the-fold performance (like perceived webpage speed) is *relative*. <http://speedperception.meteorapp.com/challenge>



The screenshot shows a web browser window with the address bar displaying "speedperception.meteorapp.com/challenge". The main heading is "How To Play". Below it, there are five numbered instructions for the challenge. Instructions 1, 2, and 3 are followed by three radio button options: "Left", "About the Same", and "Right". Instruction 4 is followed by a "Replay" button. Instruction 5 is followed by a "Next" button. At the bottom right of the page is a green "Start" button.

How To Play

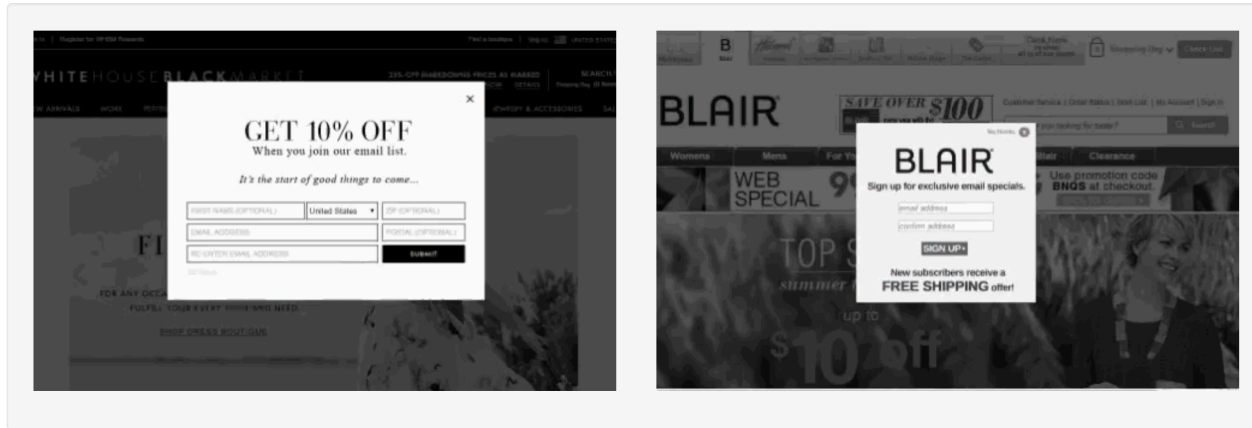
1. Watch the pair of videos that shows how two websites load. Wait until both videos turn grey.
2. Based on videos, determine which website seems to load faster.
3. Submit your feedback
 - Click on **Left** if website on the left appears to load faster.
 - Click on **About the Same** if both websites take about the same time to load.
 - Click on **Right** if website on the right appears to load faster.
4. Click on **Replay** to play them again.
5. After submitting your feedback, click on **Next** to load next pair of videos.

Start

SpeedPerception Challenge: Design (2/2)

- Premise: perception of above-the-fold performance (like perceived webpage speed) is *relative*. <http://speedperception.meteorapp.com/challenge>

Which of the 2 websites do you perceive to be faster?



Left

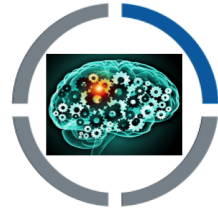
About the Same

Right

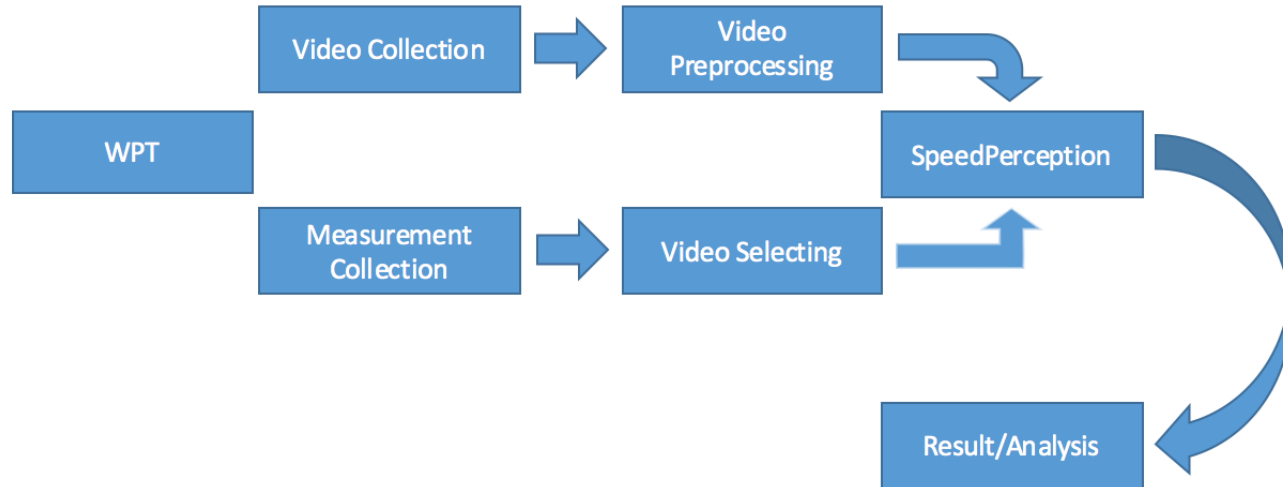
↶ Replay

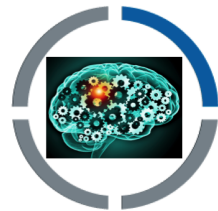
→ Next

SpeedPerception Challenge: Phase-1 Data (1/2)



- Source: Internet Retailer top-500 URLs
- WebPagetest (private instance) to collect HAR / metrics / videos (June 2016)
- Chrome browser / Cable connection speed / Desktop rendering mode
- All steps publicly available: <https://github.com/pdey/SpeedPerceptionApp>



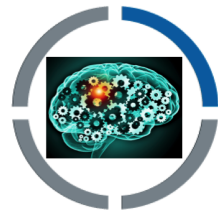


SpeedPerception Challenge: Phase-1 Data (2/2)

- After applying publicized rules for pair selection:
 - 115 URLs
 - 160 A/B pairs
- Each session: 16 test pairs + 5 honeypot* pairs = 21 pairs
- Record voluntary time to click for each pair (not publicized)

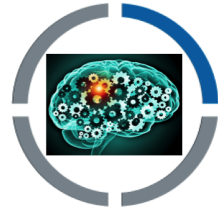
Honeypot* video pair = A video pair with known (very obvious) choice – used as a way to evaluate user input quality

SpeedPerception Challenge: Working hypotheses



- **H1:** No single metric can explain human choices with 90%+ accuracy
- **H2:** Visual metrics will perform better than non-visual/network metrics
- **H3:** User will not wait until “Visual Complete” to make their choice (despite the explicit instruction to wait until video turns grey)

SpeedPerception Challenge

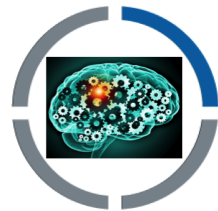


- Went public on 28th July 2016

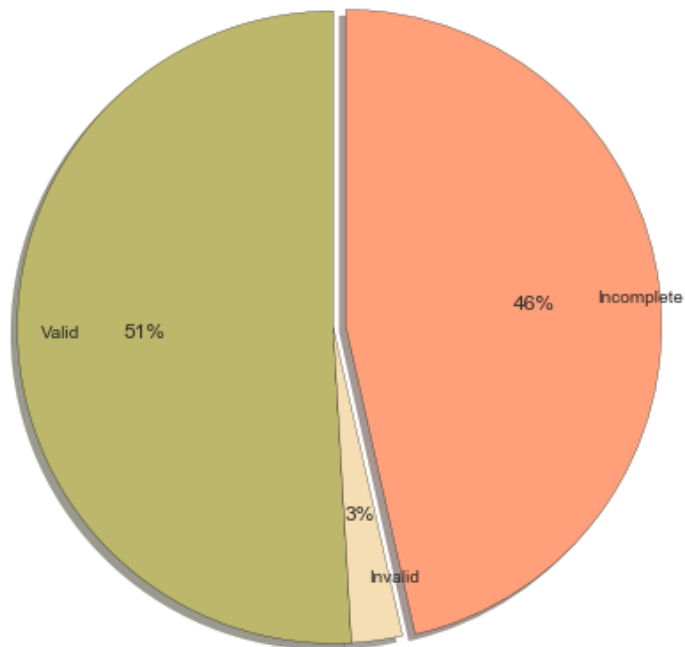


- Phase-1 ended on 30th September 2016
- Benchmark and findings available at:
 - <http://SpeedPerception.org>

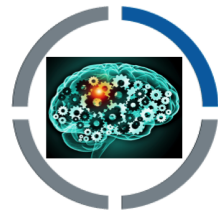
SpeedPerception Benchmark / DeepDive



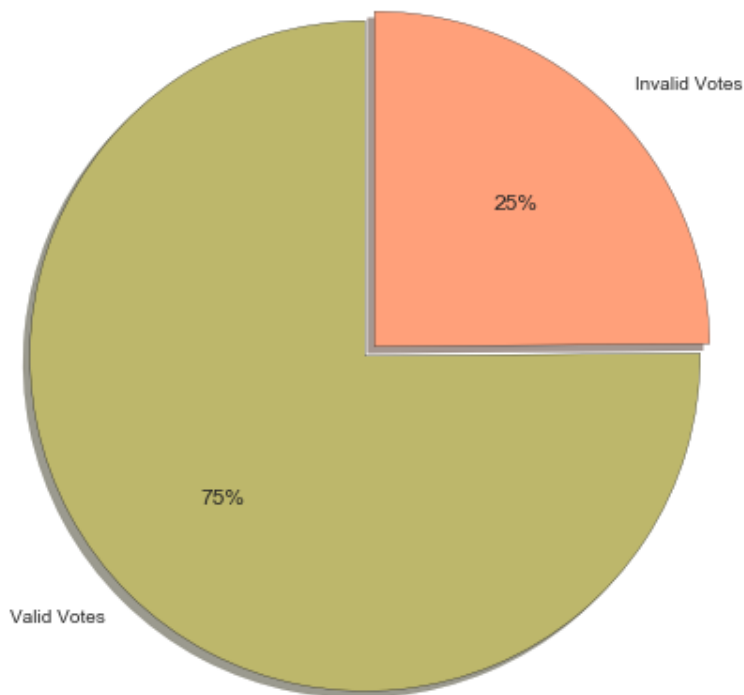
SpeedPerception Data of 5444 sessions



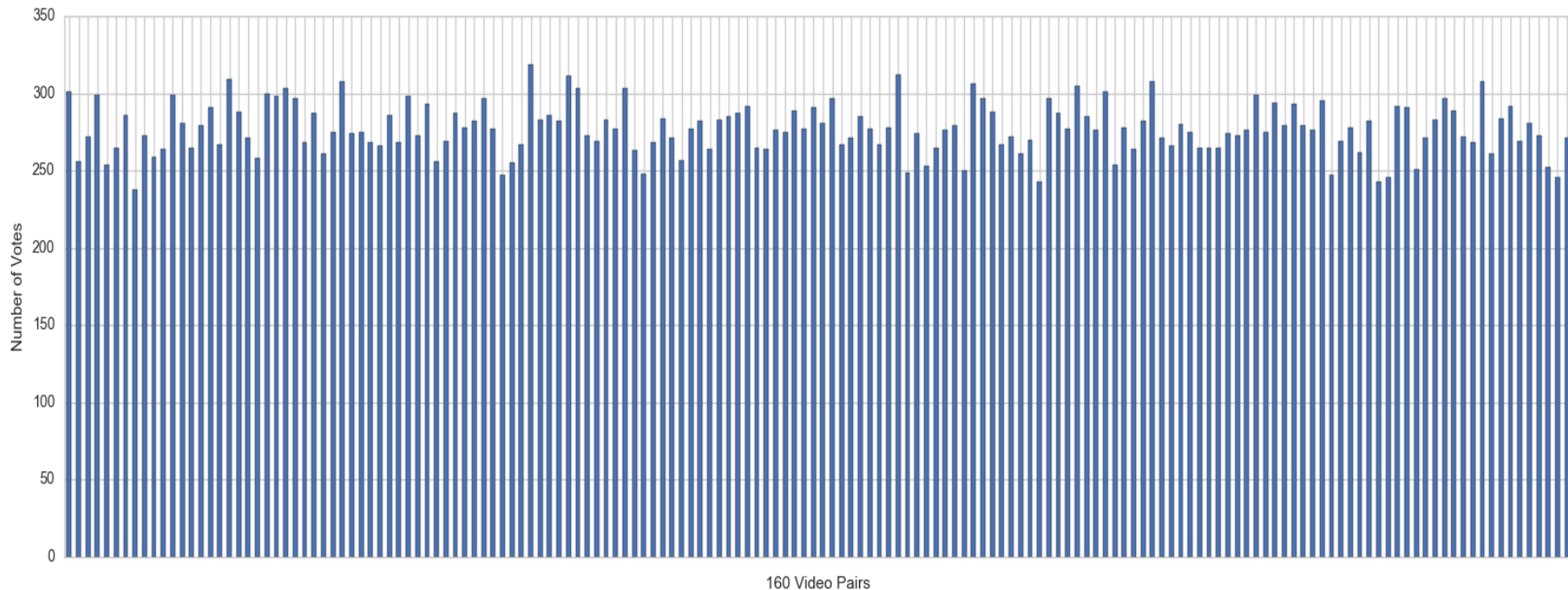
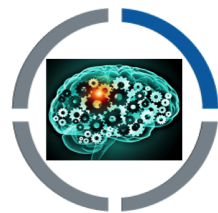
SpeedPerception Benchmark / DeepDive

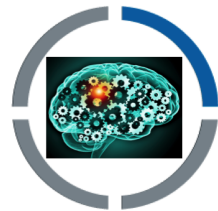


SpeedPerception Data of 77482 votes



SpeedPerception Benchmark / DeepDive



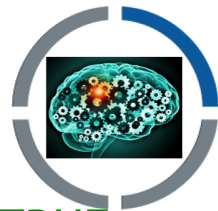


SpeedPerception Benchmark / User Feedback

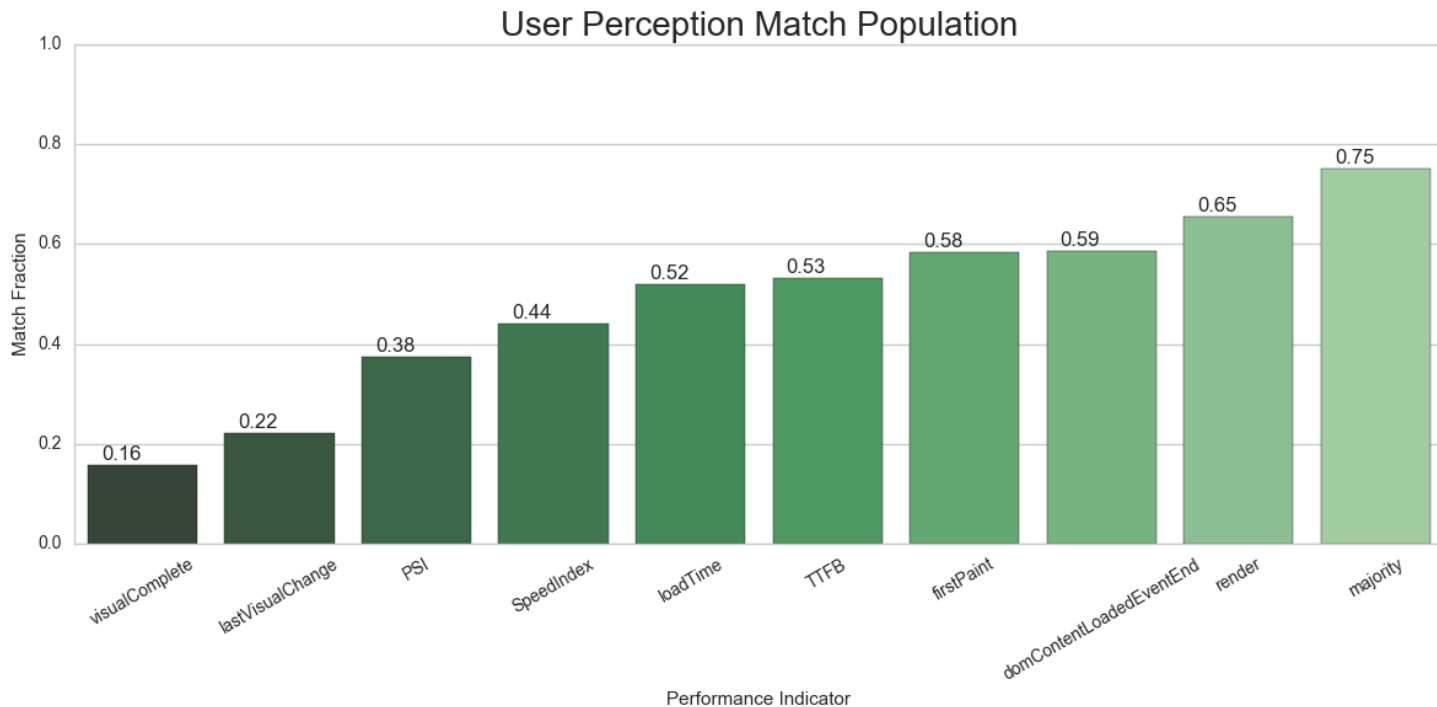
- Perception of speed and UX strongly impacted by popups / overlays



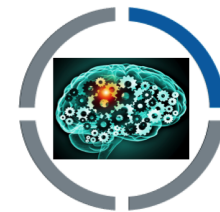
SpeedPerception Benchmark / Hypothesis # 1



- **H1:** No single metric can explain human choices with 90%+ accuracy - **TRUE**



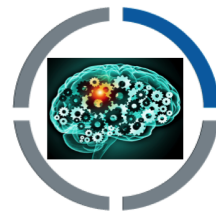
SpeedPerception Benchmark / Hypothesis # 1



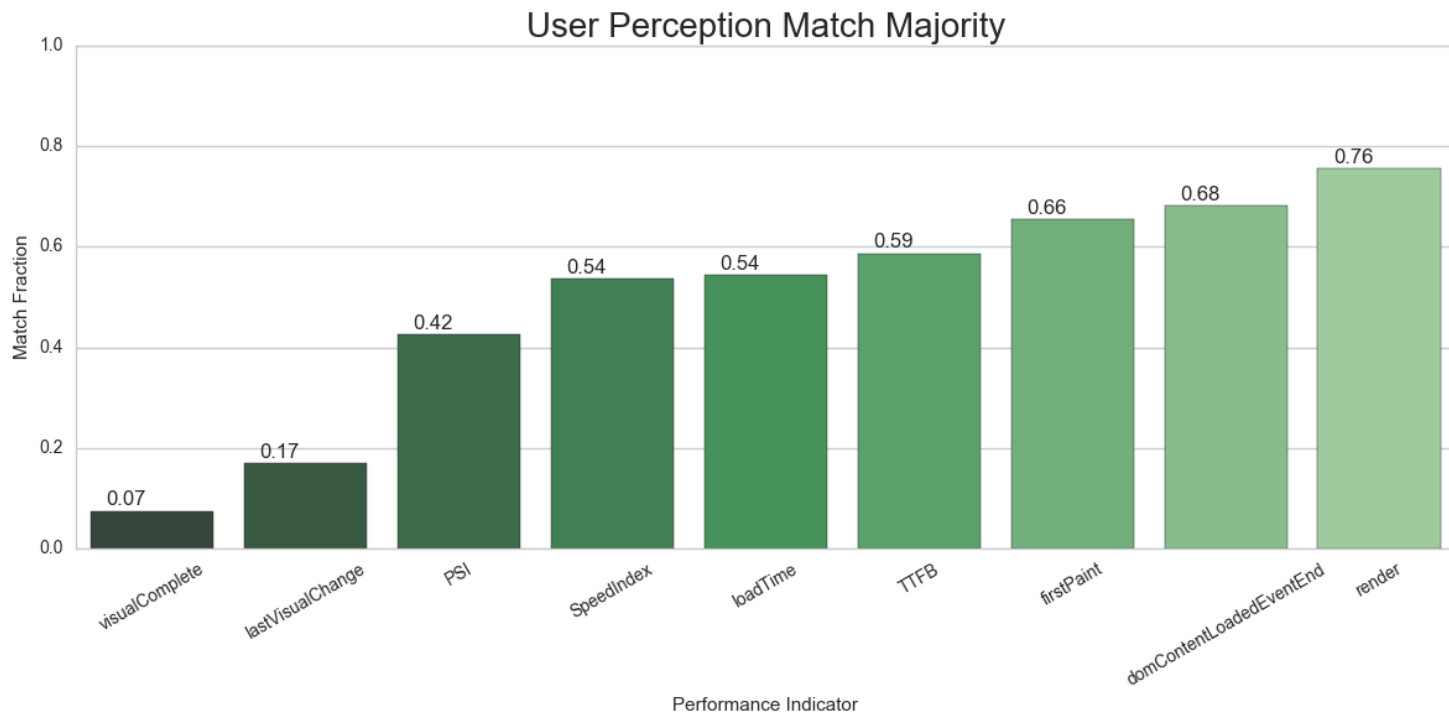
- **H1:** No single metric can explain human choices with 90%+ accuracy - **TRUE**



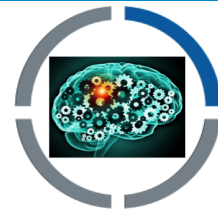
SpeedPerception Benchmark / Hypothesis # 1



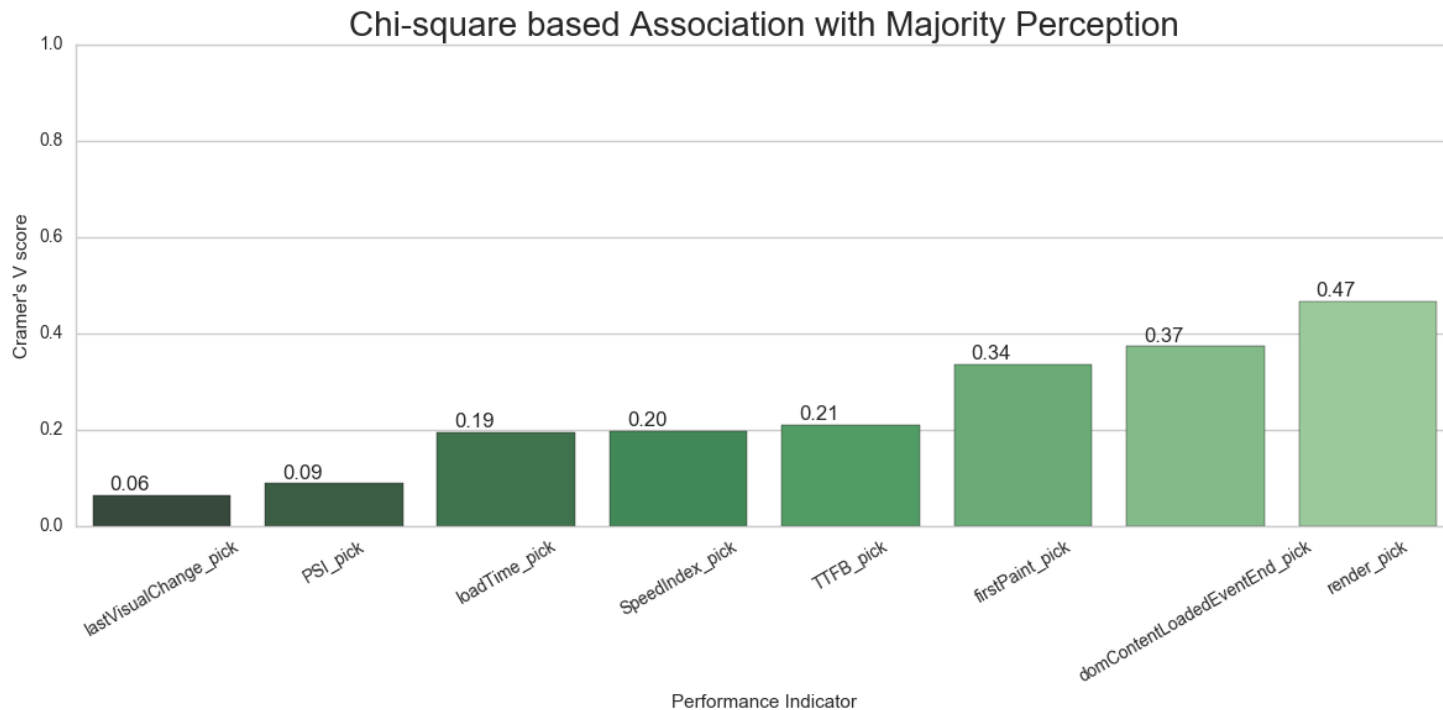
- **H1:** No single metric can explain human choices with 90%+ accuracy - **TRUE**



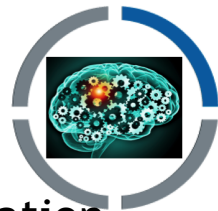
SpeedPerception Benchmark / Hypothesis # 1



- **H1:** No single metric can explain human choices with 90%+ accuracy - **TRUE**

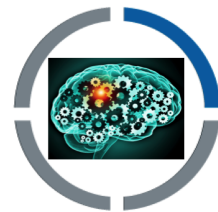


Questions to consider

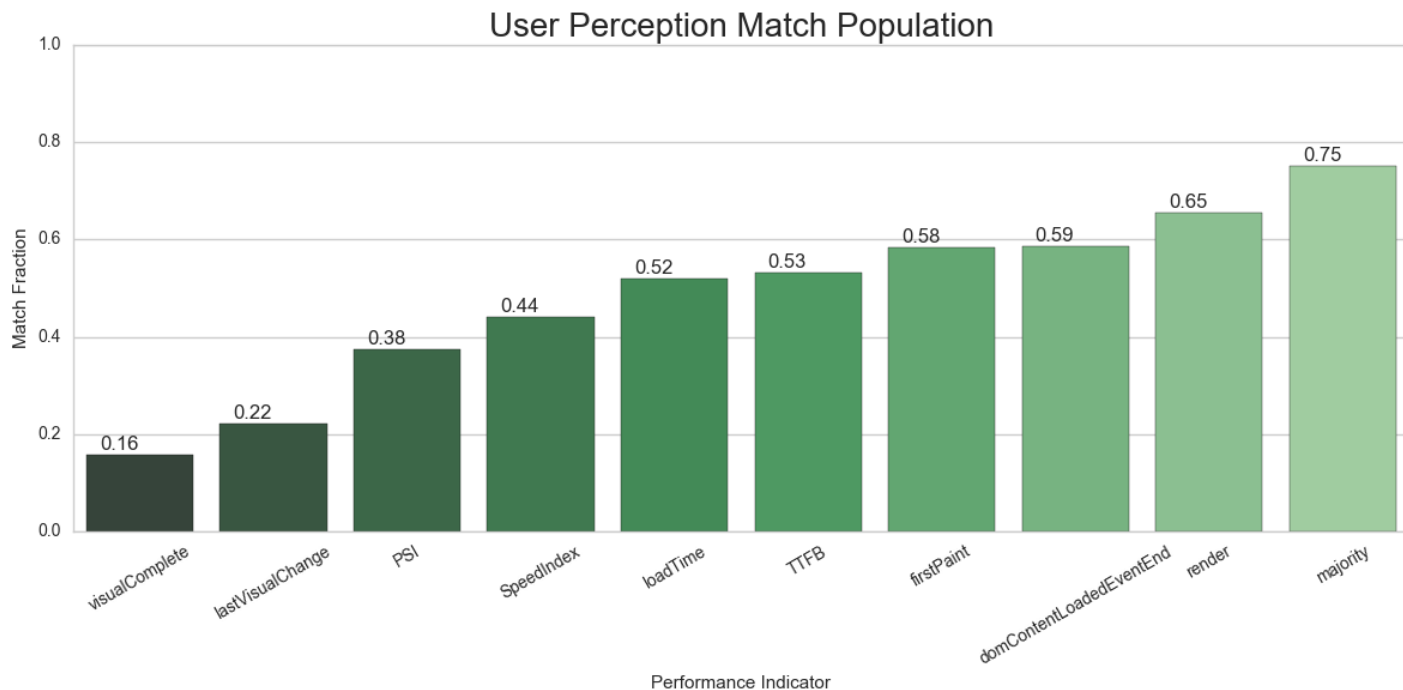


- Yes, there appears to be no one unicorn metric but, is there a combination synthetic metric (joint ML model) that will do a better job?
- People only looked two videos and made the call. Is there some additional information that we can extract from videos that will improve our models?

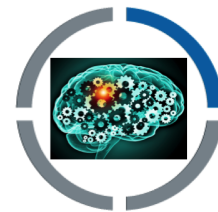
SpeedPerception Benchmark / Hypothesis # 2



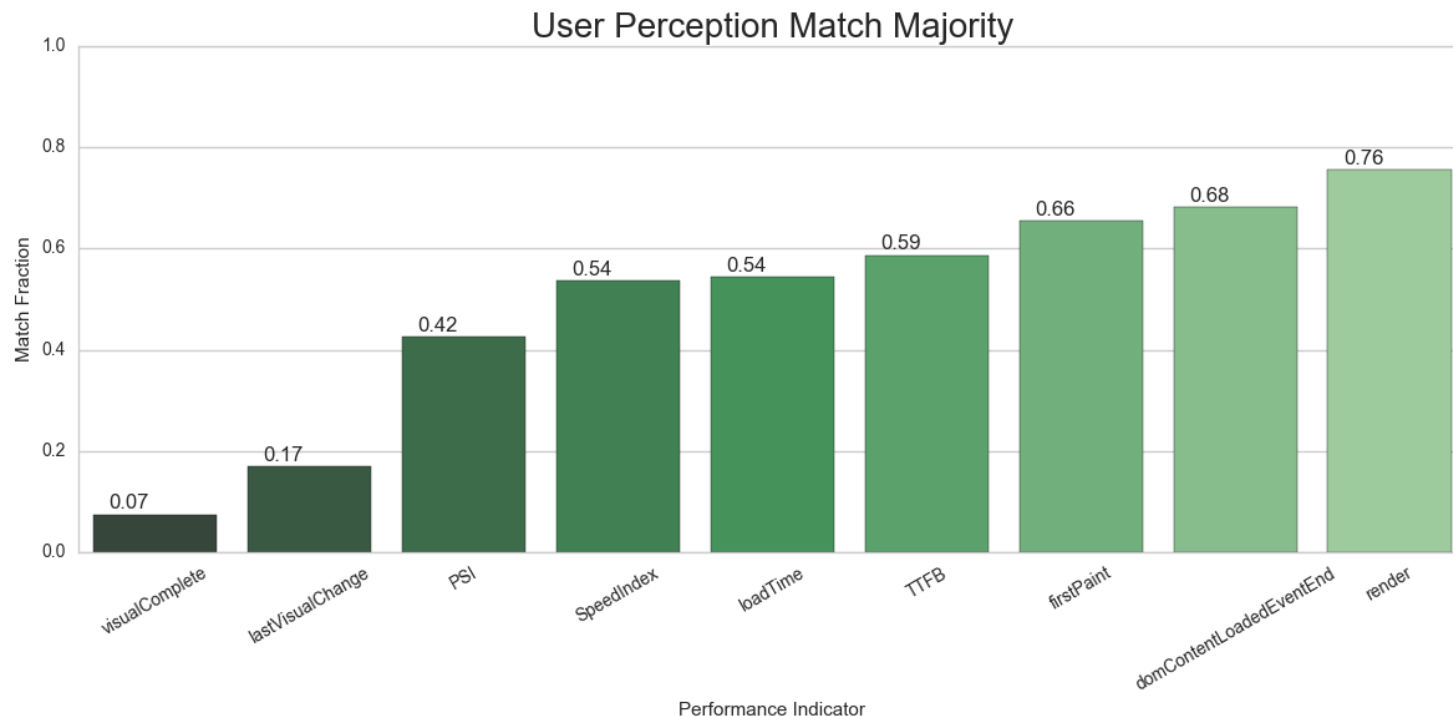
- **H2:** Visual metrics will perform better than non-visual metrics - **NOT TRUE**



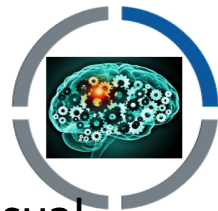
SpeedPerception Benchmark / Hypothesis # 2



- **H2:** Visual metrics will perform better than non-visual metrics - **NOT TRUE**

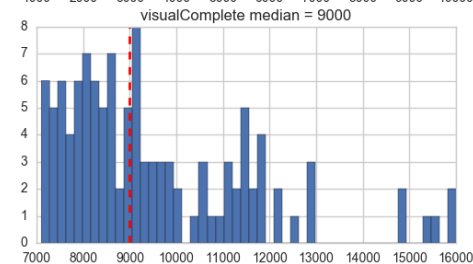
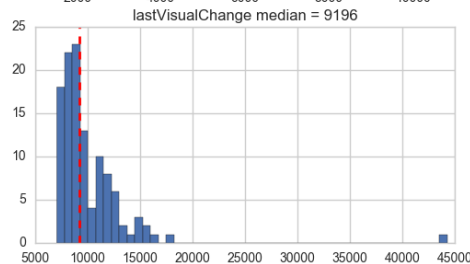
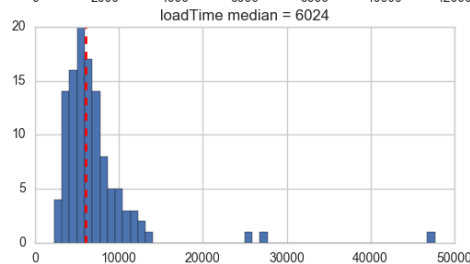
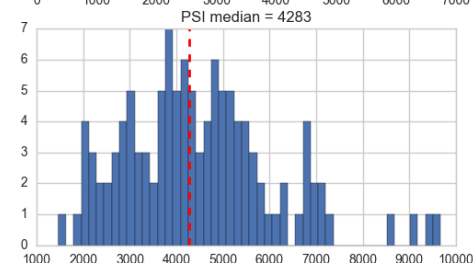
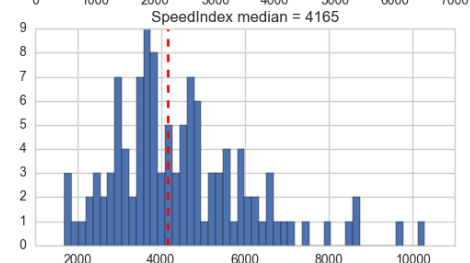
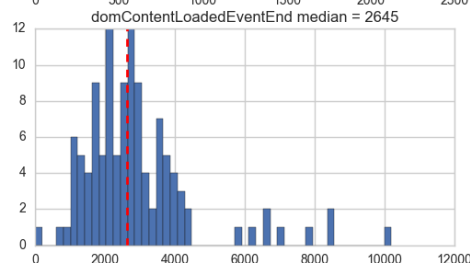
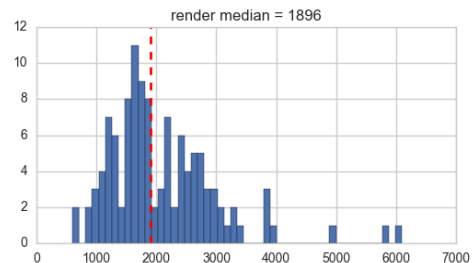
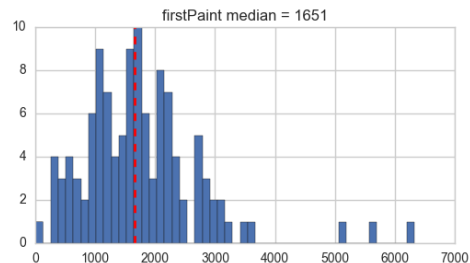
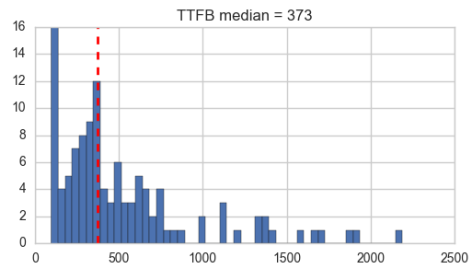
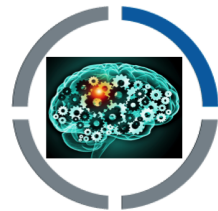


Questions to consider

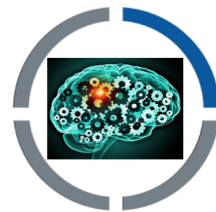


- May be the presence of visual jitter / interstitials hurt the ability of visual metrics to perform well. How can we improve them?
- Will there be different trends for video pairs that are free of visual jitter like modals and overlays?
- Is it possible to automatically predict the presence of visual jitter in websites (or website loading videos) to help choose a better set of metrics?

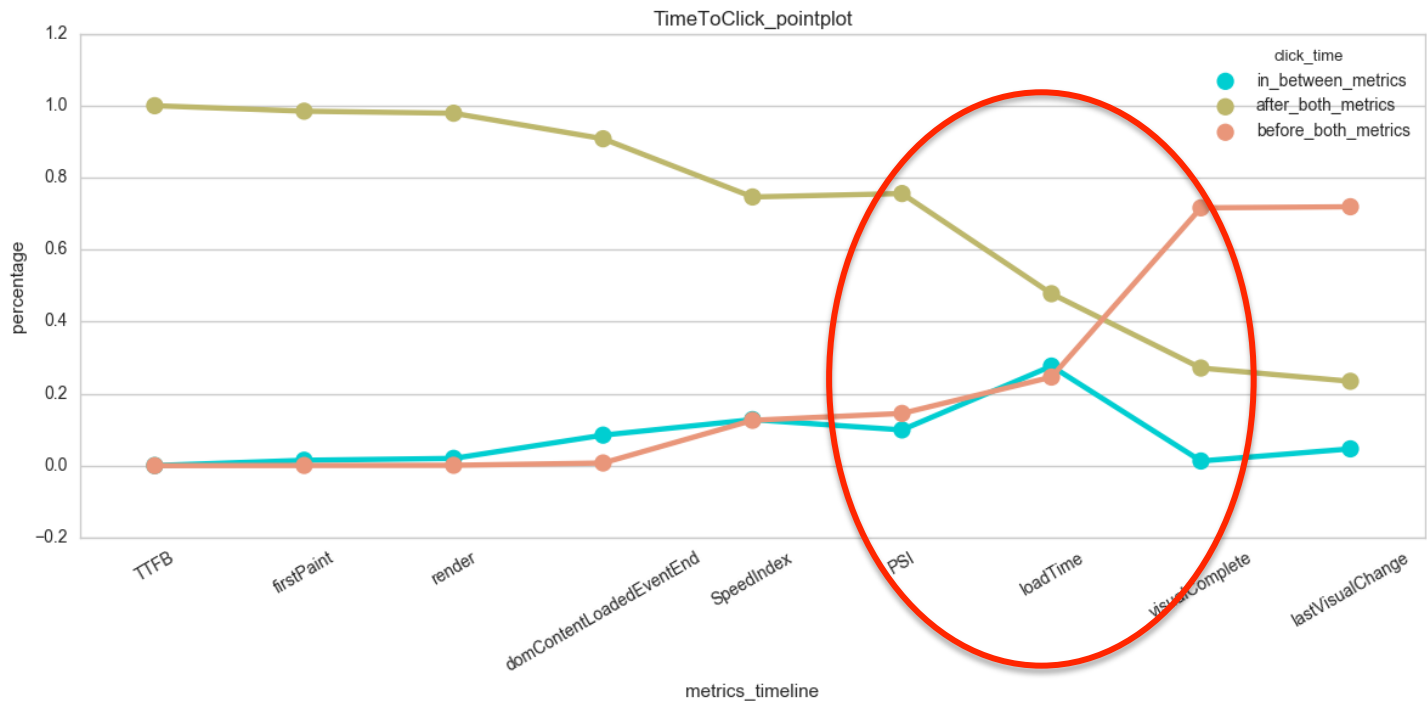
SpeedPerception Benchmark / Order of events



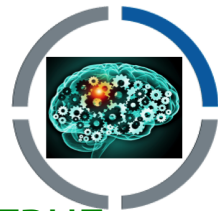
SpeedPerception Benchmark / Hypothesis # 3



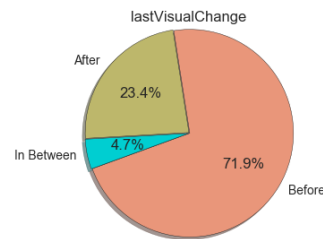
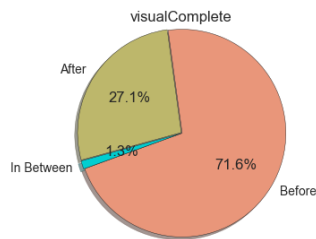
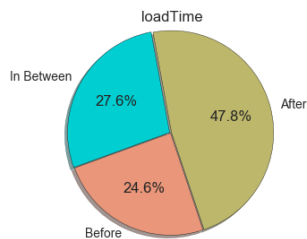
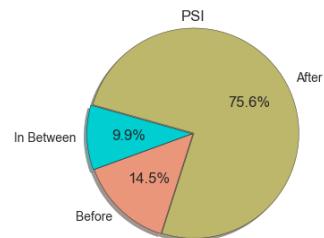
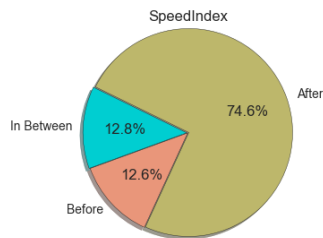
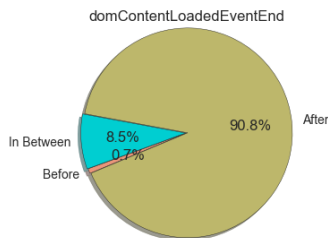
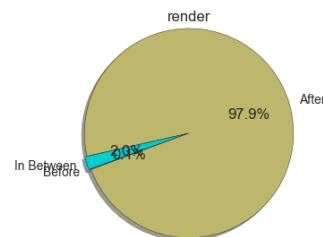
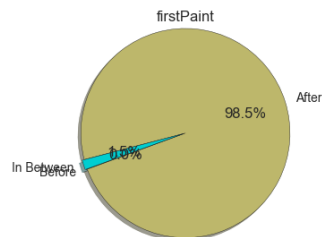
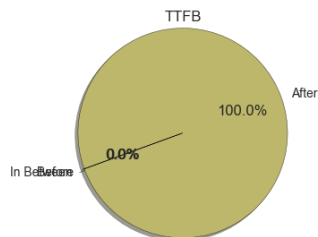
- H3:** User will not wait until “Visual Complete” to make their choice - **TRUE**



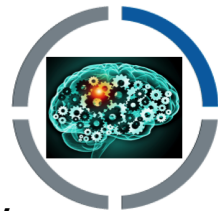
SpeedPerception Benchmark / Hypothesis # 3



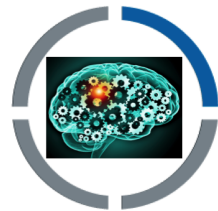
- H3:** User will not wait until “Visual Complete” to make their choice - **TRUE**



Questions to consider



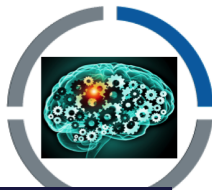
- Time to click is Time to decide plus some delay. How is the time delay influenced by the device/UA?
- How will the Time to click change depending on the presence/absence of visual jitter elements (modals / pop-ups / carousals) in the website structure?



Thoughts / Looking ahead..

- A rich source to explore & model webperf ideas with human A/B ground truth
 - *Play with the data/code and explore !!*
 - Need to look into Mobile rendering and its impact on results (Phase-2)
 - URL sources ought to go beyond IR500 list / E-commerce vertical
 - Other browsers beyond Chrome might be worth sampling
 - Would it be worth capturing & including Dev. Timeline information?
 - Better sampling of URLs to understand impact of interstitials / visual jitter
-
- Suggestions? Send to @perceptPA or parvez@ieee.org

SpeedPerception Challenge: Phase-2 planning underway



- Benchmark, analysis code and findings from Phase-1 available at:
 - <http://SpeedPerception.org>
 - Play with it, and explore !!
- Send feedback to @perceptPA or parvez@ieee.org