Discussion of:

"Political Identity and Conjunction Fallacy: Experimental Evidence from the 2024 U.S. Presidential Election"

by Miao, Yang, Zhang, and Zhong (2025)

Cameron Peng, LSE and CEPR

2026 ABFER

The Linda problem (Tversky and Kahneman, 1983)

Linda is 31, single, outspoken, very bright. Philosophy major; deeply concerned with discrimination and social justice; participated in anti-nuclear demonstrations.

Which is more probable?

(a) Linda is a bank teller

(b) Linda is a bank teller and active in the feminist movement

The Linda problem (Tversky and Kahneman, 1983)

Linda is 31, single, outspoken, very bright. Philosophy major; deeply concerned with discrimination and social justice; participated in anti-nuclear demonstrations.

Which is more probable?

(a) Linda is a bank teller

(b) Linda is a bank teller and active in the feminist movement

By the conjunction rule, $P(A \cap B) \leq P(A)$: adding a conjunct can only narrow the set.

The Linda problem (Tversky and Kahneman, 1983)

Linda is 31, single, outspoken, very bright. Philosophy major; deeply concerned with discrimination and social justice; participated in anti-nuclear demonstrations.

Which is more probable?

(a) Linda is a bank teller

(b) Linda is a bank teller and active in the feminist movement

By the conjunction rule, $P(A \cap B) \leq P(A)$: adding a conjunct can only narrow the set.

Yet empirically, ~85% pick (b): the canonical conjunction fallacy (CF).

The Linda problem (Tversky and Kahneman, 1983)

Linda is 31, single, outspoken, very bright. Philosophy major; deeply concerned with discrimination and social justice; participated in anti-nuclear demonstrations.

Which is more probable?

(a) Linda is a bank teller

(b) Linda is a bank teller and active in the feminist movement

By the conjunction rule, $P(A \cap B) \leq P(A)$: adding a conjunct can only narrow the set.

Yet empirically, ~85% pick (b): the canonical conjunction fallacy (CF).

Why does this happen? The description is engineered: every trait maximizes feature overlap with "feminist," and minimizes overlap with "bank teller."

Three explanations for the Linda problem

- Representativeness (T&K 1983): people judge probability by similarity to the description.
  - "feminist teller" fits Linda's description better than "teller" alone ⇒ rated more probable
  - itself a form of problem substitution (Kahneman & Frederick 2002): answer "how representative?" instead of "how probable?"

Three explanations for the Linda problem

- Representativeness (T&K 1983): people judge probability by similarity to the description.
  - "feminist teller" fits Linda's description better than "teller" alone ⇒ rated more probable
  - itself a form of problem substitution (Kahneman & Frederick 2002): answer "how representative?" instead of "how probable?"

- Averaging (Yates & Carlson 1986): people average the two probabilities rather than multiply.

$$P(A \cap B) \approx w \cdot P(A) + (1 - w) \cdot P(B), \qquad w \in (0, 1).$$

When $P(\text{feminist}) \gg P(\text{teller})$, the average exceeds $P(\text{teller})$.

Three explanations for the Linda problem

- Representativeness (T&K 1983): people judge probability by similarity to the description.
  - "feminist teller" fits Linda's description better than "teller" alone ⇒ rated more probable
  - itself a form of problem substitution (Kahneman & Frederick 2002): answer "how representative?" instead of "how probable?"

- Averaging (Yates & Carlson 1986): people average the two probabilities rather than multiply.

$$P(A \cap B) \approx w \cdot P(A) + (1 - w) \cdot P(B), \qquad w \in (0, 1).$$

  When $P(\text{feminist}) \gg P(\text{teller})$, the average exceeds $P(\text{teller})$.

- Pragmatic reinterpretation (Hertwig & Gigerenzer 1999): subjects read "bank teller" as "bank teller *and not a feminist*".
  - actual comparison is "teller and feminist" vs. "teller and *not* a feminist": no rule is violated

Other CFs: a statistical link (Mr. F)

T&K (1983) also documented CF without an engineered stereotype: just a statistical link between conjuncts.

Mr. F is selected at random from the U.S. adult male population. Which is more probable?

(a) Mr. F has had one or more heart attacks

(b) Mr. F has had one or more heart attacks and is over 55

Other CFs: a statistical link (Mr. F)

T&K (1983) also documented CF without an engineered stereotype: just a statistical link between conjuncts.

Mr. F is selected at random from the U.S. adult male population. Which is more probable?

(a) Mr. F has had one or more heart attacks

(b) Mr. F has had one or more heart attacks and is over 55

A clear majority pick (b). Same violation, but no engineered description: just biographical facts and a strong statistical link, since most heart-attack victims are over 55.

Other CFs: a statistical link (Mr. F)

T&K (1983) also documented CF without an engineered stereotype: just a statistical link between conjuncts.

Mr. F is selected at random from the U.S. adult male population. Which is more probable?

(a) Mr. F has had one or more heart attacks

(b) Mr. F has had one or more heart attacks and is over 55

A clear majority pick (b). Same violation, but no engineered description: just biographical facts and a strong statistical link, since most heart-attack victims are over 55.

Suggested mechanism: people answer the easier conditional $P(\text{over 55} \mid \text{heart attack})$, which is near 1, instead of the joint $P(\text{both})$.

Other CFs: forecasting items

Carlos Alcaraz in the Wimbledon final. Which is more probable?

(a) Alcaraz wins the match

(b) Alcaraz loses the first set and wins the match

⇒ Subjects rate (b) more probable. “Comeback” is a coherent and exciting narrative.

Other CFs: forecasting items

Carlos Alcaraz in the Wimbledon final. Which is more probable?

(a) Alcaraz wins the match

(b) Alcaraz loses the first set and wins the match

⇒ Subjects rate (b) more probable. “Comeback” is a coherent and exciting narrative.

1983 forecasting items. Which is more probable?

(a) A complete suspension of U.S.-Soviet diplomatic relations in 1983

(b) A Russian invasion of Poland and a complete suspension of U.S.-Soviet relations in 1983

⇒ Subjects rate (b) more probable. The invasion supplies a plausible cause.

Other CFs: forecasting items

Carlos Alcaraz in the Wimbledon final. Which is more probable?

(a) Alcaraz wins the match

(b) Alcaraz loses the first set and wins the match

⇒ Subjects rate (b) more probable. “Comeback” is a coherent and exciting narrative.

1983 forecasting items. Which is more probable?

(a) A complete suspension of U.S.-Soviet diplomatic relations in 1983

(b) A Russian invasion of Poland and a complete suspension of U.S.-Soviet relations in 1983

⇒ Subjects rate (b) more probable. The invasion supplies a plausible cause.

Suggested mechanism: “good story”. The joint forms a coherent narrative, often with one event plausibly causing the other.

Takeaway 1: CF takes many forms, with many mechanisms

Three takeaways from the CF literature, before we turn to the paper.

First: the bias is not one thing. Different forms invite different mechanisms.

Example	**Form**	**Standard mechanism**
Linda the bank teller	description	representativeness, averaging
Mr. F's heart attack	statistical link	substitute the conditional
Alcaraz loses 1st set, wins match	within-domain narrative	"good story"
Russian invasion + diplomatic break	cross-domain causal narrative	"good story" + causal link

Takeaway 2: framing matters a lot

Reframings that leave the math unchanged often kill the CF:

- *Frequency format* ("of 100 Lindas, how many . . . "): ~85% → ~25%
- *Explicit partition* ("teller, *whether or not* a feminist"): blocks the implicit "and not a feminist" reading

> Probability judgments are constructed in the moment from framing cues, not retrieved from a stable underlying belief.

Closely related recent evidence:

- Bordalo, Conlon, Gennaioli, Kwon & Shleifer (2025): belief distributions are unstable, shaped by which features the prompt makes salient
- Fan, Liang & Peng (2024), *Inference-Forecast Gap*: same information, different elicitations ⇒ different beliefs

Takeaway 3: implications for survey design

Two lessons for belief elicitation.

Lesson 1: avoid compound questions (the rare but obvious case)

- "Probability of a U.S. recession caused by Fed tightening?"

Lesson 2: watch sequence effects (the common and subtle case)

- A prior question plants a scenario that respondents implicitly condition on next.
- E.g., "probability of a U.S. recession next year?" → "expected market returns? inflation? GDP growth?"
  - follow-ups are answered as "*if a recession hits,* what then?", narrower than intended

The current paper: design

An experimental study of CF in both choices and beliefs.

- Conducted two days before the 2024 U.S. presidential election
- $N = 1{,}171$ Prolific subjects
- Items pair an electoral outcome with an economic outcome (e.g., "Harris wins and unemployment falls")
- Both choices and beliefs
- Three between-subject treatments:
  - Main / Reverse-order: electoral first vs. economic first
  - Odd/Even: economic outcome replaced by a parity event (no valence)

The current paper: findings and interpretation

Main findings:

- CF is pervasive: ~33–44% of choices, ~30–60% of beliefs
- Congruence effect: CF is stronger when both outcomes align with the subject's identity (e.g., for a Democrat: "Harris wins and economy improves")
- In choices, the congruence effect *survives* controlling for CF in beliefs

Authors' interpretation: a preference-based mechanism, with state-dependent valence and source preferences

Comment 1: which CF form does the paper resemble most?

Recall the paper's items: "Trump wins and unemployment rises by Sept. 2025"

- Forward-looking real-world events (not features of a person)
- Two different domains (politics, economy) that can be linked by a plausible causal story

Closest match in the taxonomy: the cross-domain causal narrative form, like the Russian-invasion case.

Candidate mechanism: problem substitution (Kahneman & Frederick 2002), where people answer an easier conditional like "how compelling is the story that A leads to B?"

The next two slides present supporting evidence.

Comment 1 (cont.): Evidence 1 / standard mechanisms cannot fit the data

The two leading Linda-style mechanisms share a built-in ceiling:

$$\text{judged } P(A \cap B) \leq \max\{P(A), P(B)\}$$

- Representativeness: the joint can't fit the description better than its best piece
- Averaging: a weighted average lies between the two singles

In Linda, this ceiling holds. In the paper's data, it is broken (Fig. A9): the joint is valued above both single events in 15.97% of congruent and 10.18% of incongruent scenarios.

Neither representativeness nor averaging can produce this.

Comment 1 (cont.): Evidence 2 / substitution unifies all four findings

Substitution has no such ceiling. The conditional $P(B \mid A)$ can be near 1 for a coherent story, above both single probabilities.

One channel naturally produces all four findings:

- Pervasive CF: substitution is a default heuristic for hard joint probabilities
- Strong-form CF: conditional has no max-bound (Evidence 1)
- Congruence effect: identity shifts the substituted conditional (detailed in Comment 2)
- Odd/Even null: no valence means no compelling story to substitute

Comment 2: the role of valence (desirability)

Picking up from Comment 1: under substitution, the congruence effect comes for free, since identity shifts the conditional $P(\text{econ} \mid \text{electoral})$ that the subject substitutes for the joint.

Numerical illustration. Take a Democrat. Suppose she holds the following beliefs:

- $P(\text{Harris wins}) = 0.6$
- $P(\text{economy improves} \mid \text{Harris}) = 0.7$ (Harris is good for the economy)
- $P(\text{economy improves} \mid \text{Trump}) = 0.2 \quad \Rightarrow \quad P(\text{improves}) = 0.5$

Two contrasting scenarios:

- Congruent: Harris wins *and* the economy improves
- Incongruent: Harris wins *and* the economy declines

Comment 2 (cont.): substitution naturally produces the congruence effect

	Congruent scenario	**Incongruent scenario**
True joint probability	0.42	0.18
Smaller single prob. (min)	0.5	0.5
Larger single prob. (max)	0.6	0.6
What she substitutes: $P(B \mid A)$	0.7	0.3
Standard CF (substituted > min)?	YES	no
Strong-form CF (substituted > max)?	YES	no

An open question: if you elicited $P(B \mid A)$ *and* controlled for it, would the congruence effect disappear?
If a residual remains, it is harder to attribute to belief mechanics and points toward motivated reasoning / affect.

Comment 3: interpreting the residual

The paper's argument: regressing CF in choices on a congruence indicator and CF in beliefs, the congruence coefficient drops ~30% but stays significant. The leftover "residual" is read as evidence for a preference channel, over and above belief distortions.

The residual is real, but the belief control is doing limited work (Table 3):

	Congruence only	+ CF in beliefs	+ demographics
R^2 (own party wins)	0.013	0.024	0.056
R^2 (other party wins)	0.010	0.031	0.046

Adding CF in beliefs raises R^2 by only 1–2 percentage points.
At the individual level, congruence in choices and in beliefs barely correlate (Spearman = 0.19, Fig. A14).

Comment 3 (cont.): belief and choice CF are measured very differently

The two elicitations differ on multiple dimensions:

	Belief elicitation	**Choice elicitation**
Incentives	none (slider report)	paid for one chosen row
Format	continuous (0–100)	21 binary choices per item
Cognitive task	"assign a number"	"pick the better bet"
Anchoring	50% midpoint salient	indifference point salient

Any of these can drive a gap between belief-CF and choice-CF, without invoking preferences at all.
The Spearman = 0.19 correlation is consistent with both method noise and a genuine preference channel.

Comment 3 (cont.): narrowing the space, and a bigger question

Quick diagnostics on the existing data:

- Compare the wedge (Choice-CF − Belief-CF) across treatments; if similar in Odd/Even, it is method, not preferences.
- Cross-tabulate each (subject, item) by {Belief-CF} × {Choice-CF}; check whether the off-diagonals are symmetric (noise) or skewed toward congruent items (preferences).
- Slider diagnostics + engagement subsample: are subjects spiking at 50%? Does the residual coefficient shrink when restricted to engaged subjects?

A bigger question, worth its own paper:

why do belief and choice elicitations yield such different responses on identical events? Framing? Mental models? Something else? Worth investigating.

Wrap-up

A clean, ambitious study that brings CF into an economically meaningful, identity-loaded setting.

A single mechanism, substitution moderated by identity, can produce all four findings: pervasive CF, the strong form (joint > both), the congruence effect, and the Odd/Even null.

Two takeaways from the discussion:

- The 10–16% "joint > both pieces" rate is a fingerprint of substitution, not Linda-style mechanisms
- The belief–choice gap is striking on its own: why do the two elicitations diverge so much? Worth its own investigation.

An important agenda. Thank you!