On the generalizability of sex-differences in risk attitudes

Jan Feld, Victoria University of Wellington

Are men generally more willing to take risks than women?

Phenomena (philosophy of science)

- Phenomena are stable and general features of the world.

- For example,
  - the return to education is positive
  - people respond to incentives

- They are **the social science equivalent of physical laws**.
  - Yes, there are exceptions (people are messy), but phenomena can guide how we think about the world.

Evidence from individual study

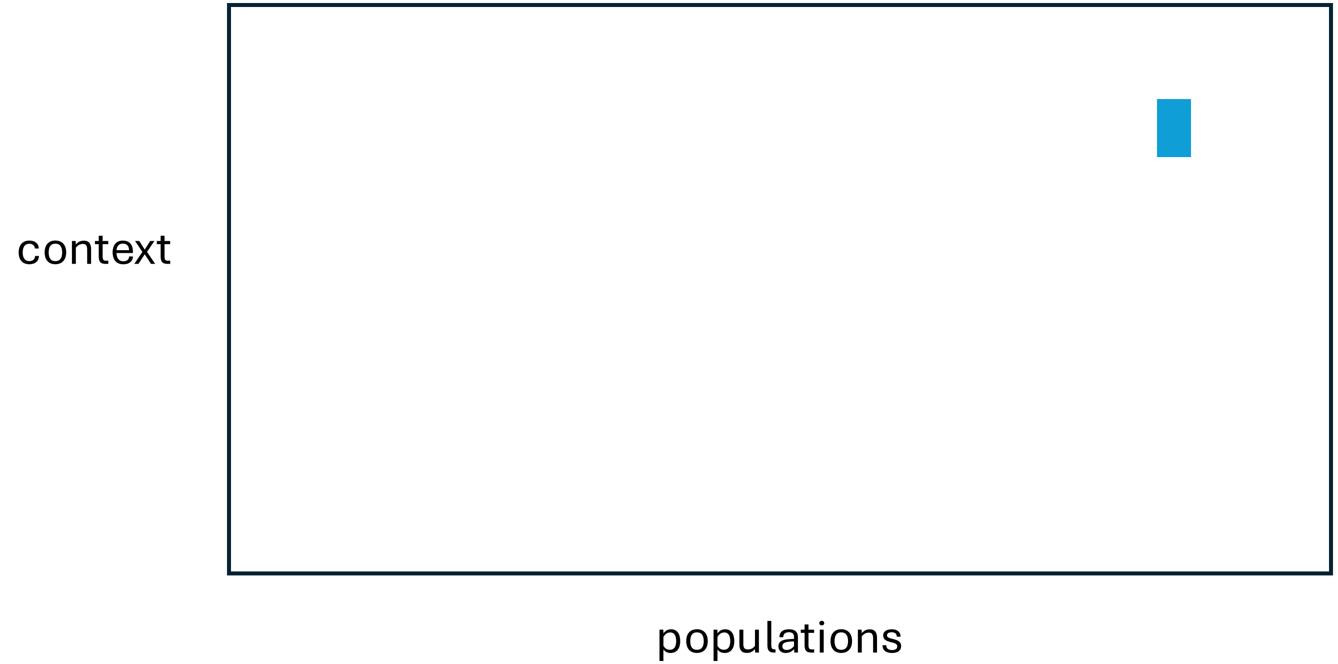
Abstract

This paper demonstrates gender differences in risk aversion and ambiguity aversion. It also contributes to a growing literature relating economic preference parameters to psychological measures by asking whether variations in preference parameters among persons, and in particular across genders, can be accounted for by differences in personality traits and traits of cognition. Women are more risk-averse than men. Over an initial range, women require no further compensation for the introduction of ambiguity but men do. At greater levels of ambiguity, women have the same marginal distaste for increased ambiguity as men. Psychological variables account for some of the interpersonal variation in risk aversion. They explain none of the differences in ambiguity. (JEL: J24, D03, D80)

From the introduction *"Our experiment is conducted on a sample of 347 15- and 16-year-old students at a Dutch high school."*

Source: Borghans, L., Heckman, J. J., Golsteyn, B. H., & Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association*, *7*(2-3), 649-658.

Individual study = single context



Legend

“positive effect”

“negative effect”

“No effect”

If the phenomenon is real,
we should find it most contexts.

A single study is not enough

- I am interested in <u>general</u> sex-differences in risk attitudes.

- A single study is not enough, because
  - Publication bias (file-drawer effect, p-hacking)
  - Results could depend on the measure
  - Results could depend on the populations

- Are meta-analyses the solution?

- They are <u>much</u> better but face similar challenges.

Challenge #1: Publication bias

- Meta-analysts spend a lot of time worrying about publication bias.

- It is hard to correct for, especially if plausible effect is small relative to the potential bias.

- <u>Example</u>: in a paper on same-sex teacher effects we apply 11 publication bias corrections (de Gendre et al., 2024)
  - Some show small positive effects, others show small negative effects.

Challenge #2: Research design distortions

- Researchers often design a study to find an effect.

- <u>Example</u>: In my paper on writing quality, I chose the study population (PhD students) for which I expected large effects.

- <u>Broader evidence</u>: Loss of effectiveness when interventions are implemented at scale (List, 2022; DellaVigna & Linos, 2022).

- Meta-analysis may be summarizing studies that are biased towards finding effects.

Challenge #3: Selected samples

- Researchers often use specific samples (e.g. students, Mturk)

- Example: Study population in most cited literature review on sex-differences in risk attitudes (Eckel & Grossmann, 2008):
    - United States (9x)
    - Switzerland (3x)
    - United Kingdom (2x)

- Some findings only hold in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations (Henrich et al., 2010)

- Meta-analyses may be summarizing studies that rely on selected samples.

What about sex-differences in risk-attitudes?

- Sex-difference in risk attitudes are likely small and plausibly context specific

- Much of the evidence is based on students in WEIRD countries (often in lab experiments)

- Potential for publication bias

What do we need?

- Many estimates about sex-differences in risk attitudes
  - which have not been distorted by the publication process (e.g. no p-hacking)
  - from diverse populations
  - from diverse measures which have not been chosen to maximize sex-differences

Meta-analysis with “quality-controlled” inputs

- Step 1: Find datasets that allow me to measure sex-differences in risk attitudes for many countries.

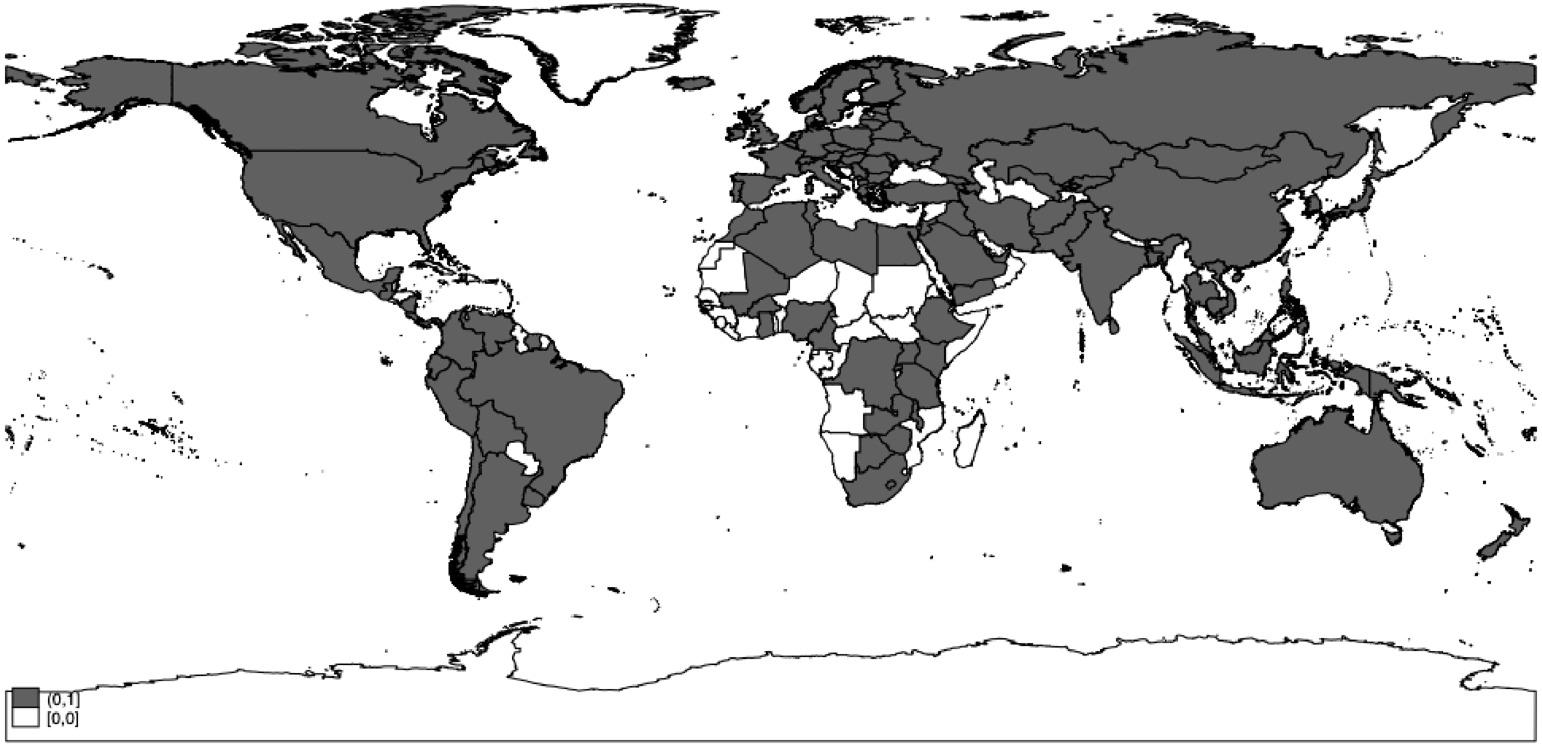
- Step 2: Estimate sex-differences in risk attitudes for each country-measure combination

- Step 3: Analyze these estimates with meta-analysis methods to correct for sampling error.

Data

- 7 multi-country datasets (e.g. World Value Survey, Global Preference Survey)
- 525k people
- 120 countries

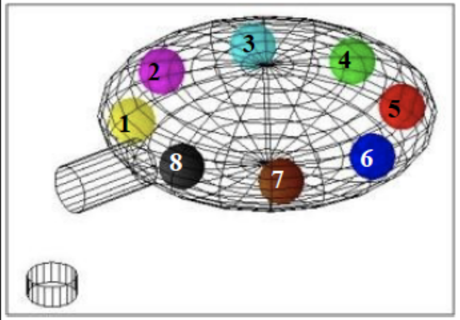
Countries any dataset



64 measures

- 42 incentivized lotteries
- 11 hypothetical lotteries
- 10 self-assessment questions
- 1 composite measure

Decision 1



Lottery	Sure	
O	O	€ 0.50 for sure
O	O	€ 1.00 for sure
O	O	€ 1.50 for sure
O	O	€ 2.00 for sure
O	O	€ 2.50 for sure
O	O	€ 3.00 for sure
O	O	€ 3.50 for sure
O	O	€ 4.00 for sure
O	O	€ 4.50 for sure

Win **€ 5** if one of the following balls is extracted:

1 2 3 4

Win **€ 0** if one of the following balls is extracted:

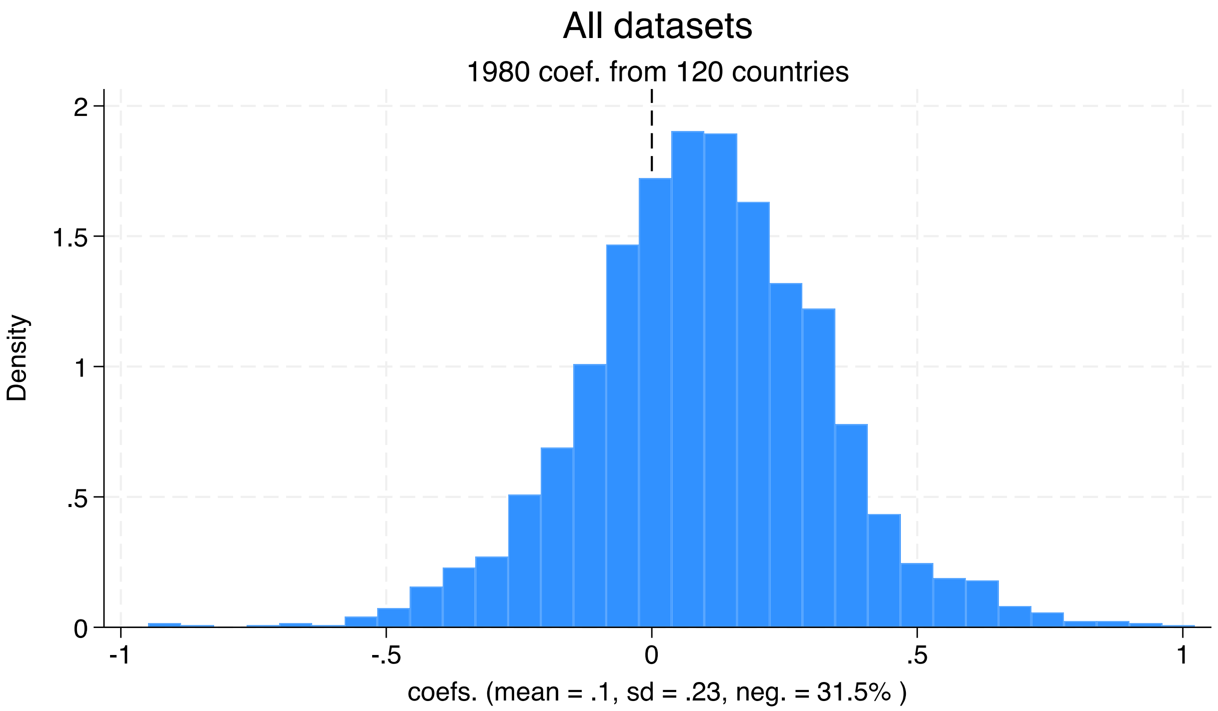
5 6 7 8

People can behave differently in different situations.
How would you rate your willingness to take risks in the following areas?
How is it …

risk averse — fully prepared to take risks

	0	1	2	3	4	5	6	7	8	9	10
– while driving?	O	O	O	O	O	O	O	O	O	O	O
– in financial matters?	O	O	O	O	O	O	O	O	O	O	O
– during leisure and sport?	O	O	O	O	O	O	O	O	O	O	O
– in your occupation?	O	O	O	O	O	O	O	O	O	O	O
– with your health?	O	O	O	O	O	O	O	O	O	O	O
– your faith in other people?	O	O	O	O	O	O	O	O	O	O	O

1,980 estimated mean sex-differences



One observation =
one country-measure combination

Example:

Average sex-differences in risk attitudes according to the **GPS measure** in **Germany**.

Positive values = men are on average more willing to take risk

<u>This distribution partly reflects sampling error</u>

Correcting for sampling error - BLUPs

- Estimate Best Linear Unbiased Predictor (BLUP) of sex-difference in effect size for each measure-country combinations

- BLUPs
  - are the weighted average of the mean and the country-specific estimate
  - shrink estimates towards the mean
  - give more weight to more precise estimates

Random effects models

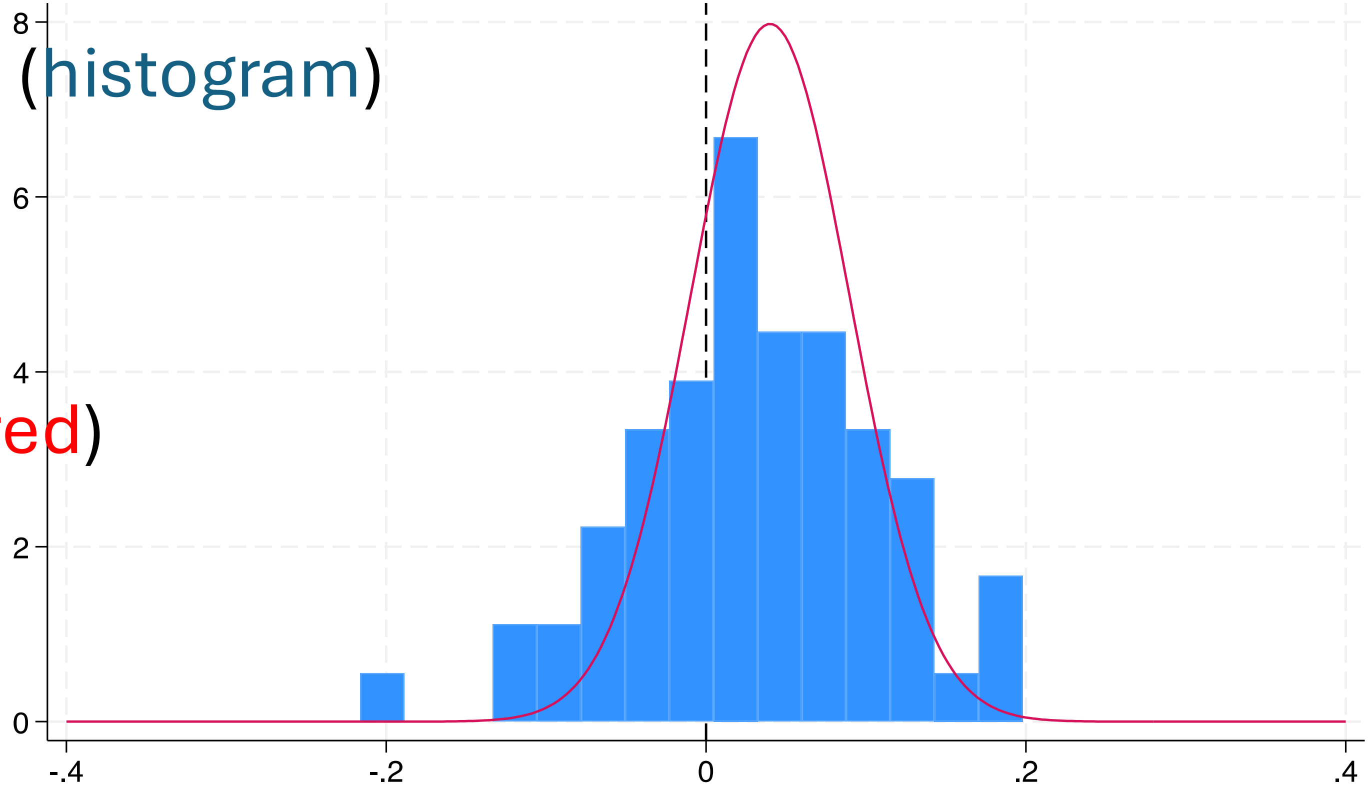
- Estimate random effects model for each measure (64 in total)

- Parameters of interest:
  - Mean (theta)
  - Variance (tau-squared)

Leveraging Normality assumption.
<u>Key statistic</u>: estimated share of countries in which men are more willing to take risks

Results Gallup measure

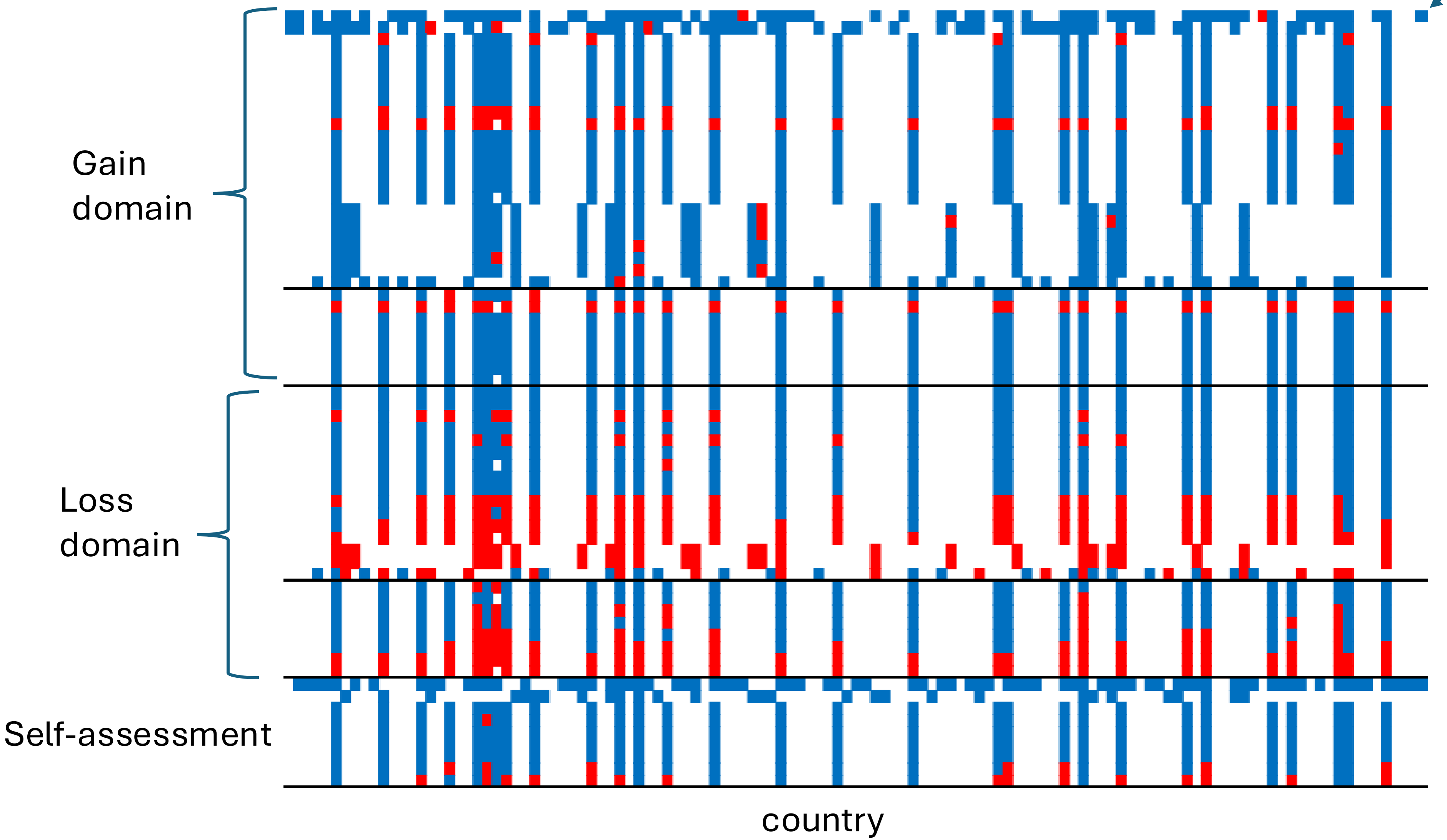
- Distribution of raw estimates (histogram)
  - Mean: 0.03 SD
  - Standard deviation : 0.08 SD
  - Positive: 66%
- Distribution of latent effect (red)
  - Mean: 0.04 SD
  - Standard deviation: 0.05 SD
  - Positive: 79%



→ Men are more risk-seeking in 79% of Gallup countries (according to Gallup measure)

Results:1,980 BLUPs

BLUP for one country-measure combination



Blue = men are predicted to be more risk-seeking

Red = women are predicted to be more risk-seeking

Shares (64 measures)

- Estimated shares of countries in which men are more willing to take risks than women
  - Blue = men are more willing to take risks in 50%+ countries
  - Red = men are more willing to take risk in less than 50% countries

Lotteries (54 measures)

	Risk					Uncertainty				
Gain	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.93
	1.00	1.00	1.00	0.99	0.94	0.89	0.67	0.09		
	0.90	0.90	0.88	0.86	0.86					
	0.84	0.84	0.83	0.79	0.65					
	0.51	0.23		(22 measures)					(8 measures)	
Loss	1.00	1.00	1.00	0.97	0.90	0.83	0.82	0.72	0.65	0.58
	0.89	0.76	0.66	0.61	0.51	0.41	0.00	0.00		
	0.33	0.33	0.32	0.04	0.01					
	0.00									
				(16 measures)					(8 measures)	

Survey questions (9 measures)

Domain	general	general	finance	finance	sport	occup.	driving	health	people
	0.99	0.95	0.98	0.92	1.00	0.98	0.88	0.66	0.52

Composite (1 measure, GPS)

0.97

Results summary

- Evidence for two phenomena
  - Men are generally more risk-seeking for financial gain lotteries
  - Men generally describe themselves as being more willing to take risks

- Men may not be generally more willing to take risks in financial loss lotteries (to be confirmed)

- Not much data on sex-differences in risk attitudes in non-financial domains (e.g. health decisions)

- The literature relies disproportionately on measures and populations which find large and consistent sex-differences

Discussion

- Meta-analysis with “quality-controlled” estimates is a powerful approach that avoids many of the shortcomings of traditional meta-analyses.
- But it is not always possible.
- But there are other things we can do get better at investigating phenomena. I have some ideas…

Things I would like to see more of (1)

- Explicitly trying to establishing phenomena (vs summarizing the literature)
- Discussions of selection into the literature.
- Ways to deal with research design distortions.
  - Idea: look a estimates that were not central to a paper.

Things I would like to see more of (2)

- Discussion of the <u>distribution</u> of the true effect in the literature.
- Explicit discussion of the context we know about generalizability (e.g. Western countries)
- Collection of additional data
- Calls for more data collection in specific contexts

Thank you!

- Please tell me why I am wrong (now or after this presentation)

We should explicitly try to establish phenomena

A typical meta-analysis

- Focusses on summarizing the literature
  - Without much consideration about studies that made-it into the literature
- Focusses on estimating average effect (overall and for different subgroups)
- Focus on publication bias
  - Ignore design "bias"

How do we find phenomena?

- Literature reviews?
- Meta-analysis?

Challengers for meta-analyses

Selected samples

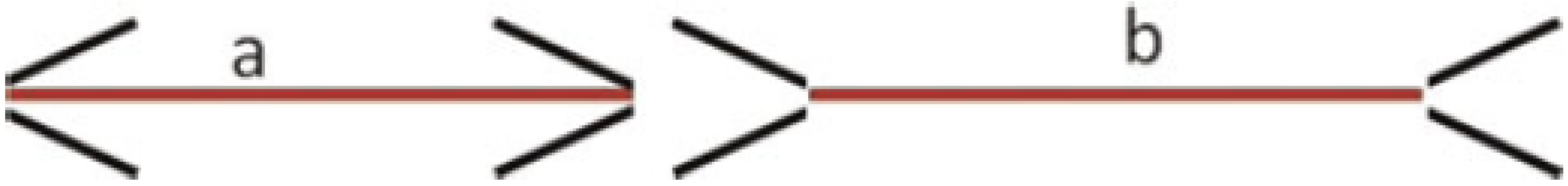
- In many literatures, studies focus on particular samples. For example:
  - University students
  - Mturk wrokers
  - The US population
  - Western, Educated, Industrialized, Rich, and Deomcratic (WEIRD) people

The WEIRD problem

- The drivers of effects (e.g. social norms, wealth, institutions) often differ between populations

- Henrich et al. (2010) show that many effects differ between Western Educated Rich Industrialized and Democratic (WEIRD) populations and non-WEIRD populations.
  - WEIRD populations are often outliers

Mueller-Leyer Illusion



The strength of this illusion varies substantially between populations and is even absent for some (San Forager, South African Miners).

Results summary (1)

- Some kinds of measures show men more willing to take risks
- Some kinds of measures show no sex-difference
- No kind of measure shows women are more willing to take risk

Focus on average effects

- In practice, meta-analysts try hard to get an unbiased estimate of the average effects in the literature (θ).

- Averages effects may hide important heterogeneity.
  - You can have a positive average even if 40% of studies have a negative effect.

- This also holds for averages across subgroups.

Design “bias”

- Researchers often design their studies with the goal of finding an effect.

- They choose context (e.g. sample, measure) that increase the chances of finding an effect.

- This is different from p-hacking.