Convenience Samples

POLSCI 4SS3

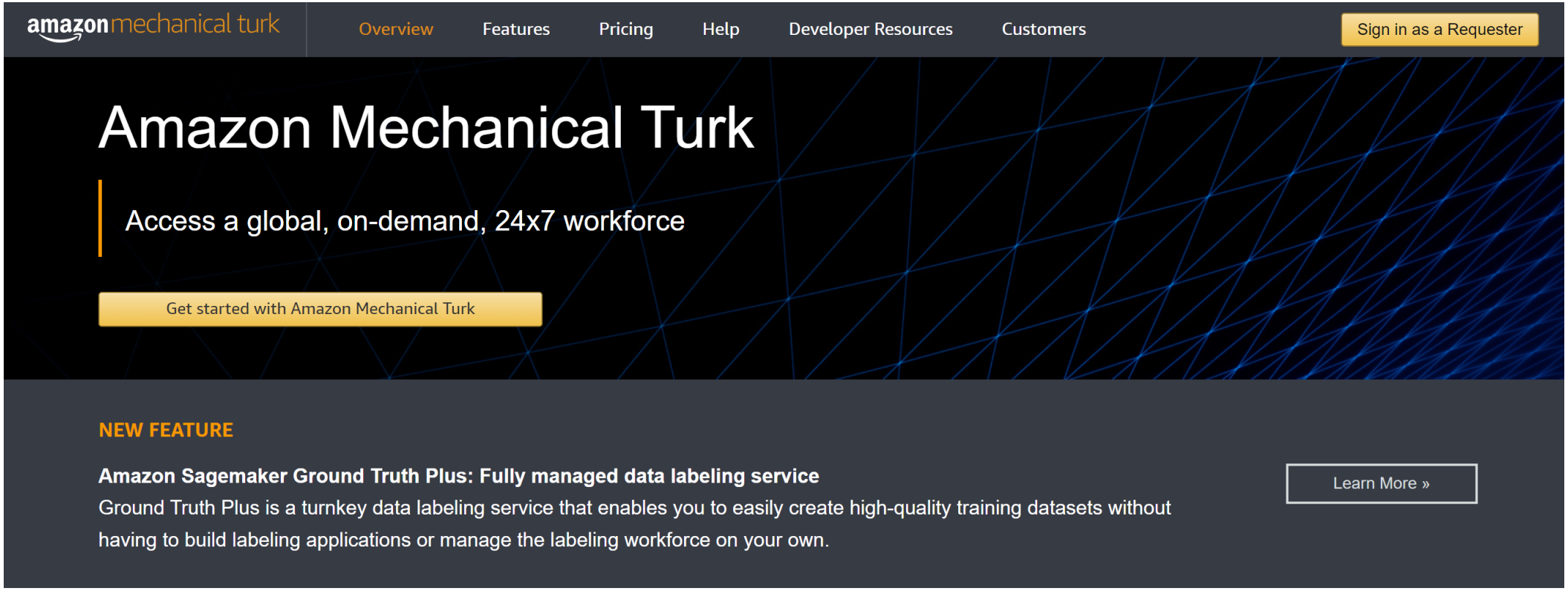
Winter 2023

Course so far

- Representative surveys as the gold standard
- Other research design help us learn more but tend to use convenience samples
- **Today:** Talk more about convenience samples

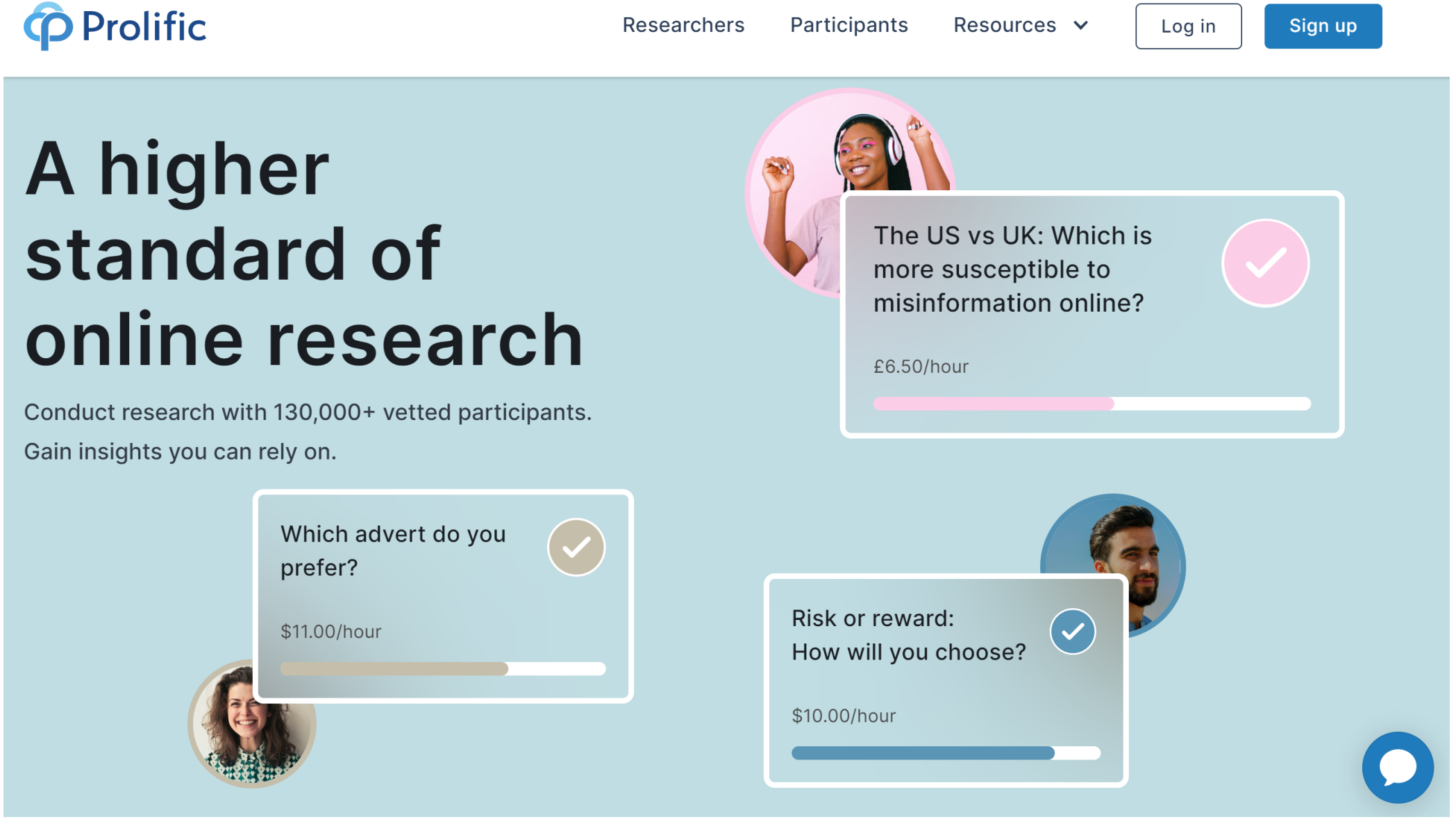
We are seeing less of this

And more of this

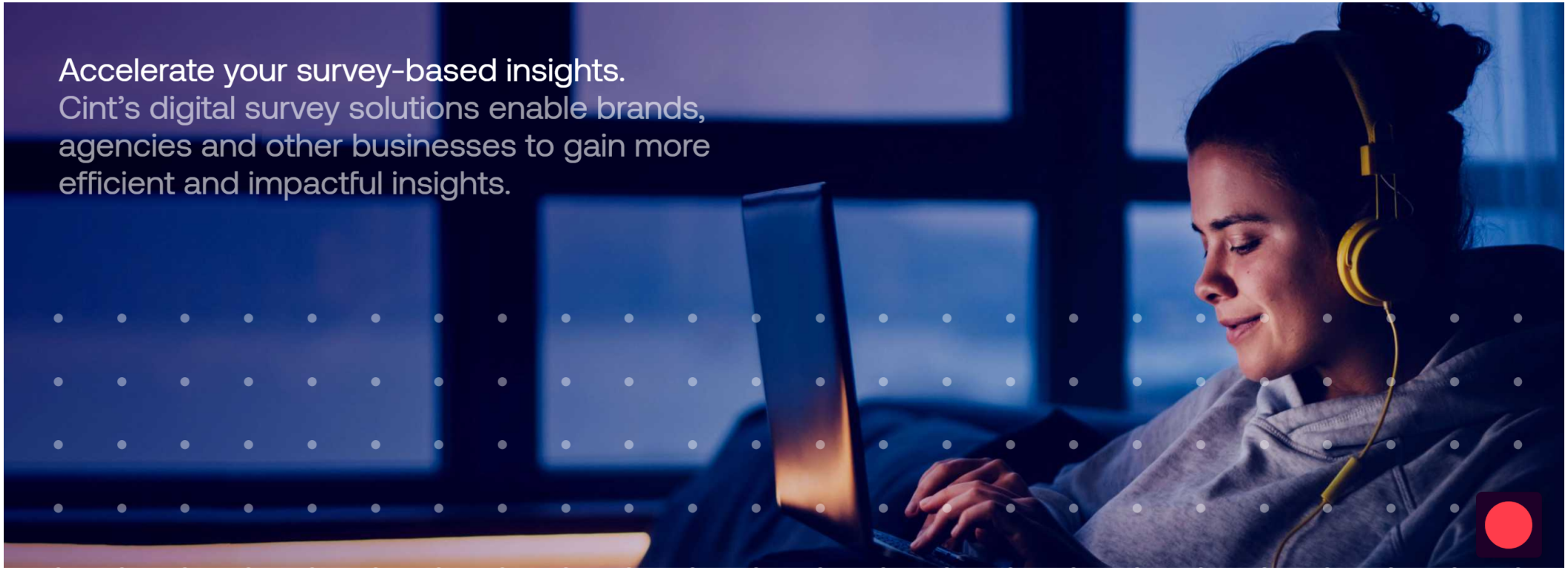


Amazon Mechanical Turk (MTurk) is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually. This could include anything from conducting simple data validation and research to more subjective tasks like survey participation, content moderation, and more. MTurk enables companies to harness the collective intelligence, skills, and insights from a global workforce to streamline business processes, augment data collection and analysis, and accelerate machine learning development.

And more of this



And more of this



Do we want surveys to be representative?

- Pros?
- Cons?
- We always want them!
- But when do we **need** them?
- Rather, when can we **get away** with not having them?

Internal an external validity

- **Validity:** Approximate truth or usefulness of an *inference*
- **Inference:** How we interpret the results of a study
- **Internal validity:** Whether inferences from a single study *cannot* be explained by other factors
- **External validity:** Whether inferences from a single study apply to a broader population or other target populations
- Convenience samples make it *easier* to achieve *internal validity* at the *expense* of *external validity*

Types of internal validity

1. X-validity `(endogenous variables)`
2. T-validity `(treatments, conditions)`
3. Y-validity `(outcome variables)`
4. C-validity `(context)`

See Egami and Hartman (2022) for more

X-validity

- Is the sample comparable to the target population?
- If not, can we claim that the differences can be ignored?
- To do that, we have to convince ourselves that:

1. Effects are the same across units

OR

2. We observe all the variables that may explain discrepancies in effects

T-validity

- Do treatments `(conditions)` reflect what participants would encounter in the real world?
- **Example:** Is thinking about *hypothetical* countries a good reflection to how people would think about *real* countries?
- Can we claim that there are no different versions of the same *treatment*?
- To do that, we need to convince ourselves that everyone would interpret vignettes in the same way
- Either because it is **realistic** enough or abstract yet **believable**

Y-validity

- Do the outcomes we measure in surveys reflect the outcomes we want to learn about in the real world?
- **Example:** Are self-reported vote intentions a good replacement for actual voting behavior?
- Can we claim that there are no different versions of the same *outcome*?
- Need to convince ourselves that measured outcomes are sufficiently valid and reliable

C-validity

- Do results generalize from other contexts?
- **Example:** If it worked with students in Sweden, will it work with students in Canada?
- Can we claim that the same units would react in the same way if the study was conducted elsewhere?
- Need to convince ourselves that **context is irrelevant** for *similar people* in *different places*

Discussion

Munger et al (2021): Accessibility and generalizability

- Replicate 3 convenience sample survey experiments with representative sample

1. Social commentary and news source credibility
2. Facebook shares and news consumption
3. Issue framing and support for gun control

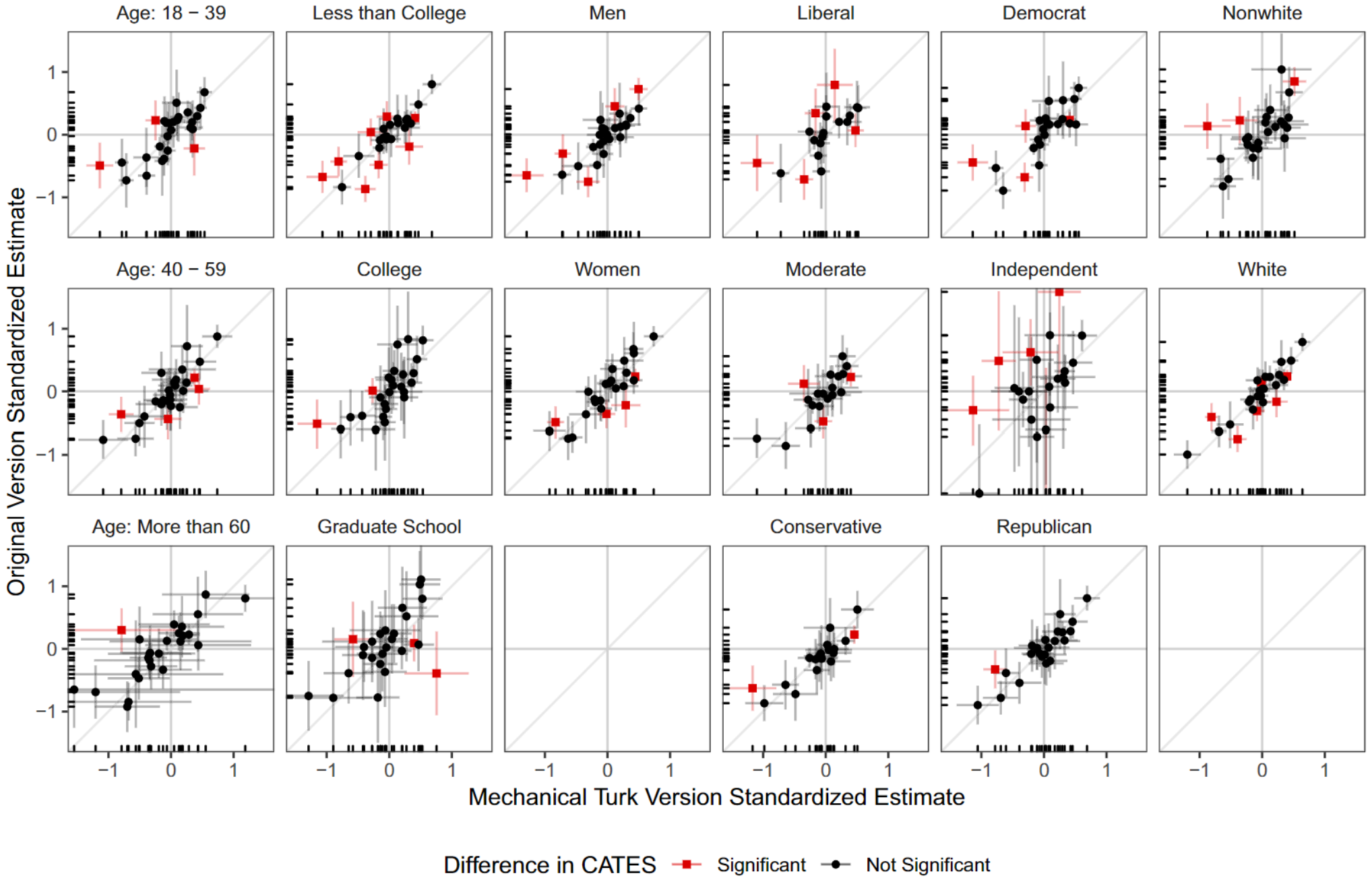
- **Argument:** Effects vary considerably by age and digital literacy

Findings

- **Replication 1:** Participants low on digital literacy did not respond differently to vignettes
- **Replication 2:** Older people clicked on whatever headline came first
- **Replication 3:** No differences because issue had nothing to do with digital literacy
- What kind of validity is this about?

Coppock et al (2018): Generalizability of heterogeneous treatment effect estimates across samples

- Replicate 27 studies from nationally-representative samples with convenience samples
- Compare how effects vary across 16 demographic characteristics


Age: 18 − 39
Less than College
Men
Liberal
Democrat
Nonwhite
Age: 40 − 59
College
Women
Moderate
Independent
White
Age: More than 60
Graduate School
Conservative
Republican
Original Version Standardized Estimate
Mechanical Turk Version Standardized Estimate
Difference in CATES
Significant
Not Significant

Explanation

- Different samples yield similar results when:

1. Treatment effects are mostly homogeneous
2. Effect heterogeneity is **orthogonal** to sample selection

- What type of validity is this about?

After Recess

Evidence-Informed Policy

Focus on: New topic!

Break time!


Lab