



Intergroup Communication Interest Group  
Standard Paper Session:  
“**Decoding Hostility: Discrimination, Hate Speech, and Outgroup Perceptions in Media and Organizations**”

# Hate Speech on Social Media: Unpacking How Toxic Language Fuels Anti-Immigrant Hostility

Juan-José IGARTUA<sup>1</sup> & Carlos Ballesteros<sup>2</sup>

<sup>1</sup>University of Salamanca (Spain) – <sup>2</sup>University of Valladolid (Spain)



This QR code provides access to the research materials available on the **Open Science Framework** (OSF)

# Hates speech & social media



**Hate speech** is broadly defined as communication that incites, promotes, or instigates **hatred, humiliation, or contempt toward individuals or groups**, typically based on characteristics such as race, ethnicity, age, disability, gender, or sexual orientation (Arcila-Calderón et al., 2024; Castaño-Pulgarín et al., 2021; Hietanen & Eddebo, 2023).



**Social media** have increasingly become into platforms for the propagation of hate speech against minority groups, such as immigrants (Müller & Schwarz, 2021).

Arcila, C., Sánchez-Holgado, P., Gómez, J., Barbosa, M., Qi, H., Matilla, A., Amado, P., Guzmán, A., López-Matías, D. & Fernández-Villazala, T. (2024). From online hate speech to offline hate crime: the role of inflammatory language in forecasting violence against migrant and LGBT communities. *Humanities and Social Sciences Communications*, 11(1), 1–14.

# Hate speech against immigrants: key features

- Hate speech targeting immigrants frequently employs **harmful stereotypes**, often depicting them as **threats** to national identity, economic stability, or social security (Essalhi-Rakrak & Pinedo-González, 2023).
- Hate speech against immigrants is not only an expression of prejudice but also **a tool for political mobilization**, where certain groups use it to galvanize support or create divisions within society (Abuín-Vences et al., 2022; Carlson, 2020; Ikeanyibe et al., 2018).



While **prior research** has predominantly focused on measuring the scope of the problem through **content analysis and computational methods** (e.g., Arcila et al, 2022; Ayo et al., 2020; Lingiardi et al., 2019; Matamoros-Fernández, 2017), there is **limited research examining the effects of hate messages on recipients**. (e.g., Hsueh et al., 2015; Pluta et al., 2023; Soral et al., 2018).

# Knowledge gaps

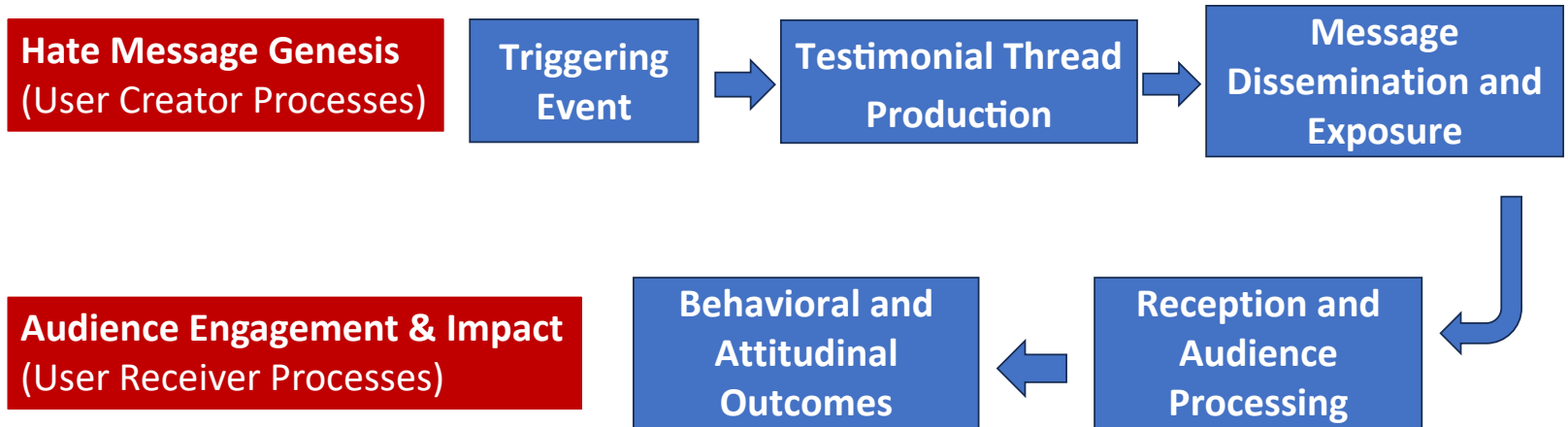
- The influence of specific message features within hate speech remains underexplored.
- The role of **linguistic toxicity**—marked by dehumanizing, aggressive, or obscene language—is poorly understood in terms of its capacity to enhance or suppress persuasive effects.
- Another understudied feature is the effects of the presence of **popularity cues**—such as the number of likes, shares, or comments—that often accompany social media posts (Dvir-Gvirsman, 2019; Sung & Lee, 2015; Woods, 2023).
- This study addresses these gaps by analyzing the **joint effect of toxic language and the perceived popularity** of hate messages.

# THREAD Model

A narrative-centered model to explain how anti-immigrant hate messages evolve and exert social influence on social media.

## Toxic Hate Responses Emerging After Disruptive events

Toxic  
Hate  
Responses  
Emerging  
After  
Disruptive events



Many online hate messages targeting immigrants emerge in response to news events and often take the form of **narrative threads**.



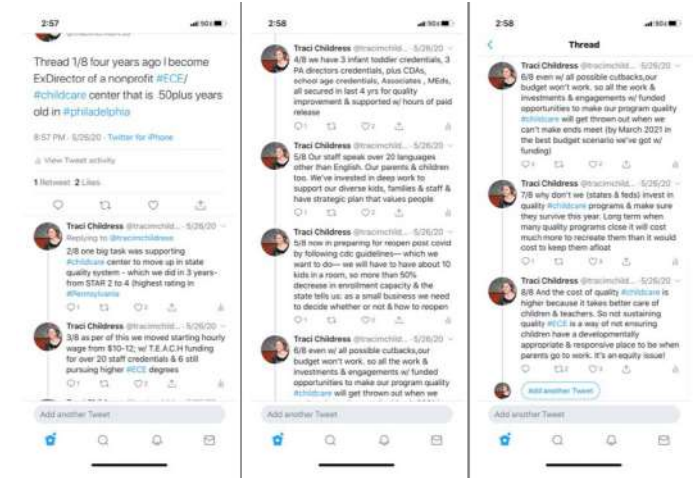


# THREAD Model

## Toxic Hate Responses Emerging After Disruptive events

### Testimonial Thread Production

- Ideologically-driven narrative production.
- Some users create **first-person testimonial threads**, connecting the public event to personal experiences.
- Users respond based on their **political ideology**.
- **Linguistic Framing**: Use of **toxic language** (e.g., dehumanizing, aggressive terms) shaped by the user's ideology.
- **Intent to Persuade or Mobilize**: The narrative is structured to resonate emotionally and ideologically with like-minded audiences.



**Toxic language** refers to the use of derogatory, dehumanizing, and hostile expressions aimed at stigmatizing, discrediting, or inciting hatred against individuals or groups based on their identity (Kim et al., 2021).

**Political conservatism** is associated with **stronger anti-immigration attitudes** (e.g., Davidov et al., 2020), suggesting that individuals with more conservative views may be more susceptible to negative messaging about immigrants.

# THREAD Model

## Toxic Hate Responses Emerging After Disruptive events

### Reception and Audience Processing

- Audience reactions depend on both:
  - **Toxicity of the message:** Presence or absence of toxic language.
  - **Political ideology of the receiver:** Influences openness to the message or rejection of it.
- Two psychological **mechanisms** mediate this processing:
  - **Narrative transportation:** Immersive engagement with the story.
  - **Identity fusion with the author of the message:** Perceived alignment or merging with the author of the testimonial.

**Narrative transportation** refers to the psychological process in which individuals become absorbed in a story (Appel et al., 2015; Green & Brock, 2000).

**Identity fusion with the author of the message** is a concept that describes a psychological state in which individuals feel a strong sense of connection with the author of a message (Swann et al., 2012).

Swann, W. B., Gómez, A., Seyle, D. C., Morales, J. F., & Huici, C. (2009). Identity fusion: the interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology*, 96(5), 995–1011.

# THREAD Model

## Toxic Hate Responses Emerging After Disruptive events

### Behavioral and Attitudinal Outcomes

The model helps explain shifts in:

- Attitudes toward immigration.
- Support for harsh policies against irregular immigration.
- Intention to share the message, to spread or endorse hate content (which contributes to virality and normalization).



Much like immersing in a **gripping novel**, deep immersion (**narrative transportation**) makes the reader more receptive to the message's persuasive power, increasing the likelihood that they will **internalize the thread's themes** and respond in alignment with its emotional and ideological cues.



The reader's **identity merges** with that of the message author: they begin to feel the author's emotions as if they were their own. In this fused state, **the reader is more likely to adopt the author's perspective**, echo their outrage, or amplify their message — as if wearing **VR goggles** that immerse them in the author's personal experience.



# Study Objectives



Political ideology of the receiver

## THREAD Model

**Audience Engagement & Impact**  
(User Receiver Processes)

Toxic Language  
Message popularity

**Message  
Dissemination and  
Exposure**

**Behavioral and  
Attitudinal  
Outcomes**

Intention to share the message  
Attitudes toward immigration  
Support for harsh policies against irregular immigration

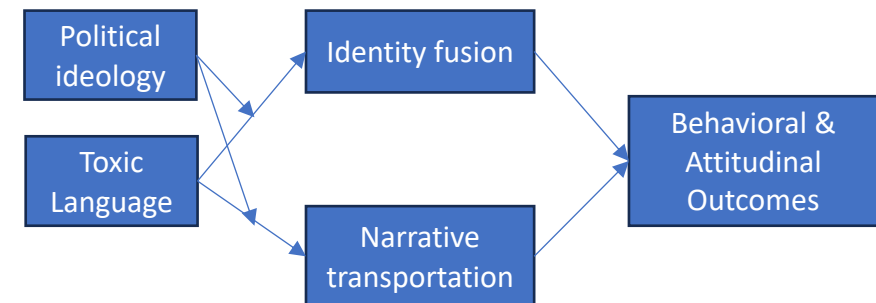
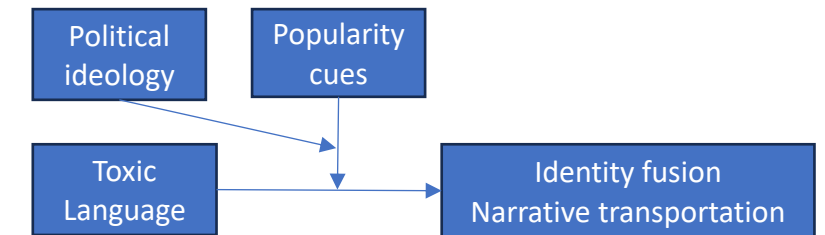
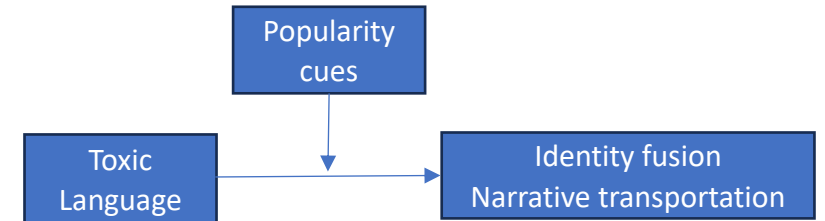
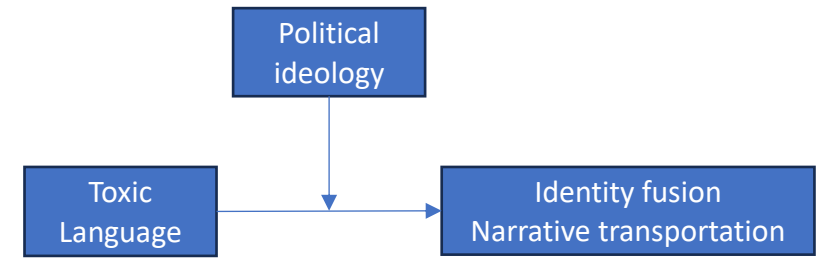
**Reception and  
Audience  
Processing**

Narrative transportation  
Identity fusion with the  
author of the message



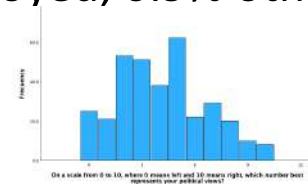
# Research questions

- **RQ1:** To what extent does the recipient's **political ideology** moderate the effect of toxic language in hate messages on identity fusion with the author and narrative transportation?
- **RQ2:** Does the presence of **popularity cues** (e.g., number of likes, shares, and comments) moderate the effect of toxic language in hate messages on identity fusion with the author and narrative transportation?
- **RQ3:** Does the recipient's **political ideology** further moderate the interaction between toxic language and **popularity cues** in shaping identity fusion with the author and narrative transportation?
- **RQ4:** Does political ideology moderate the **indirect effects** of toxic language on the intention to share the message, attitudes toward immigrants, and support for harsh policies against irregular immigration, through parallel mediation by narrative transportation and identity fusion with the author?



# Method: Online experiment with Qualtrics<sup>1</sup>

- **Participants ( $N = 339$ )**
  - Age: 18-64 years ( $M = 26.98$  years;  $SD = 11.57$ ).
  - Gender: 47.5% female, 50.7% male, 0.6% non-binary or third gender, and 1.2% “prefer not to say”.
  - Current employment status: 63.7% students, 33.3% employed workers, 2.1% unemployed, 0.9% other.
  - **Political ideology** (0 = left, 10 = right):  $M = 4.15$ ,  $SD = 2.50$  [ $Mdn = 4.00$ ,  $Mo = 5.0$ ].
- **Design:**
  - A **2 x 2 between-subjects factorial design**.
  - IV1: level of toxicity in the hate speech messages (low or high levels of toxic language).
  - IV2: number of interactions (low vs. high) associated with those posts.
  - Questionnaire with pre-test measures, testimonial message, and post-test measures.
- **Quality control procedures:** The initial sample consisted of 429 participants. However, quality control procedures were applied, considering the **reading time of the messages** (which ranged between 267 and 343 words). Only participants who spent between 80 and 300 seconds reading the message were included. Additionally, participants who completed the entire questionnaire in less than 360 seconds or more than 2,400 seconds were excluded from the analysis.





<sup>1</sup> A **pilot study** ( $N = 41$ ) was conducted to evaluate the perceived toxicity of expressions commonly used in online hate speech against immigrants and to calibrate interaction metrics for social media messages.

# Experimental Stimuli

- **Triggering Event:** A news story about the arrival of immigrants in Spain “**Hundreds of immigrants enter Melilla after a new mass assault on the border fence**”.
- **First-person testimonial messages** were presented in the form of Twitter (X) **threads**, simulating user-generated posts reacting to a **news story** about the arrival of immigrants in Spain and underscoring the **perceived threat** to Spanish society.
- **To enhance the external validity** of the study, the messages referenced **two types of threats**:
  - Unfair competition in the job market or **economic threat**.
  - Violence at the borders of Melilla or **threats to security**.

An English translation of the experimental message can be found in the OSF repository.

Economic threat + No toxic language + Low number of interactions	Economic threat + Toxic language + High number of interactions
<p>Juan Pérez @juanper_92</p> <p>Cierto que el salto a la valla de Melilla nos afecta a todos los españoles. Os cuento mi caso, que tiene traca. Abro hilo</p>  <p>voceopulli.com Cientos de inmigrantes acceden a Melilla tras un nuevo salto masivo a... La entrada masiva dejó a 49 agentes de la Guardia Civil heridos de carácter leve, al igual que 57 migrantes, de los que tres han sido ...</p> <p>1 0 3</p> <p>Juan Pérez @juanper_92 · 26 min</p> <p>Siempre he trabajado varios meses en el campo cogiendo fresas, naranjas, melones... lo que sale. Luego, con el boom de la construcción, me fui de yesero, se ganaba más y siempre había trabajo. Cuando se dejó de construir, quise volver al campo. Pero...</p> <p>0 0 1</p> <p>Juan Pérez @juanper_92 · 26 min</p> <p>...los agricultores ya no contratan a españoles, lo hacen directamente en Marruecos y Rumania. Los inmigrantes cobran menos y trabajan a destajo, mientras mi mujer, nuestra hija y yo malvivimos de su media jornada en el súper y alguna chapuza que me sale.</p> <p>0 0 0</p> <p>Juan Pérez @juanper_92 · 24 min</p> <p>Estamos viviendo los tres en un cuarto de la casa de su madre. Llevamos años apuntados en una lista del Ayuntamiento, pero parece que no tenemos derecho a un piso de protección oficial, que están llenos de moros y gitanos con todo tipo de ayudas.</p> <p>0 0 0</p> <p>Juan Pérez @juanper_92 · 24 min</p> <p>De madrugada ya no nos despierta el gallo, sino un altavoz de la "mezquita" que hay en una cochera debajo de nuestra habitación, llamando a la oración: lo hace cinco o seis veces al día. Los vecinos nos hemos quejado, pero el alcalde dice que no puedo hacer nada.</p> <p>0 0 1</p> <p>Juan Pérez @juanper_92 · 23 min</p> <p>Cuando vienen "a rezar", se junta en la calle una multitud de gente, todos con chilabas, no podemos ni salir. Vas a muchos trapichear con costo y quién sabe con qué más, a plena luz del día. Y la Policía ya no aparece por el barrio, cada vez parece más que estemos en una ciudad de África.</p> <p>0 0 0</p> <p>Juan Pérez @juanper_92 · 22 min</p> <p>Es tal la inseguridad y el acoso a las mujeres, que mi hija y mi mujer ya no pueden bajar solas a comprar o ir con sus amigas. ¿Cuántos son ilegales? ¿Algún político o juez revisa qué dicen y qué guardan en sus mezquitas?</p> <p>0 0 0</p> <p>Juan Pérez @juanper_92 · 21 min</p> <p>Mi solución: Si salían la valla, se les devuelve al otro lado, no queremos gente así aquí.</p> <p>0 0 1</p>	<p>Juan Pérez @juanper_92</p> <p>Cierto que el salto a la valla de Melilla nos afecta a todos los españoles. Os cuento mi caso, que tiene traca. Abro hilo</p>  <p>voceopulli.com Cientos de inmigrantes acceden a Melilla tras un nuevo salto masivo a... La entrada masiva dejó a 49 agentes de la Guardia Civil heridos de carácter leve, al igual que 57 migrantes, de los que tres han sido ...</p> <p>122 358 901</p> <p>Juan Pérez @juanper_92 · 26 min</p> <p>Siempre he trabajado varios meses en el campo cogiendo fresas, naranjas, melones... lo que sale. Luego, con el boom de la construcción, me fui de yesero, se ganaba más y siempre había trabajo. Cuando se dejó de construir, quise volver al campo. Pero...</p> <p>150 400 950</p> <p>Juan Pérez @juanper_92 · 26 min</p> <p>...los agricultores peseteros ya no contratan a españoles, lo hacen directamente en Marruecos y Rumania. Es tal la marabunta de inmigrantes, que cobran lo que les den y trabajan a destajo, mientras mi mujer, nuestra hija y yo malvivimos de su media jornada en el súper y alguna chapuza que me sale.</p> <p>120 364 908</p> <p>Juan Pérez @juanper_92 · 24 min</p> <p>Estamos viviendo los tres en un cuarto de la casa de su madre. Llevamos años apuntados en una lista del Ayuntamiento, pero parece que no tenemos derecho a un piso de protección oficial, que están llenos de sucios moros y gitanos con todo tipo de ayudas.</p> <p>146 394 943</p> <p>Juan Pérez @juanper_92 · 24 min</p> <p>De madrugada ya no nos despierta el gallo, sino un altavoz de la "mezquita" que hay en una cochera debajo de nuestra habitación, llamando a la oración: lo hace cinco o seis veces al día. Los vecinos nos hemos quejado, pero el alcalde sociata dice que no puede hacer nada.</p> <p>130 370 915</p> <p>Juan Pérez @juanper_92 · 23 min</p> <p>Cuando los moros vienen "a rezar", se junta en la calle un montón de escoria, todos con chilabas, no podemos ni salir. Vas a muchos trapichear con costo y quién sabe con qué más, a plena luz del día. Y la Policía traidora ya no aparece por el barrio, cada vez parece más que estemos en una ciudad de África.</p> <p>142 368 936</p> <p>Juan Pérez @juanper_92 · 22 min</p> <p>Es tal la inseguridad y el acoso a las mujeres, que mi hija y mi mujer ya no pueden bajar solas a comprar o ir con sus amigas. ¿Cuántos de estos parásitos son ilegales? ¿Algún político o juez hipócrita revisa qué dicen y qué guardan en sus sucias mezquitas?</p> <p>134 376 922</p> <p>Juan Pérez @juanper_92 · 21 min</p> <p>Mi solución: Si esta basura salía la valla, se les devuelve al otro lado, no queremos gente así en España.</p> <p>138 362 929</p>

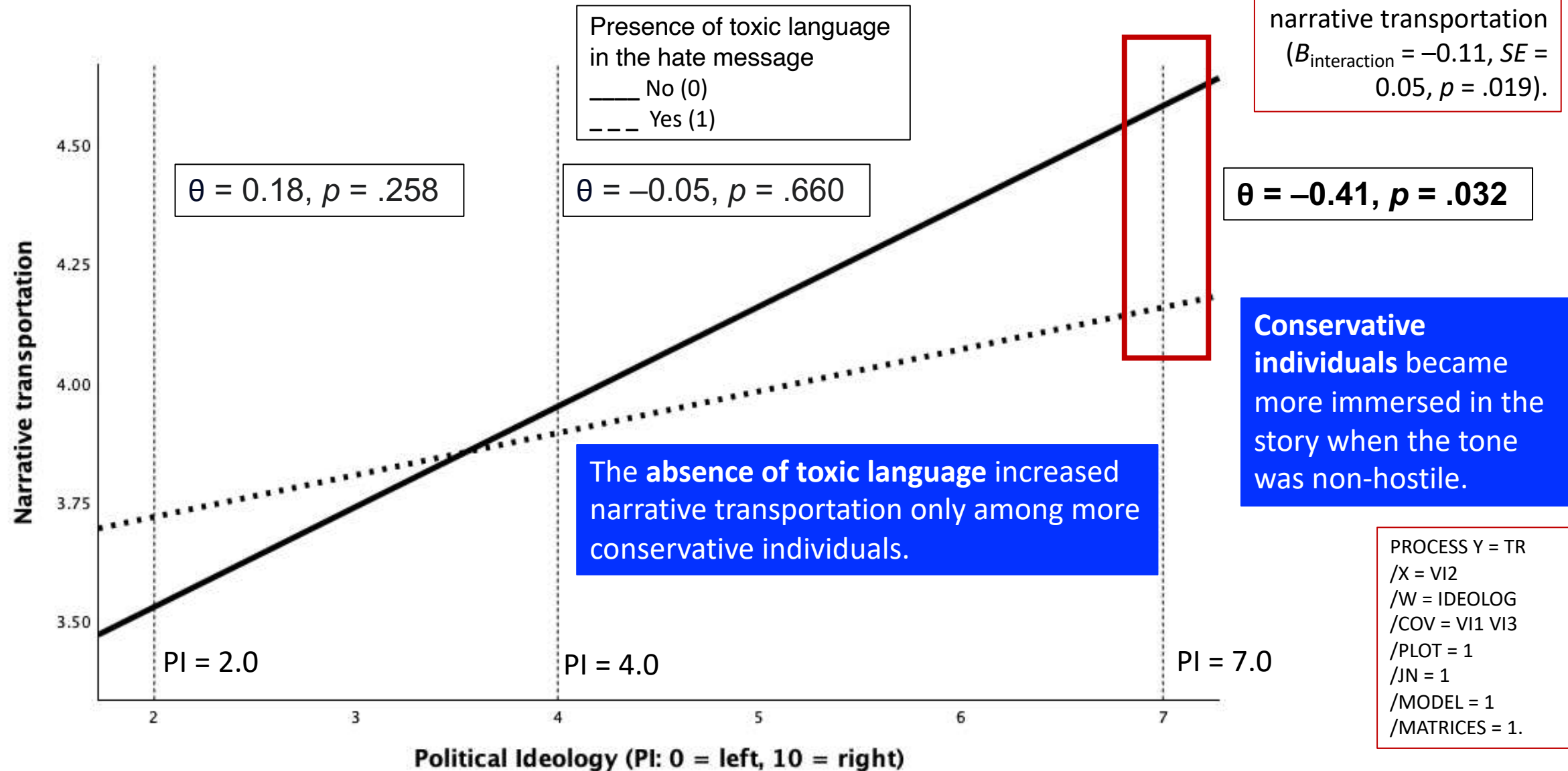
# Measures

- **Fusion with the author of the message**, assessed through the Inclusion of the Other in the Self (IOS) Scale (Gächter et al., 2015; from 1 = low, to 7 = high;  $M = 2.35$ ,  $SD = 1.74$ ).
- **Narrative transportation** (Appel et al., 2015; “I felt highly mentally involved while reading the message”, from 1 = strongly disagree, to 7 = strongly agree;  $\alpha = .70$ ,  $M = 3.93$ ,  $SD = 1.22$ ).
- **Intention to propagate or share the message** (Igartua et al., 2017; “I would be willing to share this message with others”;  $\alpha = .83$ ,  $M = 2.79$ ,  $SD = 1.28$ ).
- **Attitudes toward immigrants and towards refugees** measured using a feeling thermometer (Wojcieszak et al., 2020; from 0 = very cold feelings, to 100 = very warm feelings;  $r[337] = .80$ ,  $p < .001$ ;  $M = 60.68$ ,  $SD = 23.24$ ).
- **Support for harsh policies against irregular immigration** (Bilewicz & Soral, 2020; Saleem et al., 2017; “The government should restrict the entry of additional immigrants into Spain”, from 1 = strongly disagree, to 7 = strongly agree;  $\alpha = .85$ ,  $M = 3.44$ ,  $SD = 1.55$ ).
- **Manipulation check**. The effectiveness of the two experimental manipulations was assessed using a scale composed of six items (e.g., “The message used very aggressive language”, from 1 = strongly disagree, to 7 = strongly agree). In addition, the **emotional impact** of the message was measured by asking participants to indicate the extent to which they experienced specific emotions (e.g., anger, hostility) while reading the message (from 1 = not at all, to 7 = very much). [See results in OSF.](#)



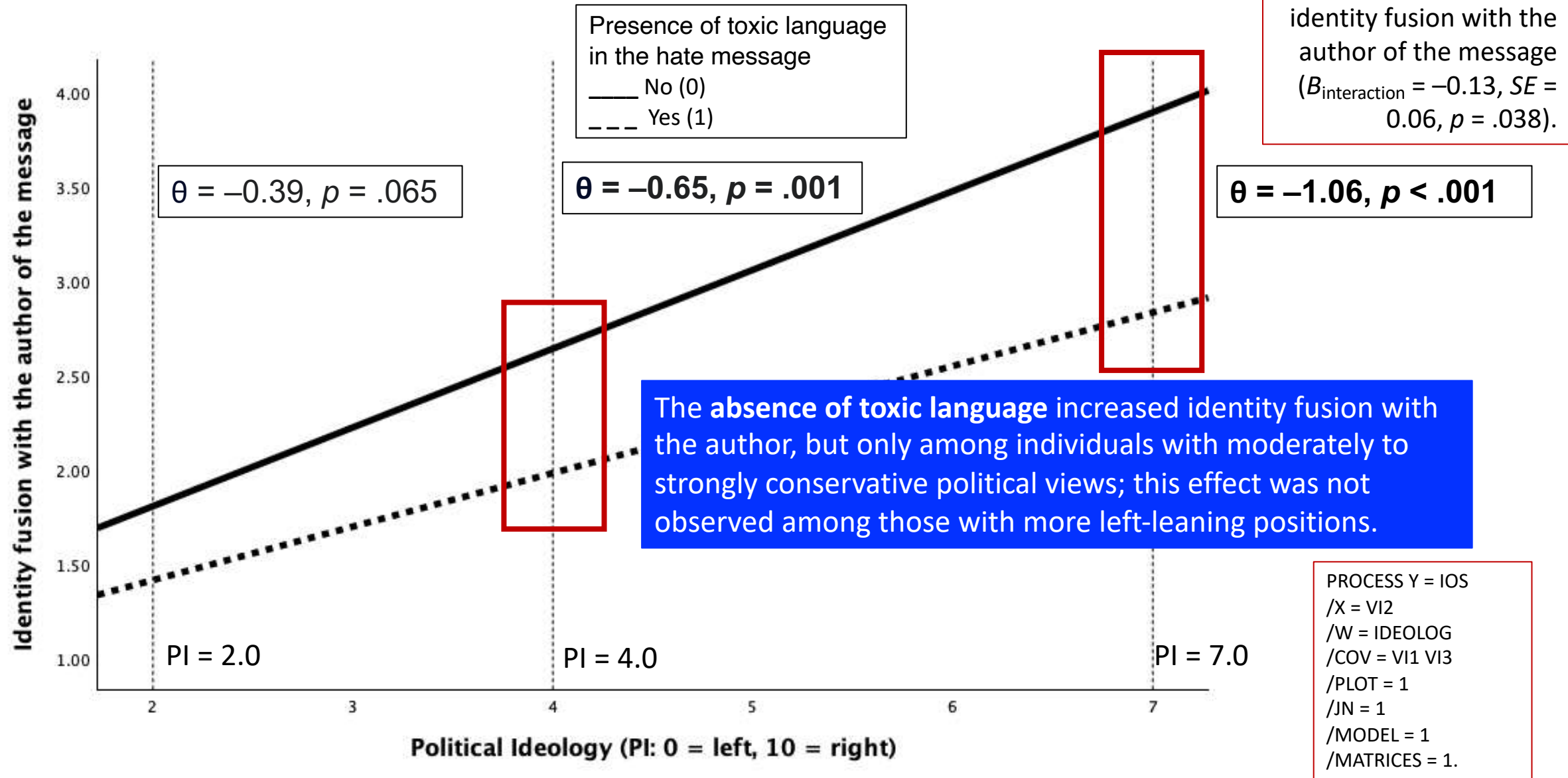
# Results: conditional effects (PROCESS)

Narrative transportation [RQ1]



# Results: conditional effects (PROCESS)

Identity fusion with the author of the message [RQ1]



Number of interactions (*popularity cues*) and political ideology as moderators

Moderation analysis using PROCESS (Model 1) [RQ2]

Moderation analysis using PROCESS (Model 3) [RQ3]



[RQ2] **Number of interactions** (*popularity cues*) did not moderate the effect of toxic language on narrative transportation ( $B_{\text{interaction}} = 0.24, SE = 0.26, p = .365$ ).

[RQ2] **Number of interactions** (*popularity cues*) did not moderate the effect of toxic language on identity fusion with the author of the message ( $B_{\text{interaction}} = 0.48, SE = 0.37, p = .192$ ).

[RQ3] Moreover, the recipient's **political ideology** did not moderate how toxic language and popularity cues shaped narrative transportation ( $B_{\text{three-way interaction}} = -0.13, SE = 0.10, p = .173$ ).

[RQ3] Moreover, the recipient's **political ideology** did not moderate how toxic language and popularity cues shaped identity fusion with the author of the message ( $B_{\text{three-way interaction}} = -0.00, SE = 0.12, p = .967$ ).

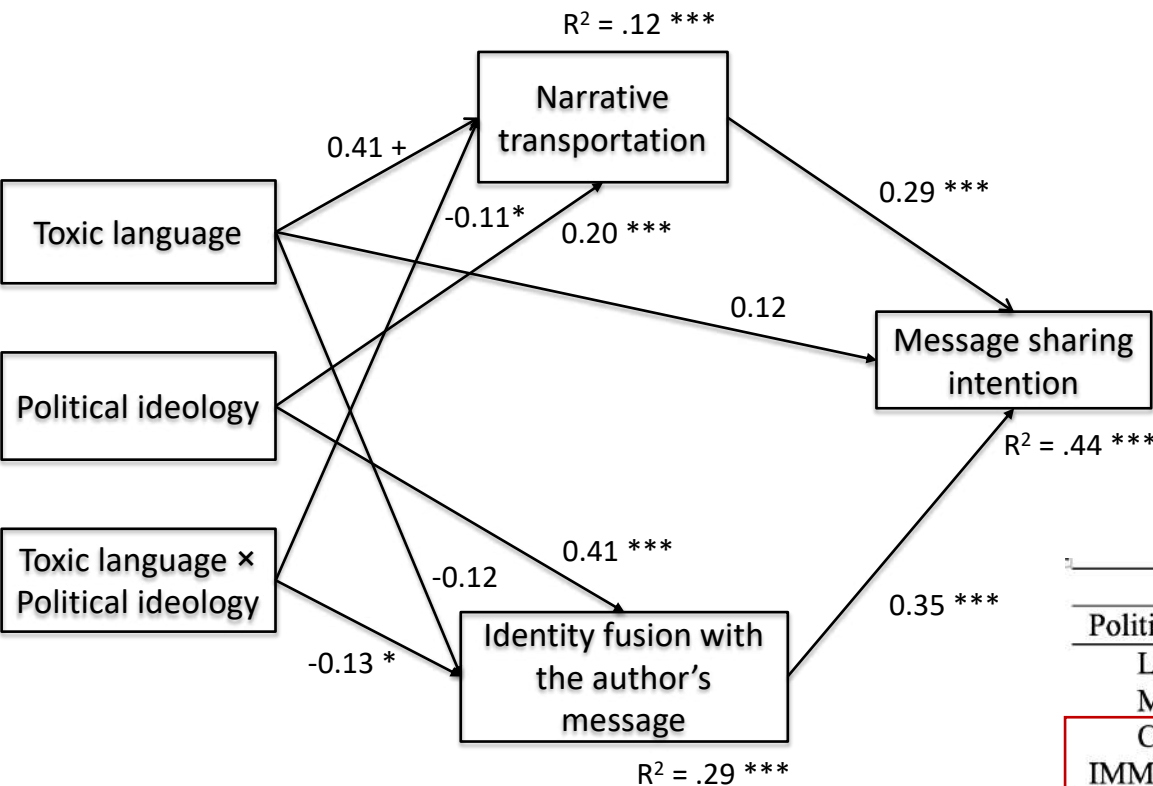
```
PROCESS Y = TR
/X = VI2
/W = VI3
/COV = VI1
/PLOT = 1
/MODEL = 1
/MATRICES = 1.
```

```
PROCESS Y = TR
/X = VI2
/W = VI3
/Z = IDEOLOG
/COV = VI1
/PLOT = 1
/JN = 1
/MODEL = 3
/MATRICES = 1.
```

The experimental condition (presence of toxic language in the hate message) was set up as a dummy variable (0 = without toxic language, 1 = with toxic language). The number of interactions associated with the post was also set up as a dummy variable (0 = low, 1 = high). Identity fusion with the author of the message (1 = low, 7 = high). Political ideology was assessed on an 11-point scale (from 0 = left to 10 = right). The type of threat (0 = economic threat, 1 = security threat) was included as a covariate.

# Full moderated mediation model [RQ4]

Dependent variable: Message sharing intention



## Model 7

Correlation between identity fusion and narrative transportation:  
 $r(337) = .49, p < .001$

PROCESS Y = COMP  
/X = VI2  
/M = TR IOS  
/W = IDEOLOG  
/COV = VI1 VI3  
/MODEL = 7  
/BOOT = 10000  
/MATRICES = 1  
/SEED = 21102024.

Toxic language → Narrative transportation → Message sharing intention				
Political ideology	Values	Indirect effect (SE)	Boot LLCI	Boot ULCI
Liberal	2.00	0.05 (0.05)	-0.045	0.159
Moderate	4.00	-0.01 (0.03)	-0.093	0.056
Conservative	7.00	-0.12 (0.06)	-0.255	-0.005
IMM = -0.03 (0.01) [95% CI: -0.071, -0.003]				

Toxic language → Identity fusion → Message sharing intention				
Political ideology	Values	Indirect effect (SE)	Boot LLCI	Boot ULCI
Liberal	2.00	-0.14 (0.06)	-0.284	-0.012
Moderate	4.00	-0.23 (0.06)	-0.372	-0.119
Conservative	7.00	-0.38 (0.12)	-0.631	-0.156
IMM = -0.04 (0.02) [95% CI: -0.103, 0.004]				

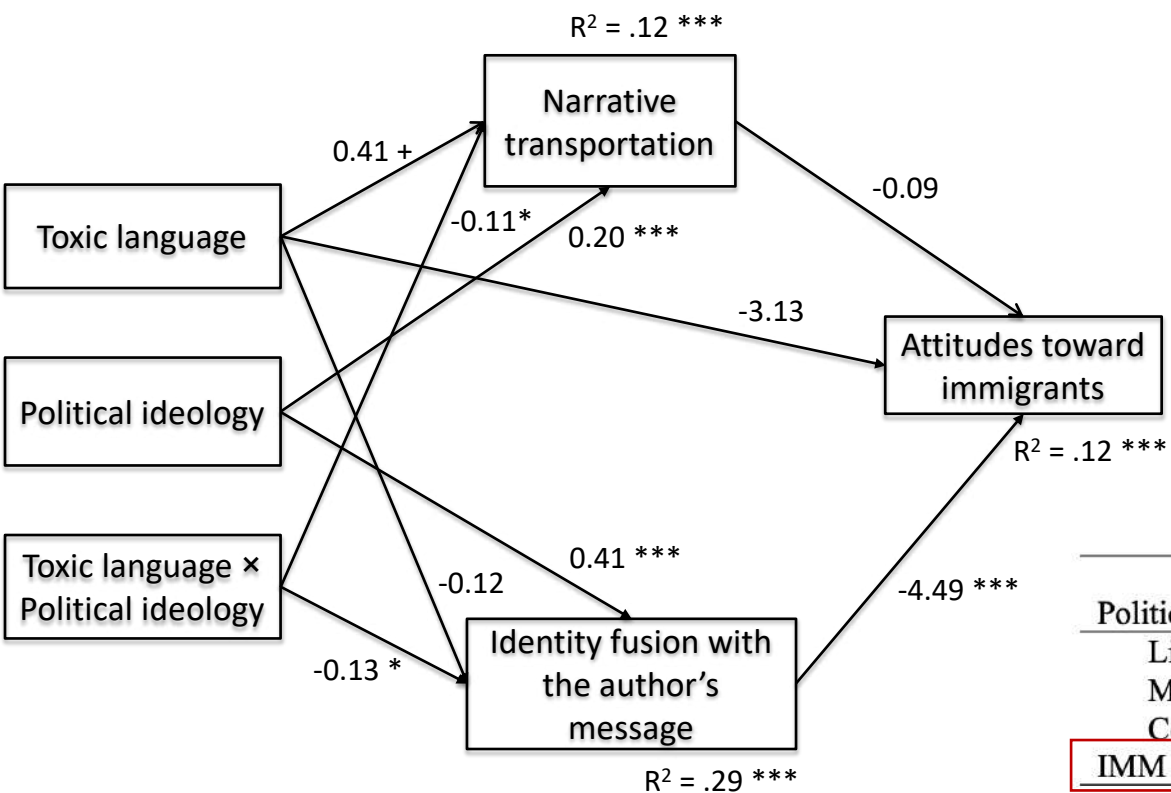
Conditional indirect effects of toxic language (0 = low, 1 = high) on message sharing intention through narrative transportation and identity fusion with the author of the message.

# Full moderated mediation model [RQ4]

Dependent variable: Attitudes toward immigrants

Model 7

PROCESS Y = ACTINREF  
/X = VI2  
/M = TR IOS  
/W = IDEOLOG  
/COV = VI1 VI3  
/MODEL = 7  
/BOOT = 10000  
/MATRICES = 1  
/SEED = 21102024.



Correlation between identity fusion and narrative transportation:  $r(337) = .49, p < .001$

Toxic language → Narrative transportation → Attitudes toward immigrants				
Political ideology	Values	Indirect effect (SE)	Boot LLCI	Boot ULCI
Liberal	2.00	-0.01 (0.28)	-0.696	0.554
Moderate	4.00	0.00 (0.16)	-0.355	0.339
Conservative	7.00	0.03 (0.53)	-1.070	1.197
IMM = 0.01 (0.15) [95% CI: -0.292, 0.341]				

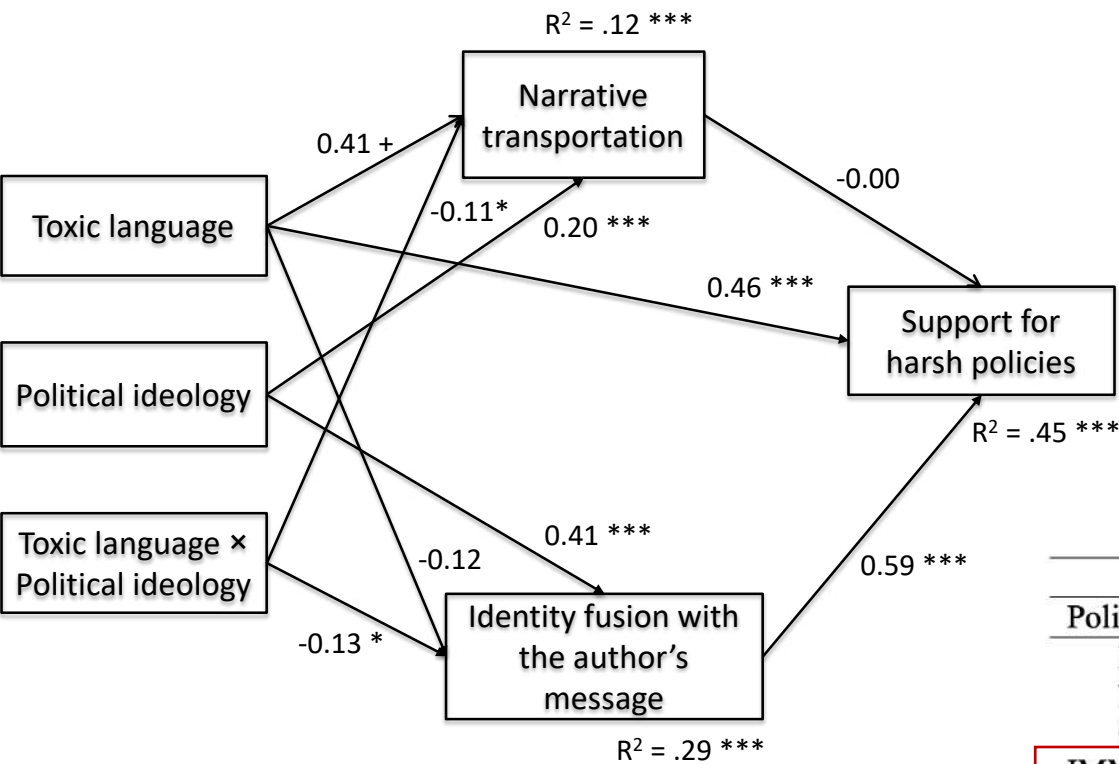
Toxic language → Identity fusion → Attitudes toward immigrants				
Political ideology	Values	Indirect effect (SE)	Boot LLCI	Boot ULCI
Liberal	2.00	1.78 (0.94)	0.148	3.777
Moderate	4.00	2.99 (0.94)	1.328	5.023
Conservative	7.00	4.80 (1.64)	1.860	8.329
IMM = 0.60 (0.34) [95% CI: -0.063, 1.319]				

Conditional indirect effects of toxic language (0 = low, 1 = high) on attitudes toward immigrants through narrative transportation and identity fusion with the author of the message.



# Full moderated mediation model [RQ4]

Dependent variable: Support for harsh policies against irregular immigration



## Model 7

Correlation between identity fusion and narrative transportation:  
 $r(337) = .49, p < .001$

PROCESS Y = PI  
/X = VI2  
/M = TR IOS  
/W = IDEOLOG  
/COV = VI1 VI3  
/MODEL = 7  
/BOOT = 10000  
/MATRICES = 1  
/SEED = 21102024.

Conditional indirect effects of toxic language (0 = low, 1 = high) on support for harsh policies against irregular immigration through narrative transportation and identity fusion with the author of the message.

Toxic language → Narrative transportation → Support for harsh policies				
Political ideology	Values	Indirect effect (SE)	Boot LLCI	Boot ULCI
Liberal	2.00	-0.00 (0.01)	-0.032	0.030
Moderate	4.00	0.00 (0.00)	-0.018	0.016
Conservative	7.00	0.00 (0.02)	-0.056	0.054
IMM = 0.00 (0.00) [95% CI: -0.015, 0.015]				

Toxic language → Identity fusion → Support for harsh policies				
Political ideology	MR	Indirect effect (SE)	Boot LLCI	Boot ULCI
Liberal	2.00	-0.23 (0.11)	-0.470	-0.021
Moderate	4.00	-0.39 (0.10)	-0.602	-0.208
Conservative	7.00	-0.63 (0.18)	-1.007	-0.280
IMM = -0.08 (0.04) [95% CI: -0.166, 0.008]				

# Conclusions and Discussion

- Hate messages using **less toxic language** can be **more persuasive** by fostering identity fusion with the author of the message.
- **Identity fusion** drives sharing, anti-immigrant attitudes, and support for harsh policies, regardless of ideology.
- **Narrative transportation** increases sharing mainly among conservative individuals (who are generally more critical of immigration; Davidov et al., 2020).
- Findings align with **Social Judgment Theory** (Dal Cin et al., 2004; Perloff, 2017), which posits that people evaluate persuasive messages based on their pre-existing attitudes. For conservative individuals, non-toxic hate narratives may fall within their ***latitude of acceptance***, thereby increasing receptivity and reducing resistance.
- **Popularity cues have little impact** compared to message content and ideology (Sundar, 2008).
- The **THREAD model** provides a robust framework for understanding online hate's psychological impact.

# Thank you for your attention!

Juan-José IGARTUA  
University of Salamanca ([jigartua@usal.es](mailto:jigartua@usal.es))

Carlos Ballesteros  
University of Valladolid

All **materials** related to the research, including the pilot and main study instruments, datasets, syntax files, testimonial messages, and Electronic Supplementary Material (ESM), are available through the Open Science Framework (OSF):

