Memory as a foundation for beliefs

Belief distortions in the long run, Stereotypes

Tilman Fries

Why talk about stereotypes?

Less information sometimes *improves* decisions

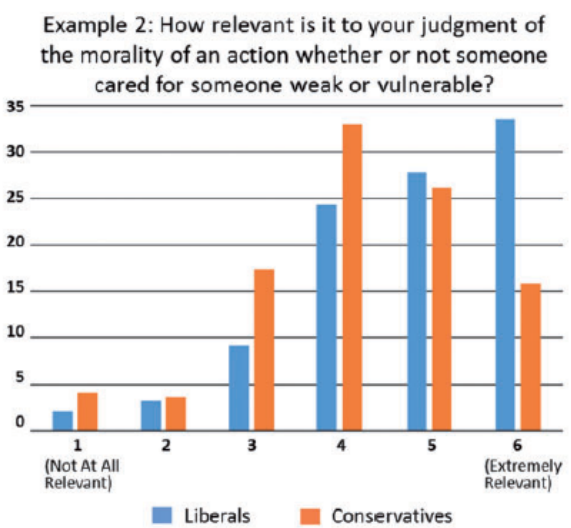
- E.g., the introduction of blind orchestral auditions increased female hires (Goldin & Rouse, AER 2000).

This implies **pre-existing stereotypes** can distort decisions.

How do such stereotypes form and persist?

This lecture: robust empirical patterns and a new memory-based model.

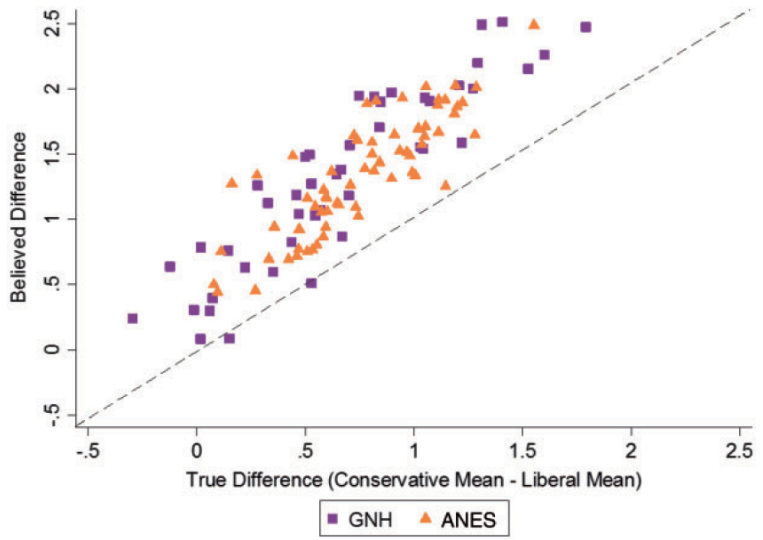
Stereotypes I: Moral-value differences, real vs. perceived



Bordalo et al. (2016) present evidence about liberal vs. conservative positions.

Actual gaps exist…

Stereotypes I: Moral-value differences, real vs. perceived



…but *believed* gaps are much larger.

Belief dots lie above 45° line → **exaggeration** of the partisan gap.

Stereotypes II: Representativeness heuristic

> Linda is 31, single, outspoken, very bright. Philosophy major, active in social-justice & anti-nuclear demos.
> Which is **more probable**?
> 1. Linda is a bank-teller
> 2. Linda is a bank-teller **and** a feminist

Stereotypes II: Representativeness heuristic

> Linda is 31, single, outspoken, very bright. Philosophy major, active in social-justice & anti-nuclear demos.
> Which is **more probable**?
> 1. Linda is a bank-teller
> 2. Linda is a bank-teller **and** a feminist

Tversky and Kahneman (PR, 1982) report that 85 % pick (2).

Stereotypes II: Representativeness heuristic

> Linda is 31, single, outspoken, very bright. Philosophy major, active in social-justice & anti-nuclear demos. Which is **more probable**?
> 1. Linda is a bank-teller
> 2. Linda is a bank-teller **and** a feminist

Picking (2) violates **sum rule**:

- The event "bank-teller + feminist" is a subset of the event "bank-teller". By the sum rule, $P(\text{bank-teller} + \text{feminist}) \leq P(\text{bank-teller}).$

The Linda example is interesting because it intuitively cues associations in our brain:

- *"Linda cares about social justice and is against nuclear power, so she **must be** a feminist."*

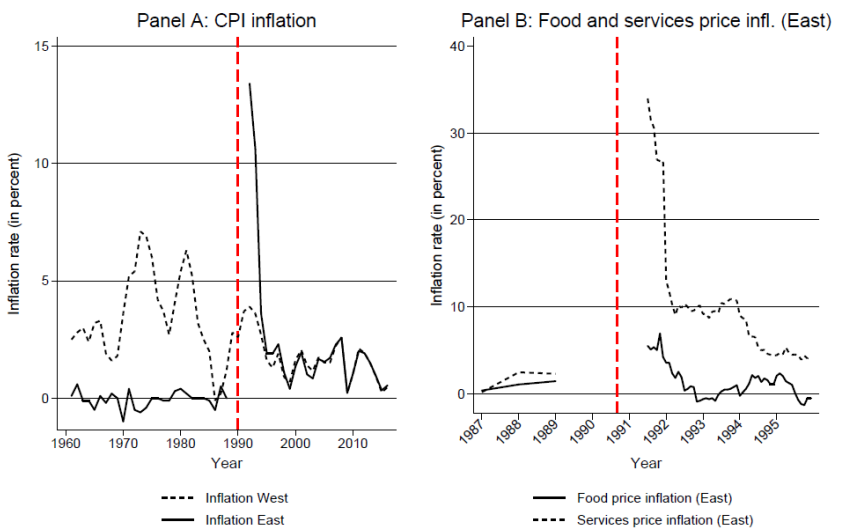
The same kind of thinking seems to occur in other contexts.

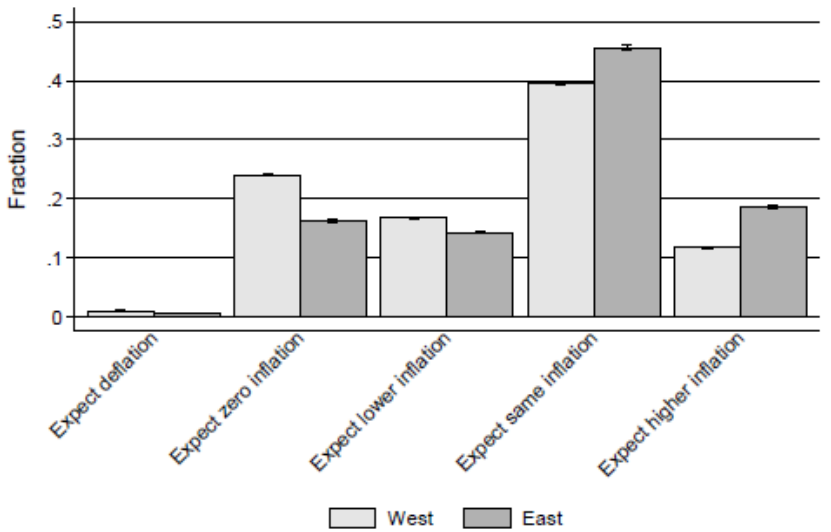
- For example, people may reason: *"He supports right-wing populism so he **must be** a misogynist, bigot, and have a worryingly high consumption of energy drinks."*

How do these associations occur? To understand this, we need to learn a bit about human memory…

Why talk about memory?

Good evidence that memory influences decisions.





Goldfayn-Frank and Wohlfahrt (JME, 2020): Inflation expectations in Germany were strongly shaped by the unification experience. This changes how people consume and invest.

Modeling memory

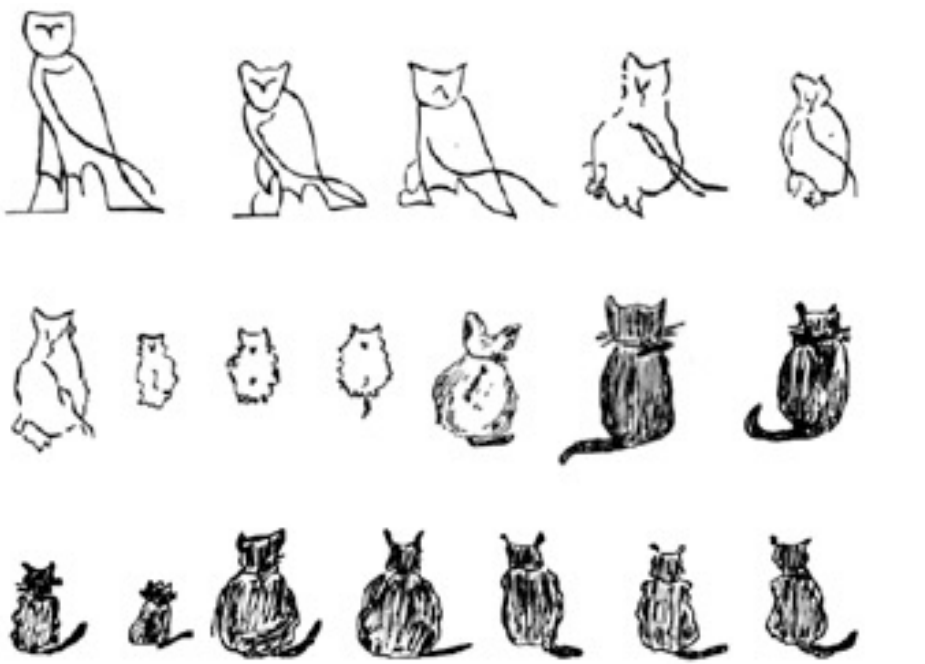
Suppose we aim to have a psychologically realistic model of how people remember. The first thing that we might want to assume is that people forget things.

This is the **sieve model** of memory. Individuals learn and store information, but over time, some details slip through and are lost.

Memory as a sieve?

The sieve model explains *some* memory features but misses out big parts. We do not only forget randomly, but also **suppress, reconstruct, and modify.**

This research started with Bartlett (1930), who documents systematic distortions over time.



In one experiment, participants reconstruct a painting of an owl, transforming it into a cat.

A better analogy may be **memory as Google search:**

1. Google is imperfect, you do not always find what you search for.
2. But what you find can change day-by-day, just as you might sometimes remember what you thought you had forgotten.
3. Google does not always provide you with exactly what you are looking for.
   - E.g., googling *"weather in munich"* might return ad results on summer clothes.
   - Just as when thinking about *summer,* you may make associations to *ice cream*.

A better analogy may be **memory as Google search.**

Psychologists broadly think about memory along these lines. They call it **associative memory.**

We will link associative memory to the existence and persistence of stereotypes.

Associative memory and stereotypes?

Question: What is the % of red hair among the Irish?

The basic argument for why associative memory makes us overestimate this %:

1. When I think about (red hair, Irish), a lot of reasons come to mind providing good evidence that many Irish persons are red haired.
2. When I think about (other hair, Irish), it is easy for my memory to make associations with *other-haired* people that are *not* Irish.

Associative memory and stereotypes?

1. When I think about (red hair, Irish), a lot of reasons come to mind providing good evidence that many Irish persons are red haired.
2. When I think about (other hair, Irish), it is easy for my memory to make associations with *other-haired* people that are *not* Irish.

   Many non-Irish other-haired people exist.

   Relatively fewer non-Irish red-haired people exist.

   Therefore, *memory interference* is larger when I think about (other hair, Irish) than about (red hair, Irish) $\rightarrow$ This makes me overestimate the %.

A minimal formal set-up

Bordalo, Conlon, Gennaioli, Kwon, Shleifer (QJE, 2022)

Consider an agent with a **memory database** E which contains entries with *attribute vector* $a = (a_1, a_2)$.

- For example: $a_1 =$ hair color (red/other), $a_2 =$ nationality (Irish/non-Irish).

The agent's **task** is to estimate a conditional share $P(a_1|a_2)$.

- For example $P(a_1 = \text{red hair}|a_2 = \text{Irish})$.

The memory database

Think of the memory database as capturing the joint distribution of (a_1, a_2). For example, the database may be:

	Red Hair	**Other hair**
Irish	1%	9%
Non-Irish	1%	89%

- Searching for $(\text{Red Hair}, \text{Irish})$ returns $E(\text{Red Hair}, \text{Irish}) = 1\%$.
- An unbiased of $P(\text{Red Hair} \mid \text{Irish})$ is $\frac{0.01}{0.01+0.09} = 10\%$.

Assumption 1 — Sampling and Counting

Sampling and counting. To form belief $P(a_1 = x | a_2 = y),$ an agent sends two requests to their memory database, $q_1,$ asking for T samples as evidence of $(a_1 = x \text{ and } a_2 = y)$ and $q_2,$ asking for $T > 0$ samples as evidence of $(a_1 \neq x \text{ and } a_2 = y).$ The agent then keeps all the results that cohere with each request. Denote the observed coherent results by R_{q_1} and R_{q_2}, respectively. In a second step, the agent combines these requests to obtain the estimate:

$$\hat{P}(a_1 = x | a_2 = y) = \frac{R_{q_1}}{R_{q_1} + R_{q_2}}.$$

Sampling & Counting Example

Recall the memory database, which captures the joint distribution:

	Red Hair	**Other hair**
Irish	1%	9%
Non-Irish	1%	89%

Suppose the agent samples with $T = 100.$

The agent first draws 100 samples from their database and keeps all results with (Red Hair, Irish) …

The agent then draws 100 samples from their database and keeps all results with (Other Hair, Irish) …

If this sampling is unbiased, then
$R_{q_1} \approx 0.1\% \cdot T = 10$ and $R_{q_2} \approx 0.9\% \cdot T = 90$.

The estimate based on unbiased sampling is approx. equal to the truth: $P(\text{red hair, Irish}) \approx \frac{10}{10+90} = 10\%$

If this sampling is unbiased, then $R_{q_1} \approx 1\% \cdot T = 1$ and $R_{q_2} \approx 9\% \cdot T = 9$.

The estimate based on unbiased sampling is approx. equal to the truth: $P(\text{red hair}|\text{Irish}) \approx \frac{1}{1+9} = 10\%$.

This sampling approach to belief formation is very common in cognitive psychology.

- It approximates the recall process, where people sort through examples in their memory.
- You can also think of this as **simulation.** Rather than sorting through past memories of meeting Irish people, people may simulate examples of how Irish people may be like.

Assumption 2 — Cued recall

Cued recall. When cued with a query q, the probability of recalling an entry with attributes $\mathbf{a}$ is equal to

$$r_q(\mathbf{a}) = \frac{S(\mathbf{a}, q)\ E(\mathbf{a})}{\sum_{\mathbf{a}' \in \mathcal{A}} S(\mathbf{a}', q)\, E(\mathbf{a}')}.$$

Above, S is a *similarity function* and $\mathcal{A}$ is the set of all possible attribute combinations.

Under cued recall, the agent is more likely to recall items similar to their query.

Cued recall: Example

Suppose that the similarity function is:

$$S(\mathbf{a}, q) = \begin{cases} 1 & , q_1 = a_1 \text{and } q_2 = a_2 \\ 1/2 & , q_1 = a_1 \text{and } q_2 \neq a_2 \text{ or } q_1 \neq a_1 \text{and } q_2 = a_2 \\ 0 & , q_1 \neq a_1 \text{and } q_2 \neq a_2 \end{cases}$$

Then, in our leading example, after searching for $q_1 = (\text{red hair}, \text{irish})$:

$$r_{q_1}(\text{red}, \text{irish}) = \frac{1 \cdot 0.01}{1 \cdot 0.01 + 1/2 \cdot 0.09 + 1/2 \cdot 0.01 + 0 \cdot 0.9} \approx 0.167$$

Cued recall: Example

Then, in our leading example, after searching for $q_1 = (\text{red hair}, \text{irish})$:

$$r_{q_1}(\text{red}, \text{irish}) = \frac{1 \cdot 0.01}{1 \cdot 0.01 + 1/2 \cdot 0.09 + 1/2 \cdot 0.01 + 0 \cdot 0.89} \approx 0.167$$

After searching for $q_2 = (\text{other hair}, \text{irish})$:

$$r_{q_2}(\text{other}, \text{irish}) = \frac{1 \cdot 0.09}{1 \cdot 0.09 + 1/2 \cdot 0.01 + 1/2 \cdot 0.89 + 0 \cdot 0.01} \approx 0.167$$

Combining both assumptions

Recall that, under sampling and counting, the agent samples T times for search terms q_1 and q_2 and forms the estimate $\hat{P}(\text{red hair}|\text{irish}) = \frac{R_{q_1}}{R_{q_1}+R_{q_2}}$.

By the law of large numbers, if T is large, the agent's belief is approx. equal to

$$\hat{P}(\text{red hair}|\text{irish}) \approx \frac{r_{q_1}(\text{red hair, irish})}{r_{q_1}(\text{red hair, irish})+r_{q_2}(\text{other hair, irish})}.$$

In our example, we have

$$\hat{P}(\text{red hair}|\text{irish}) \approx \frac{0.167}{0.167+0.167} = 50\%.$$

- The agent widely overestimates $P(\text{red hair}|\text{irish})$!

Model Mechanism

When searching (red hair, irish) , the agent either finds what they are looking for, or there is *interference* by (i) other hair Irish, (ii) red hair non-Irish.

When searching $(\text{other hair, irish})$, the agent either finds what they are looking for, or there is *interference* by (i) red hair Irish, (ii) other hair non-Irish.

Model Mechanism

When searching (red hair, irish) , the agent either finds what they are looking for, or there is *interference* by (i) other hair Irish, (ii) red hair non-Irish.

When searching (other hair, irish) , the agent either finds what they are looking for, or there is *interference* by (i) red hair Irish, (ii) other hair non-Irish.

1. In both searches, the terms interfere with one another.

Model Mechanism

When searching (red hair, irish) , the agent either finds what they are looking for, or there is *interference* by (i) other hair Irish, (ii) red hair non-Irish.

When searching $(\text{other hair, irish})$, the agent either finds what they are looking for, or there is *interference* by (i) red hair Irish, (ii) other hair non-Irish.

1. In both searches, the terms interfere with one another.
2. The searches are also interfered by the other group's members.

Model Mechanism

When searching (red hair, irish) , the agent either finds what they are looking for, or there is *interference* by (i) other hair Irish, (ii) red hair non-Irish.

When searching $(\text{other hair, irish})$, the agent either finds what they are looking for, or there is *interference* by (i) red hair Irish, (ii) other hair non-Irish.

$\rightarrow$ Because there are many more other hair non-Irish than red hair non-Irish, the search interference is much higher for q_2 than for q_1. This leads the agent to overestimate.

Model Summary

In the model, memory is probabilistic.

- The agent cannot access the whole memory database at once.

Searching the memory database is imperfect.

- The agent does not only end up with evidence for which they searched for.
- However, *cued recall* skews the search towards memories more similar to the search term.

Model Results

The model predicts that individuals overestimate features that are **diagnostic** of a certain group

- This is because there's less interference when recalling **diagnostic** features.
- E.g., red hair is relatively overrepresented (i.e., diagnostic) among the Irish.
- Other examples: Old people among Floridians, beer drinkers among Bavarians, rational calculators among economists.

→ Cued recall *amplifies* pre-existing differences.

Model Extensions

The model can be extended among several domains to make it more realistic. Two important domains are:

- **Imperfect memory and learning:** The agent may forget parts of their memory database or consciously add entries to it.
- **Skewed memory databases:** The memory database entries may be biased through: Cultural norms, media bias, subjective experiences, etc.

Even under such extension the main result still applies: Cued recall will *amplify* any pre-existing group differences present in the memory database.

Laboratory evidence for cued recall

Design – Bordalo et al. (2022)

People see words or numbers that are either colored orange or blue:

apple table 17 car music 42

river dog book 99

40 such items → filler task → ask for $P(\text{word} \mid \text{orange})$?

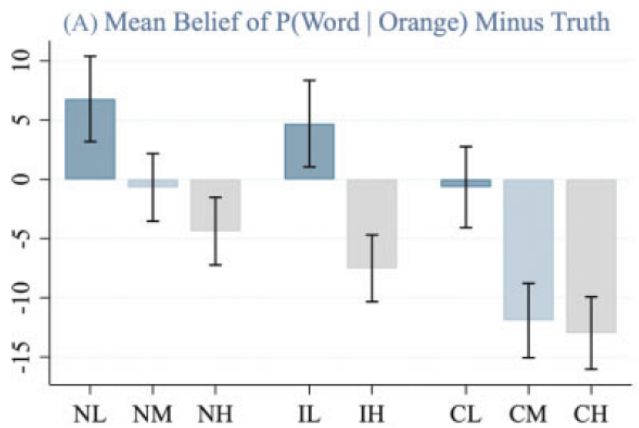
Treatments

Across treatments, they increase (i) the share of words among blue, (ii) the share of words among orange.

Treatment	Distribution	Distribution of irrelevant data	Sample size (N)	Elicited belief
Neutral	50% orange words, 50% orange numbers	NL: 0% blue words	147	P(word | orange)
		NM: 50% blue words	146	
		NH: 100% blue words	151	
Intermediate	55% orange words, 45% orange numbers	IL: 0% blue words	158	P(word | orange)
		IH: 100% blue words	154	
Common	70% orange words, 30% orange numbers	CL: 0% blue words	154	P(word | orange)
		CM: 30% blue words	149	
		CH: 100% blue words	144	

- Increasing blue words tests for **interference**.
- Increasing orange words tests whether recall is tied to the true underlying frequencies.

Results



1. More blue-word clutter reduces the recalled orange-word share.
2. Raising true orange-word share raises estimates, but **less than 1-for-1**.
   - ✔ Matches model predictions.

Field evidence — Media narratives

Memory associations induced by the media

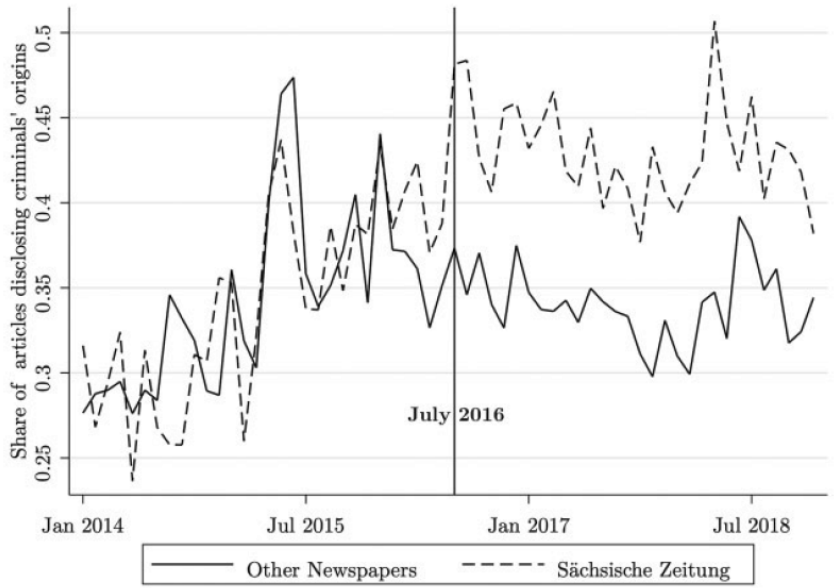
Keita, Renault, and Valette (EJ, 2023)

Testable model prediction: What kind of features people ascribe to groups depends on what they learned in the past.

When cued with a feature, individuals will draw on their memory to “fill in” the group that is likely to have this feature.

- The media potentially influences these associations by how they report about events.

KRV: Setting



KRV use an editorial change at the *Sächsische Zeitung* as a natural experiment. In 2016, the newspaper consistently began to disclose an offender's nationality whenever reporting on a crime.

KRV: Findings

	(1) Immigration	(2) Crime	(3) Health	(4) Environment	(5) Pers. Situation	(6) Economy
$July16_t \times E_l^{SZ}$	−0.087***	0.071***	0.049	0.030	−0.018	−0.018
	(0.026)	(0.009)	(0.033)	(0.028)	(0.023)	(0.024)
No. observations	110,364	110,260	110,269	110,225	110,200	110,112
Adjusted R^2	0.100	0.105	0.076	0.032	0.071	0.034
Average $Attitudes_{ilt}$	0.342	0.395	0.176	0.292	0.136	0.135

Regression outcome variable: “How important is topic X to you?”

Main dependent variable: SZ market share w/ post editorial change interaction

KRV: Findings

	(1) Immigration	(2) Crime	(3) Health	(4) Environment	(5) Pers. Situation	(6) Economy
$July16_t \times E_l^{SZ}$	−0.087***	0.071***	0.049	0.030	−0.018	−0.018
	(0.026)	(0.009)	(0.033)	(0.028)	(0.023)	(0.024)
No. observations	110,364	110,260	110,269	110,225	110,200	110,112
Adjusted R^2	0.100	0.105	0.076	0.032	0.071	0.034
Average $Attitudes_{ilt}$	0.342	0.395	0.176	0.292	0.136	0.135

→ Readers **downgraded** immigration worries, **upgraded** crime worries.

This suggests that the consistent disclosure of offender nationality introduced *counter-stereotypical* memory associations.

Takeaways

Stereotypes and belief distortions can arise from how memory works, not just from biased data.

Memory is **associative;** it amplifies diagnostic features of groups, leading to stereotypical beliefs.

Laboratory and field evidence support the model: memory interference and media cues shape beliefs.

Understanding memory-based belief formation helps explain persistent stereotypes.