A Stylometric Inquiry into Hyperpartisan and Fake News

Martin Potthast*, **Johannes Kiesel**†, Kevin Reinartz†, Janek Bevendorff†, Benno Stein†
*Leipzig University, †Bauhaus-Universität Weimar

webis.de

ACL, July 16th, 2018


oh!
Fake news is not our friend.
We're committed to reducing its spread; so we're working with more fact-checkers globally, improving our technology, and giving you background information on the articles in your News Feed.
Find out more: fb.me/changesAU
facebook
Melbourne Weath
TODAY
14°
Exit

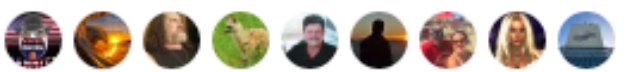
Donald J. Trump
@realDonaldTrump

Follow

....it is very possible that those sources don't exist but are made up by fake news writers. #FakeNews is the enemy!

5:45 AM - 28 May 2017

12,418 Retweets **56,686** Likes



19K 12K 57K

What are Fake News?

Disinformation displayed as news articles

What are Fake News?

Disinformation displayed as news articles

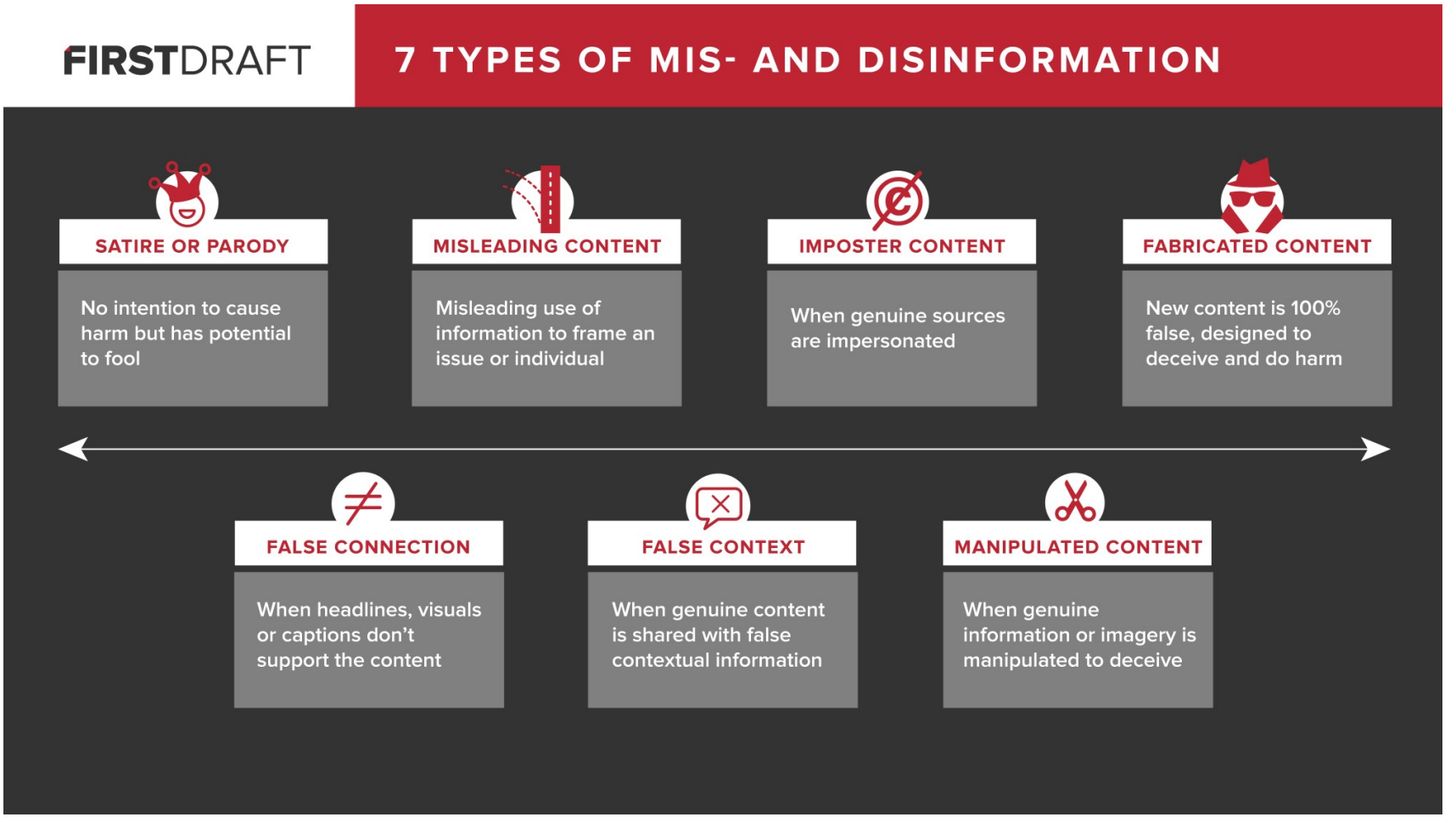


Image: Claire Wardle, First Draft

What are Fake News?

Disinformation displayed as news articles

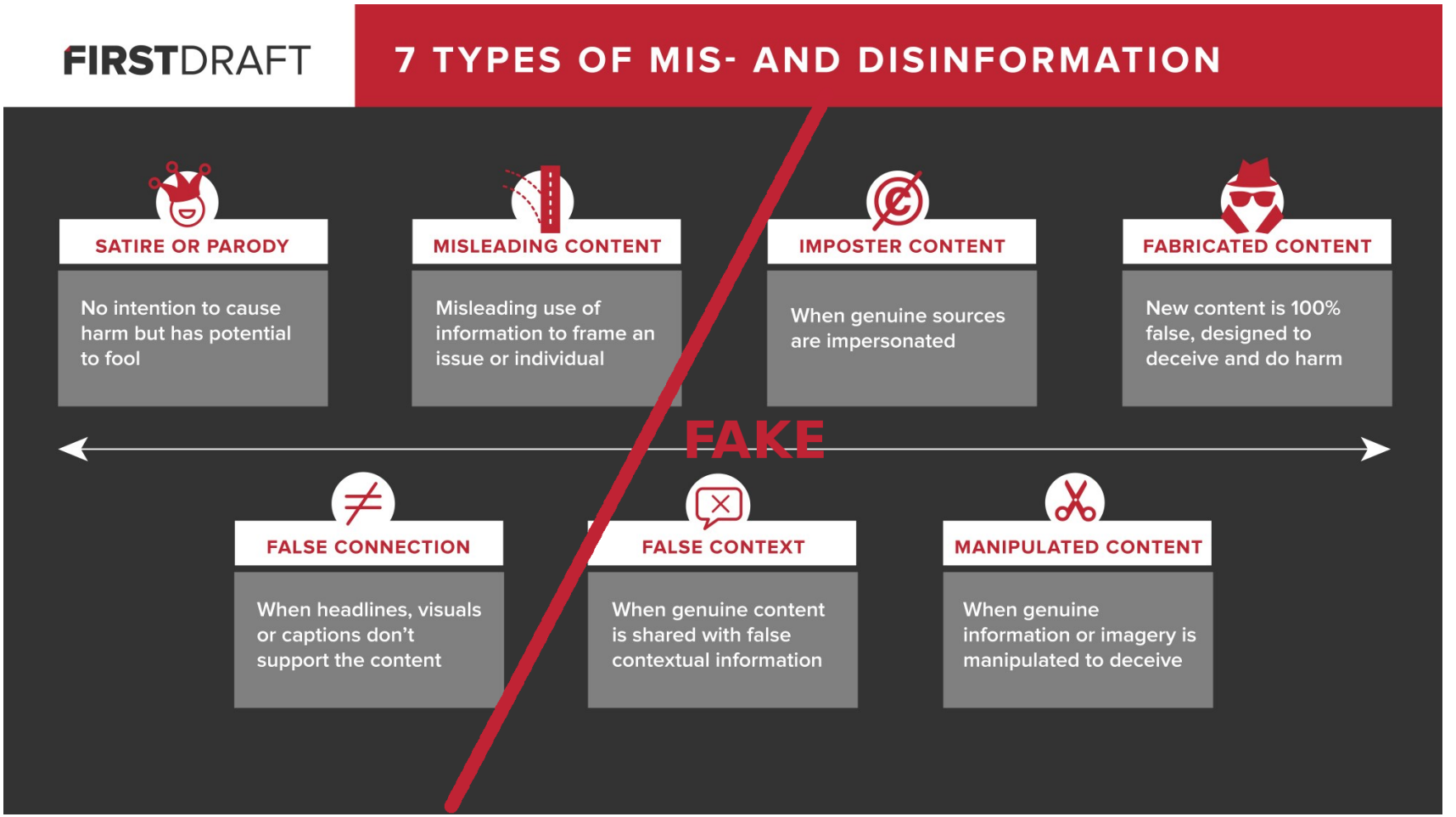


Image: Claire Wardle, First Draft

A Stylometric Inquiry into Hyperpartisan “News” and “News” in False Context and/or with Content that is Impostered, Manipulated, and/or Fabricated

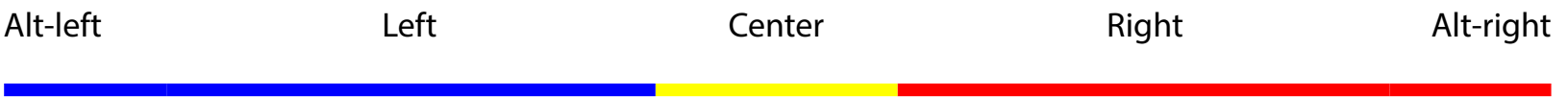
Martin Potthast*, **Johannes Kiesel**†, Kevin Reinartz†, Janek Bevendorff†, Benno Stein†
*Leipzig University, †Bauhaus-Universität Weimar

webis.de

ACL, July 16th, 2018

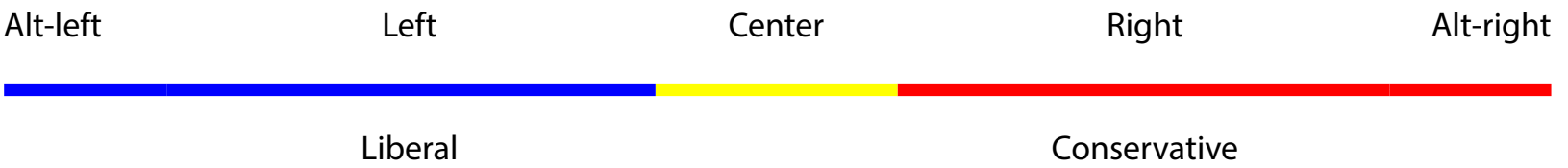
The Political Spectrum

The left-right political spectrum is a system of classifying political positions, ideologies and parties. Left-wing politics and right-wing politics are often presented as opposed, although either may adopt stances from the other side. [Wikipedia]



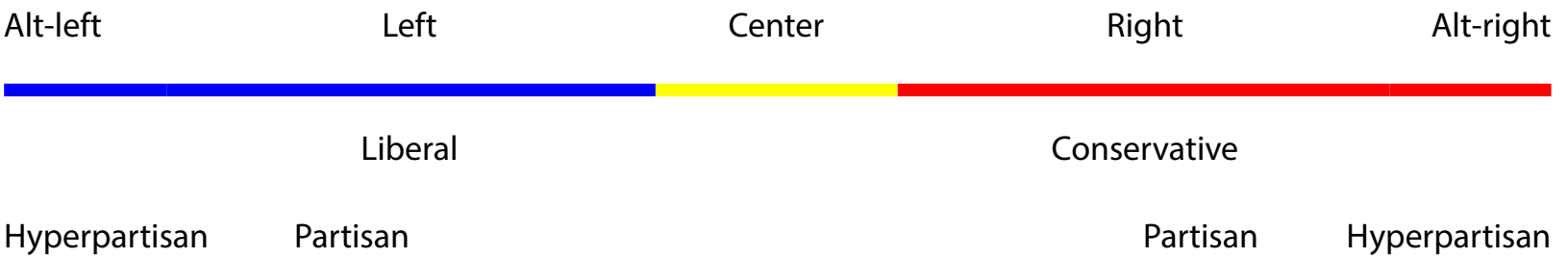
The Political Spectrum

The left-right political spectrum is a system of classifying political positions, ideologies and parties. Left-wing politics and right-wing politics are often presented as opposed, although either may adopt stances from the other side. [Wikipedia]



The Political Spectrum

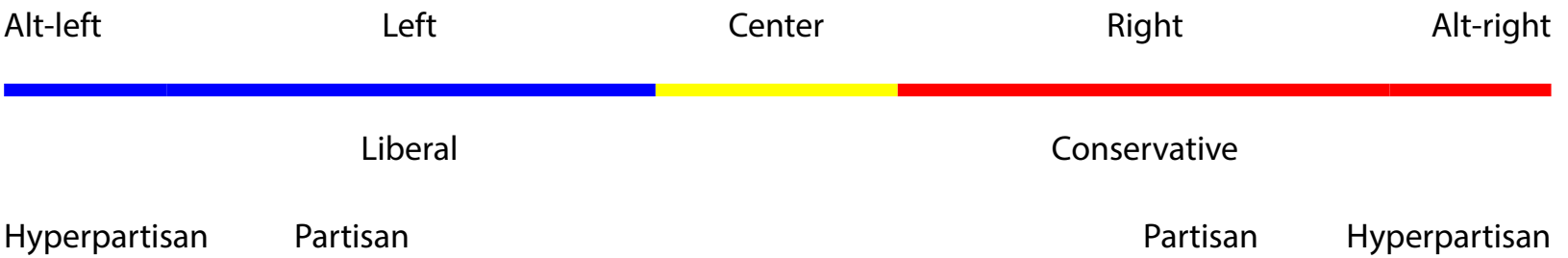
The left-right political spectrum is a system of classifying political positions, ideologies and parties. Left-wing politics and right-wing politics are often presented as opposed, although either may adopt stances from the other side. [Wikipedia]



Partisan: someone with a psychological identification with one major party. [Wikipedia]

The Political Spectrum

The left-right political spectrum is a system of classifying political positions, ideologies and parties. Left-wing politics and right-wing politics are often presented as opposed, although either may adopt stances from the other side. [Wikipedia]



Partisan: someone with a psychological identification with one major party. [Wikipedia]

News media reporting on politics can be aligned on this spectrum as well.

We are observing an increasing number of hyperpartisan news publishers.

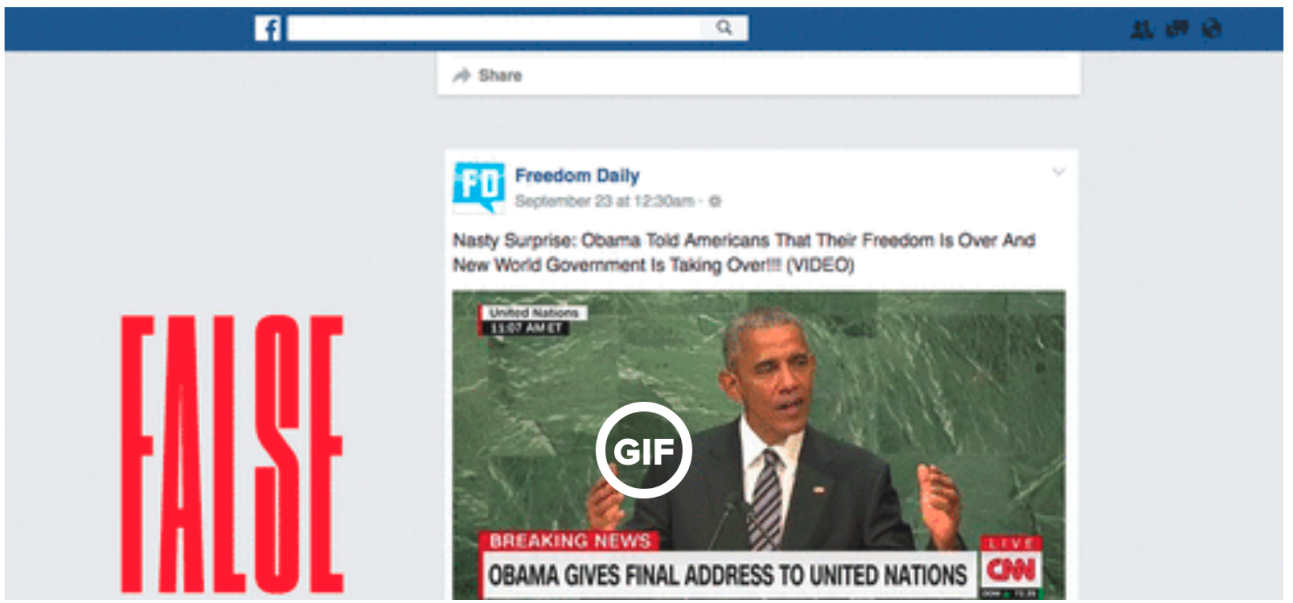
Fake News and Hyperpartisan News

BuzzFeed NEWS / REPORTING TO YOU

Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate

A BuzzFeed News analysis found that three big right-wing Facebook pages published false or misleading information 38% of the time during the period analyzed, and three large left-wing pages did so in nearly 20% of posts.

By Craig Silverman (BuzzFeed Founding Editor, Canada), Lauren Strapagiel (BuzzFeed Staff), Hamza Shaban (BuzzFeed News Reporter), Ellie Hall (BuzzFeed News Reporter), Jeremy Singer-Vine (BuzzFeed News Reporter)



Why are Fake News Published by Hyperpartisan Pages?

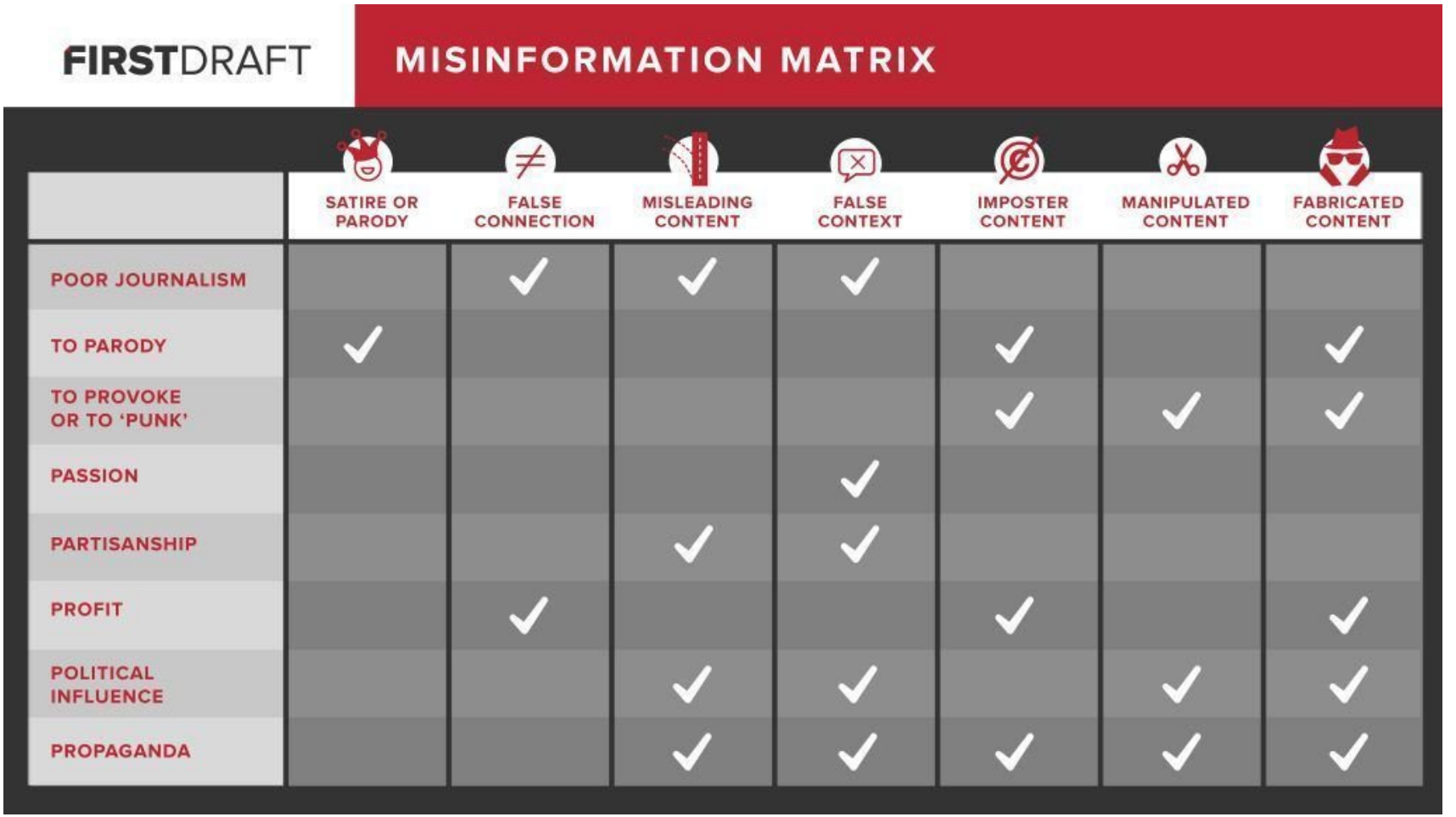


Image: Claire Wardle, First Draft

Why are Fake News Published by Hyperpartisan Pages?

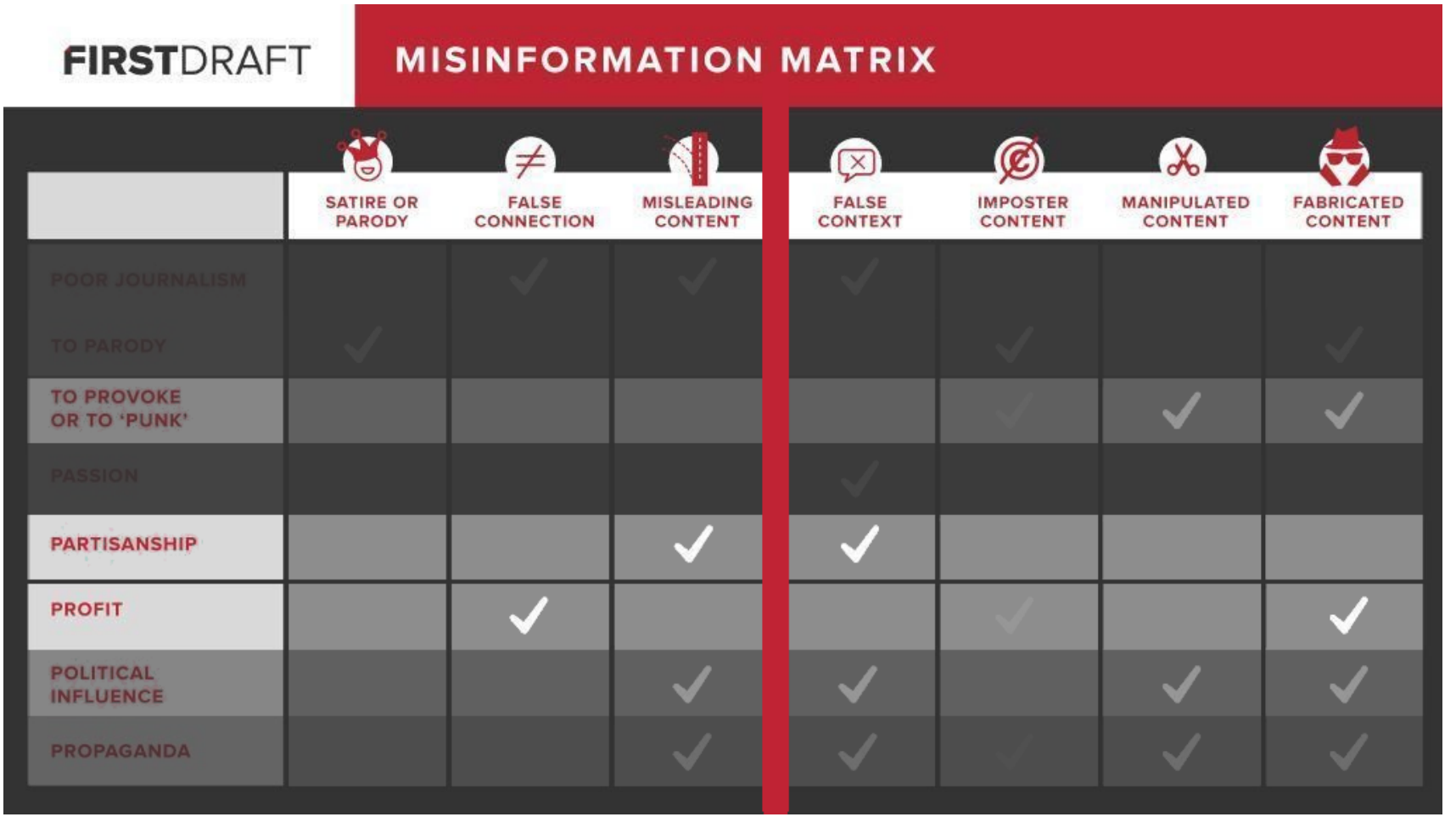


Image: Claire Wardle, First Draft

Fake News Detection

Knowledge-based

- Requires political knowledge base
- Unavailable ahead of time
- We cannot trust the web

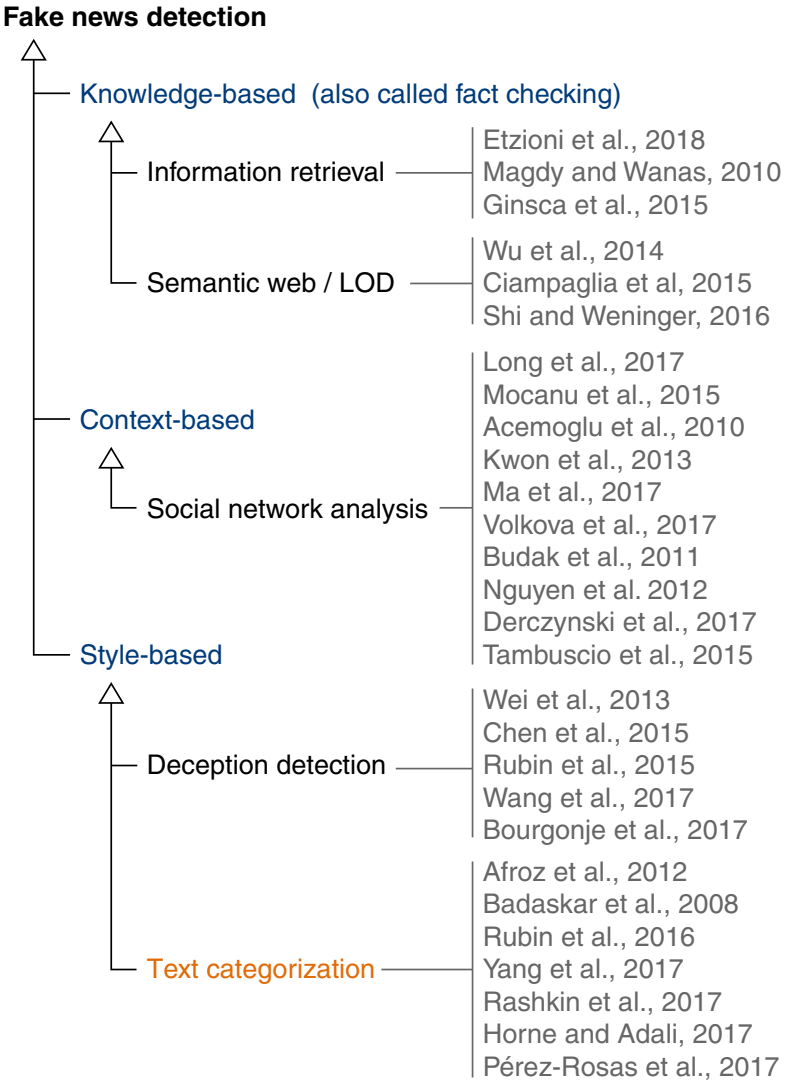
Context-based

- Limited to social media platforms
- Part of damage already done

Style-based

- Allows for pre-posting check
- Real-time reaction possible
- Hard to mask
- But are style differences sufficient?

Taxonomy of Approaches



Fake News and Hyperpartisan News Corpus Construction

Fake News and Hyperpartisan News Corpus Construction

Orientation Publisher	**Fact-checking results**				
	true	mix	false	n/a	Σ
Center	806	8	0	12	826
ABC News	90	2	0	3	95
CNN	295	4	0	8	307
Politico	421	2	0	1	424
Left-wing	182	51	15	8	256
Addicting Info	95	25	8	7	135
Occupy Democrats	59	25	7	0	91
The Other 98%	28	1	0	1	30
Right-wing	276	153	72	44	545
Eagle Rising	106	47	25	36	214
Freedom Daily	49	24	22	4	99
Right Wing News	121	82	25	4	232
Σ	1264	212	87	64	1627

Annotations provided by journalists at BuzzFeed

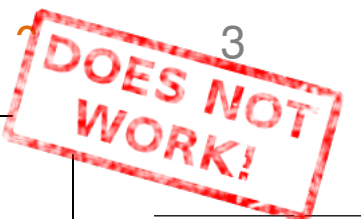
Fake News and Hyperpartisan News Selected Results

Orientation / Publisher	Fact-checking results: true	mix	false	n/a	Σ
Center	806	8	0	12	826
ABC News				3	95
CNN					307
Politico					424
Left-wing				8	256
Addicting Info				7	135
Occupy Democr				0	91
The Other 98%				1	30
Right-wing	276	153	72	44	545
Eagle Rising	106	47	25	36	214
Freedom Daily	49	24	22	4	99
Right Wing News	121	82	25	4	232
Σ	1264	212	87	64	1627

Fake News Detection

Precision ≈ 42%

Recall ≈ 41%



Annotations provided by journalists at BuzzFeed

Fake News and Hyperpartisan News Selected Results

Orientation Publisher	**Fact-checking results** true	mix	false	n/a	Σ
Center	806	8	0	12	826
ABC N					95
CNN					307
Politic					424
Left-w					256
Addict					135
Occup					91
The O					30
Right-wing	276	153	72	44	545
Eagle Rising	106	47	25	36	214
Freedom Daily	49	24	22	4	99
Right Wing News	121	82	25	4	232
Σ	1264	212	87	64	1627

Orientation Detection

Precision ≈ 21% Precision ≈ 56%

Recall ≈ 20% Recall ≈ 59%

Annotations provided by journalists at BuzzFeed

Fake News and Hyperpartisan News Selected Results

Orientation Publisher	**Fact-checking results** true	mix	false	n/a	Σ
Center	806	8	0	12	826
ABC News					95
CNN					307
Politico					424
Left-wing					256
Addicting In					135
Occupy Der					91
The Other 9					30
Right-wing	276	153	72	44	545
Eagle Rising	106	47	25	36	214
Freedom Daily	49	24	22	4	99
Right Wing News	121	82	25	4	232
Σ	1264	212	87	64	1627

Hyperpartisanship Detection

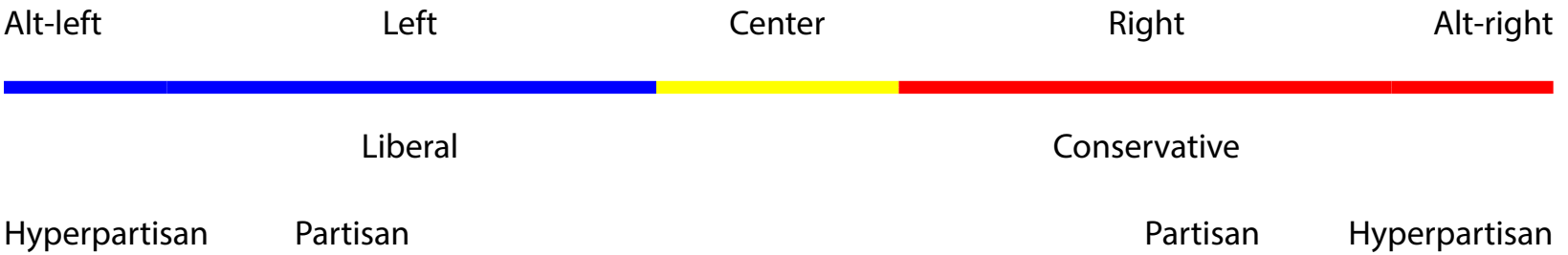
Precision ≈ 69%

Recall ≈ 89%

Annotations provided by journalists at BuzzFeed

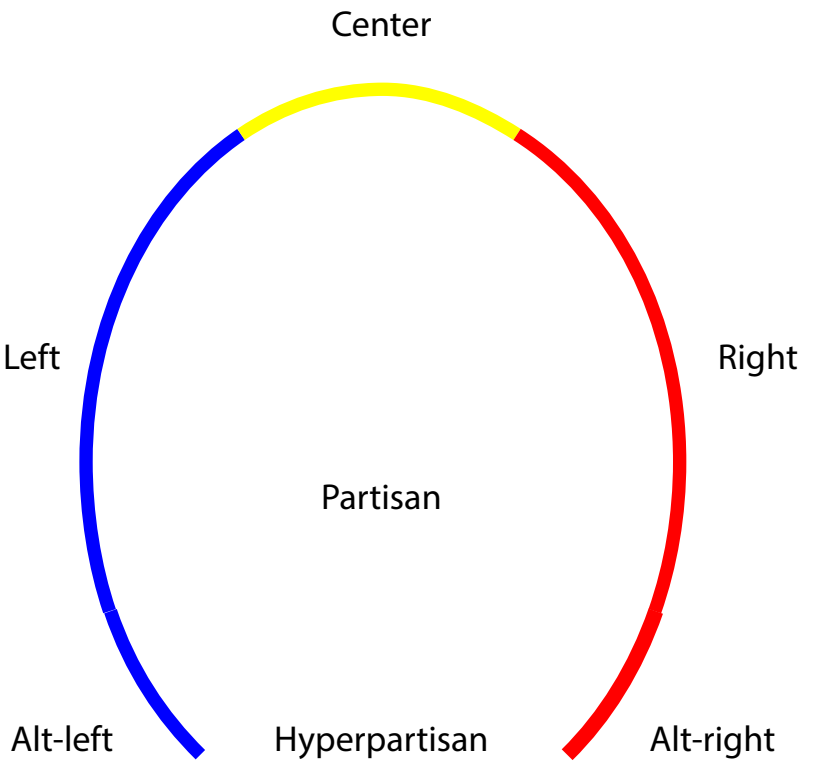
Fake News and Hyperpartisan News

How can it be that the alt left and the alt right cannot be distinguished from the mainstream, when both together (hyperpartisan news) can be?



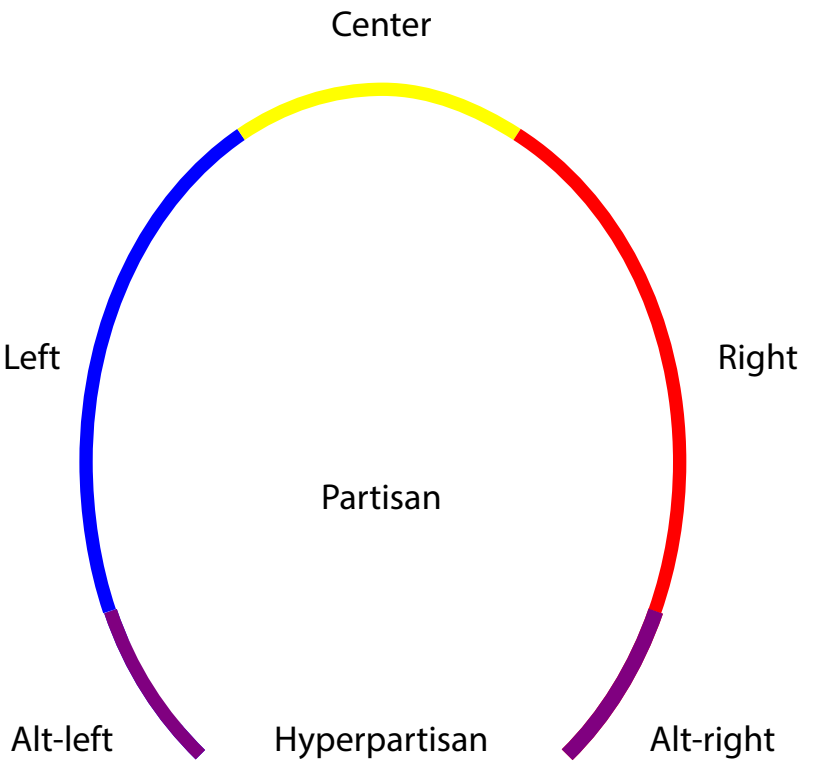
Fake News and Hyperpartisan News

How can it be that the alt left and the alt right cannot be distinguished from the mainstream, when both together (hyperpartisan news) can be?



Fake News and Hyperpartisan News

How can it be that the alt left and the alt right cannot be distinguished from the mainstream, when both together (hyperpartisan news) can be?



The horseshoe theory asserts that the alt left and the alt right, rather than being at opposite and opposing ends of a linear political continuum, in fact closely resemble one another, much like the ends of a horseshoe. [Wikipedia]

Horseshoe Validation Experiment I Leave-out Classification

left-wing

center

right-wing

Horseshoe Validation Experiment I Leave-out Classification

left-wing | center | right-wing

- Classifier is trained to distinguish left-wing and center articles
- Right-wing articles are used for testing

Horseshoe Validation Experiment I Leave-out Classification

left-wing

center

right-wing

74% | 26%

- Classifier is trained to distinguish left-wing and center articles
- Right-wing articles are used for testing
- Majority of right-wing articles are classified as left-wing rather than center

Horseshoe Validation Experiment I Leave-out Classification

left-wing

center

right-wing

74% | 26%

- Classifier is trained to distinguish left-wing and center articles
- Right-wing articles are used for testing
- Majority of right-wing articles are classified as left-wing rather than center

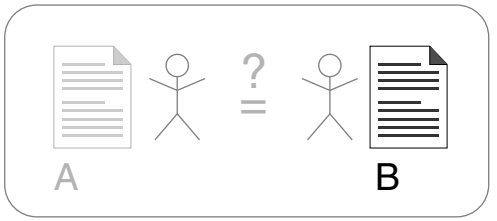
left-wing

34% | **66%**

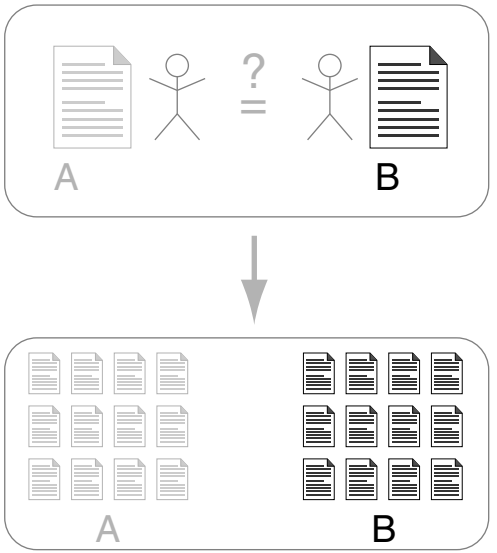
center

right-wing

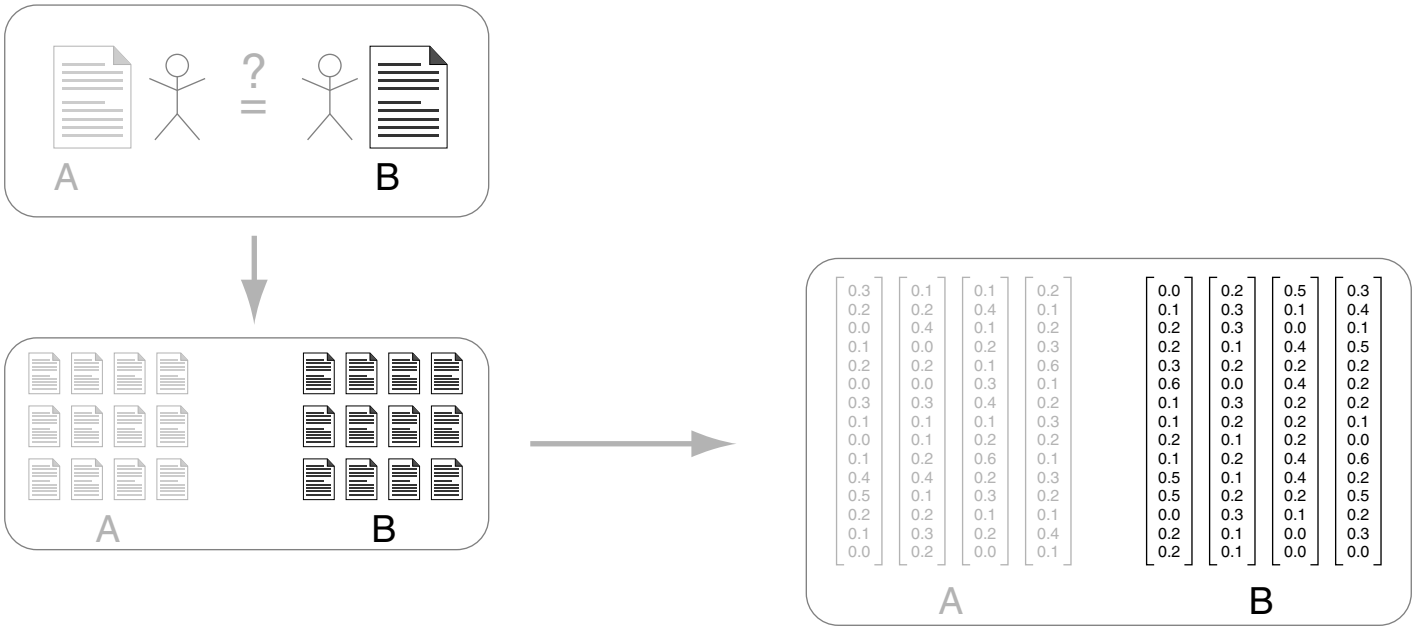
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



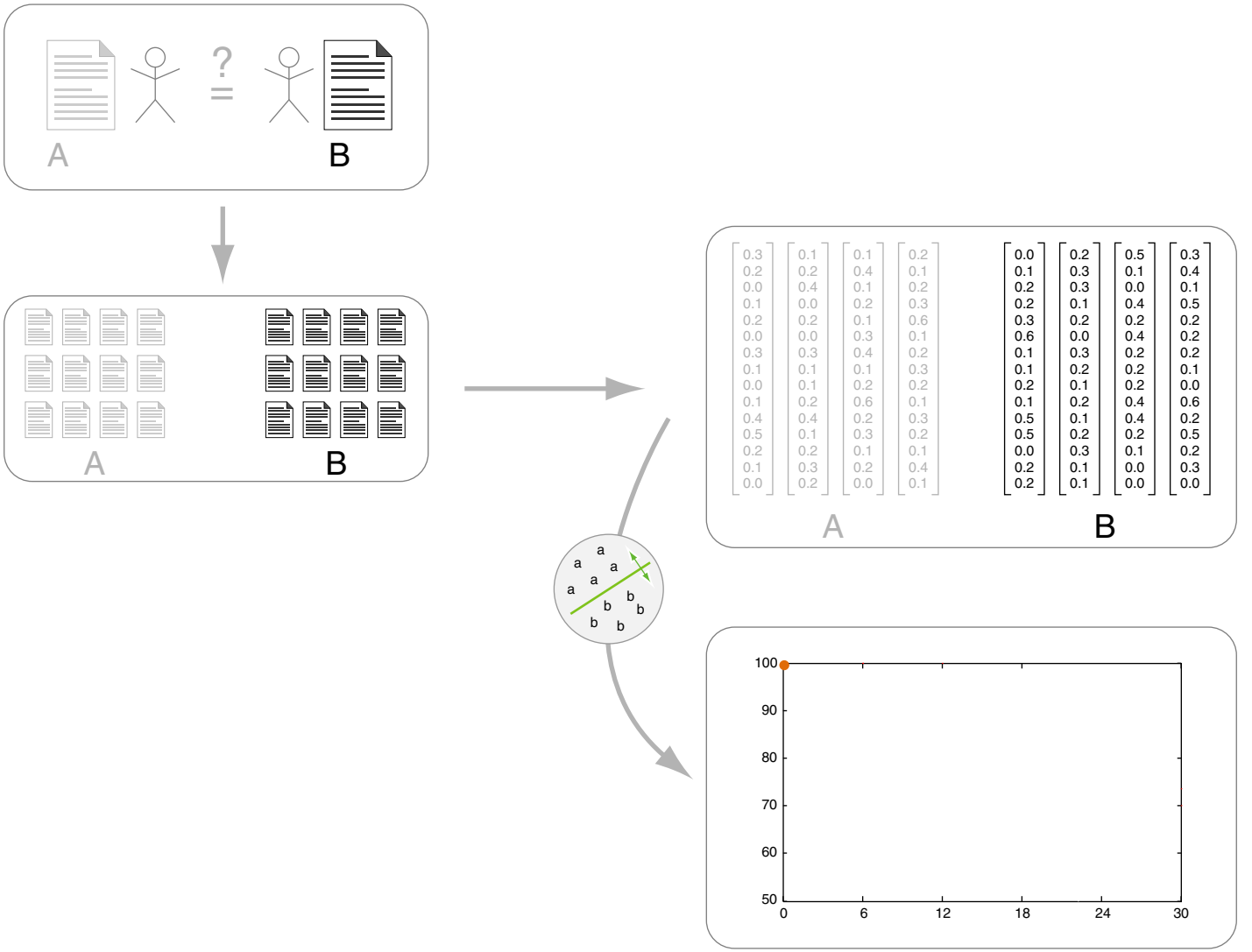
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



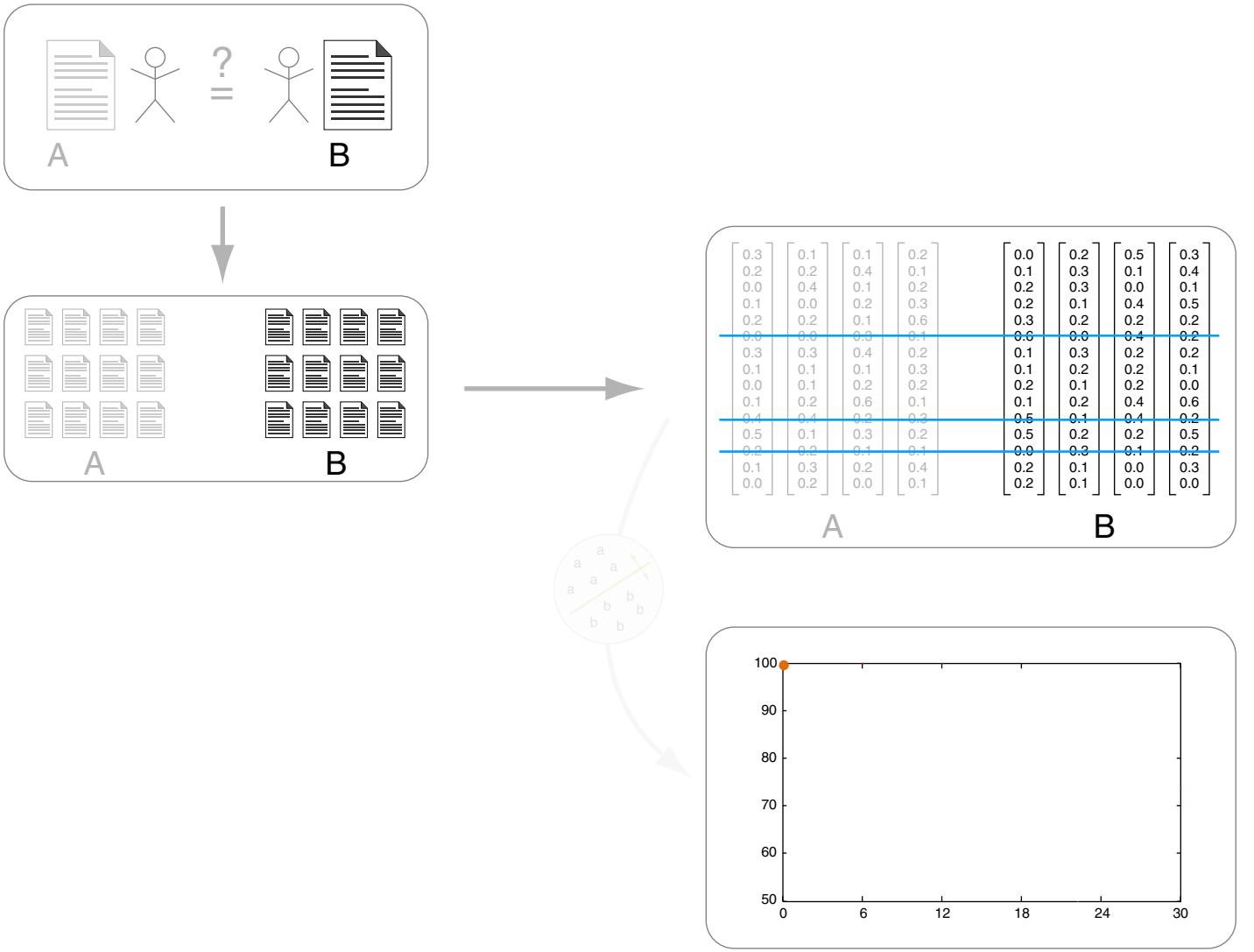
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



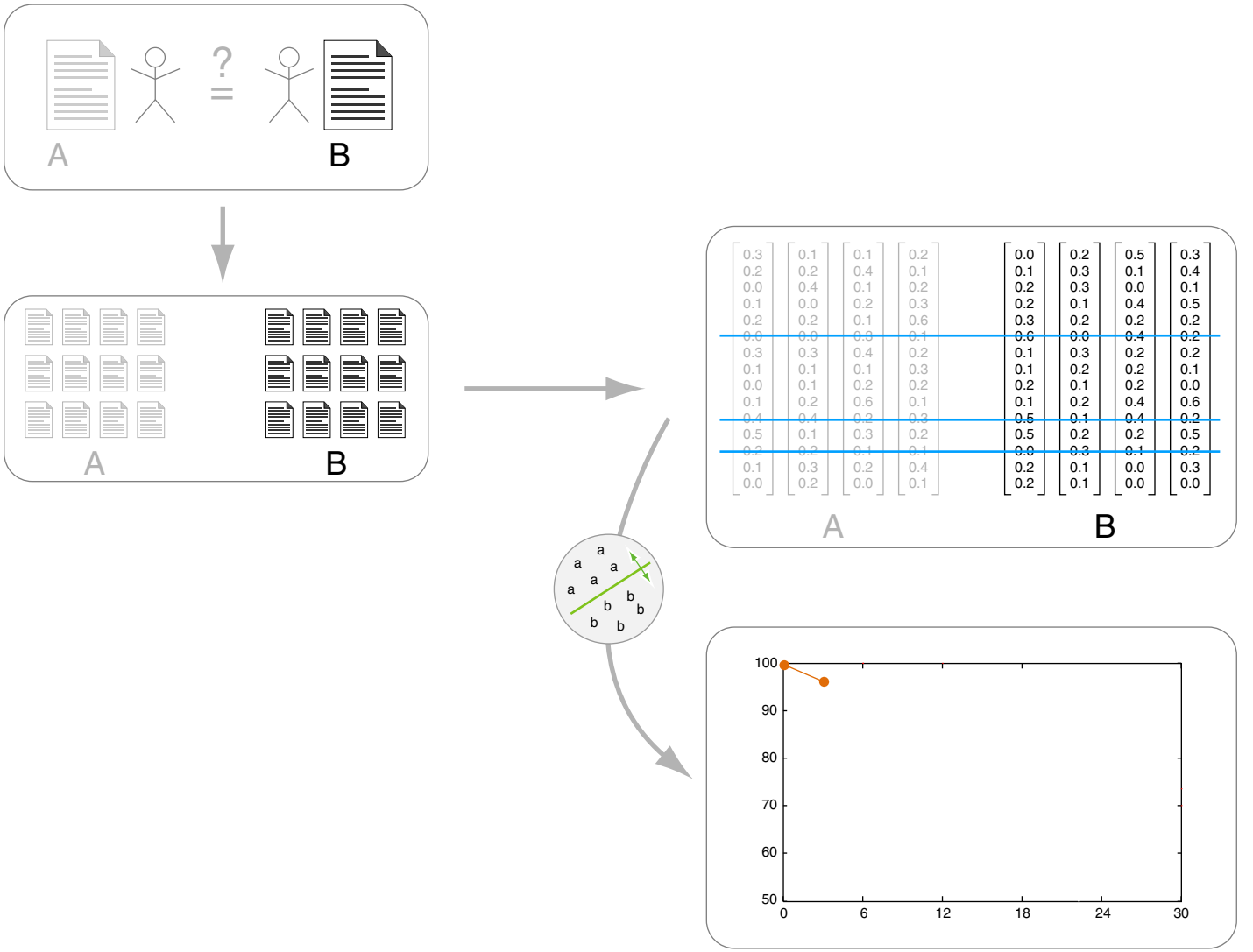
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



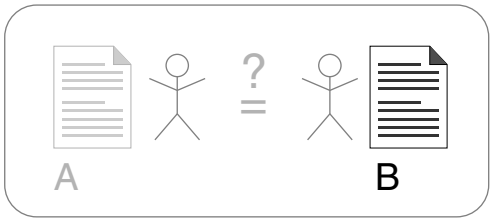
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]

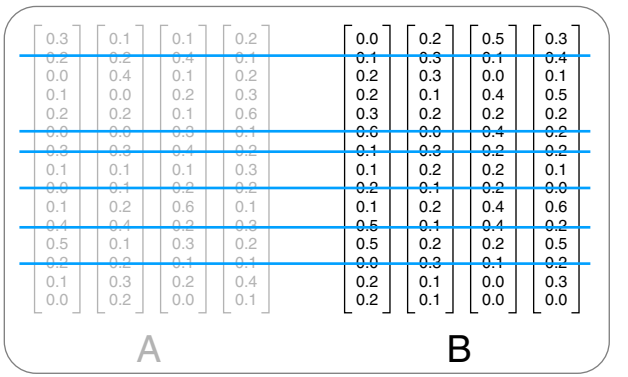


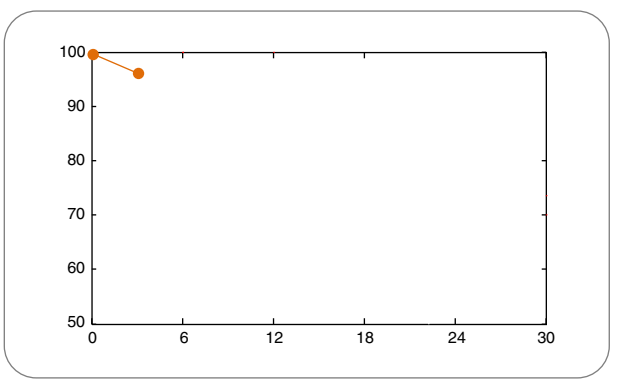
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



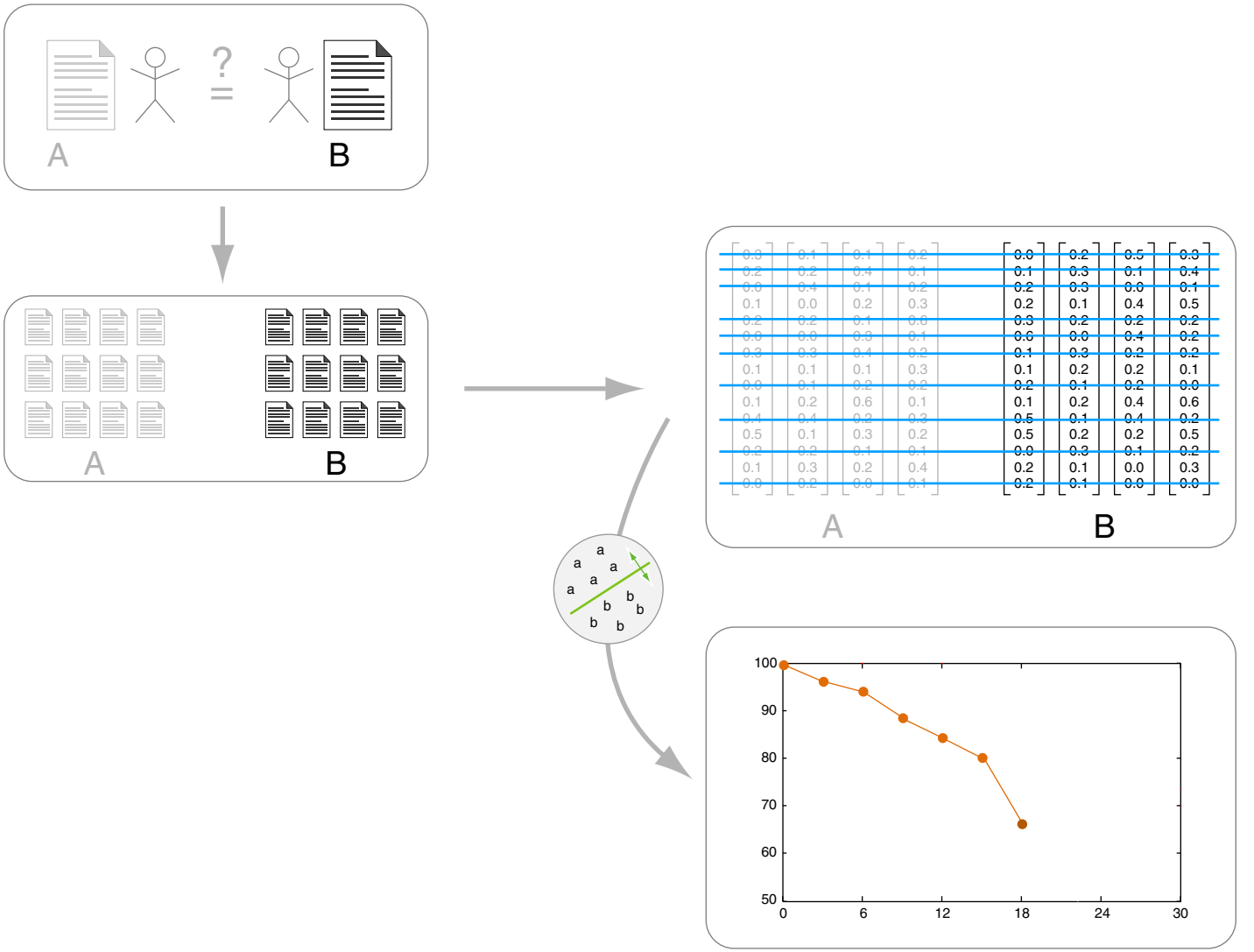
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]





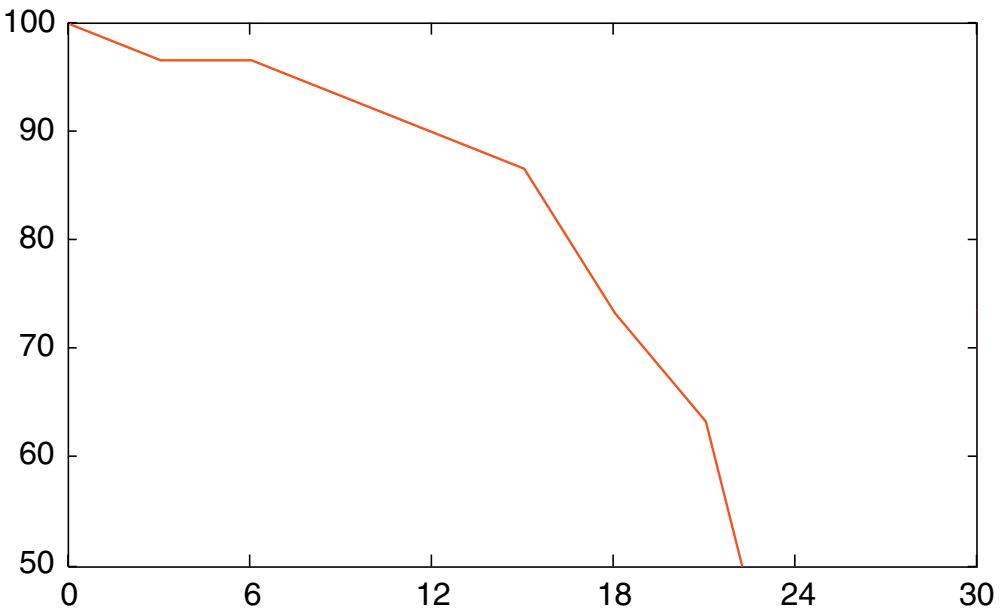


Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]

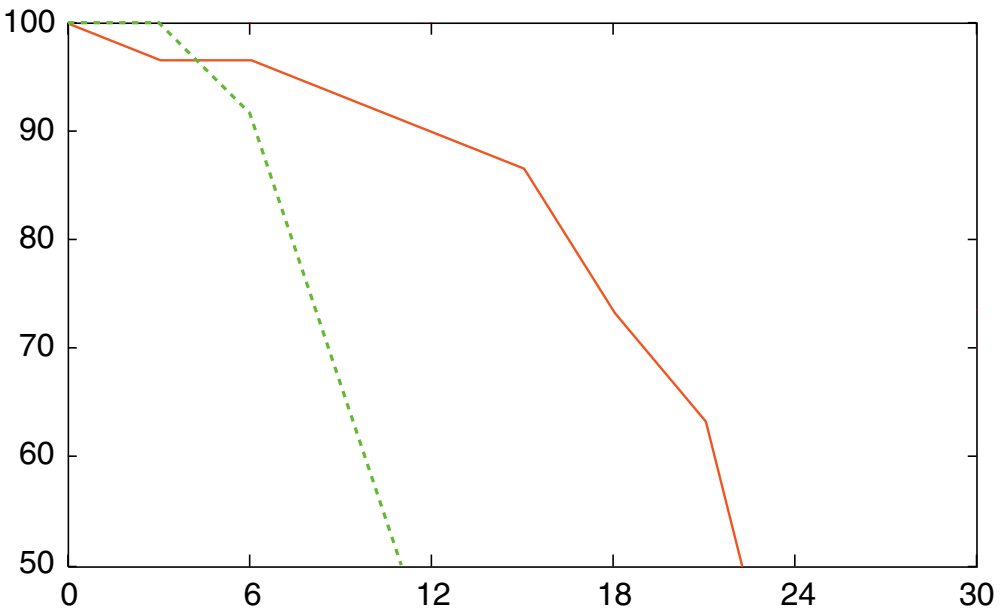
Typical learning characteristic for . . .



different authors ($A \neq B$)

Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]

Typical learning characteristic for . . .

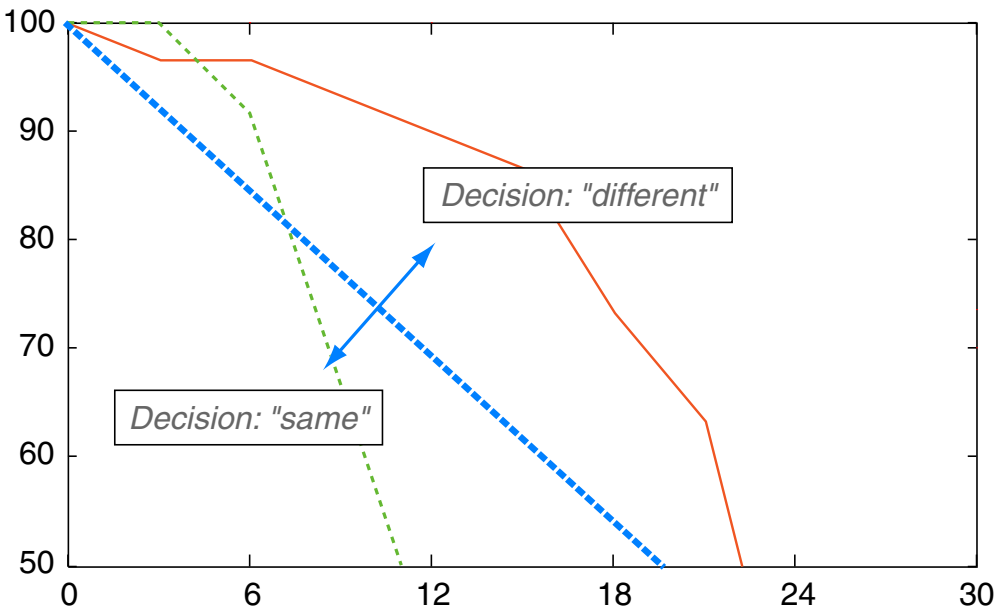


different authors ($A \neq B$)

same author ($A = B$)

Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]

Typical learning characteristic for . . .



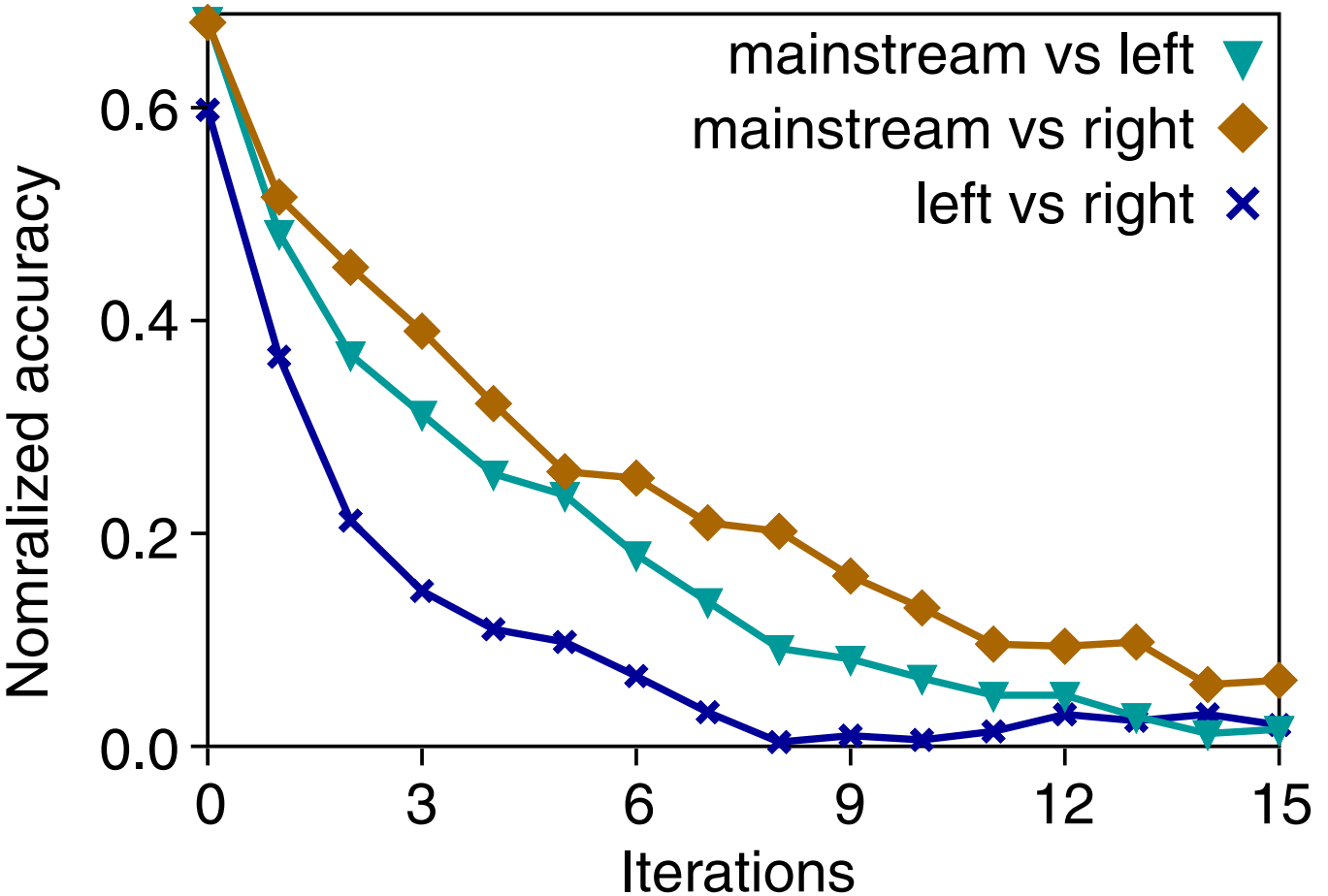
different authors ($A \neq B$)

same author ($A = B$)

The typical learning characteristic can be learned. ➜ “Meta Learning”

We apply Unmasking to distinguish style genres.

Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



Summary and Outlook

- Hyperpartisan news pages produce relatively many fake news articles
- Hyperpartisan news can be distinguished quiet well based on style
- Style-based detection allows for real-time detection

➜ Political extremism in news can be ousted or at least flagged

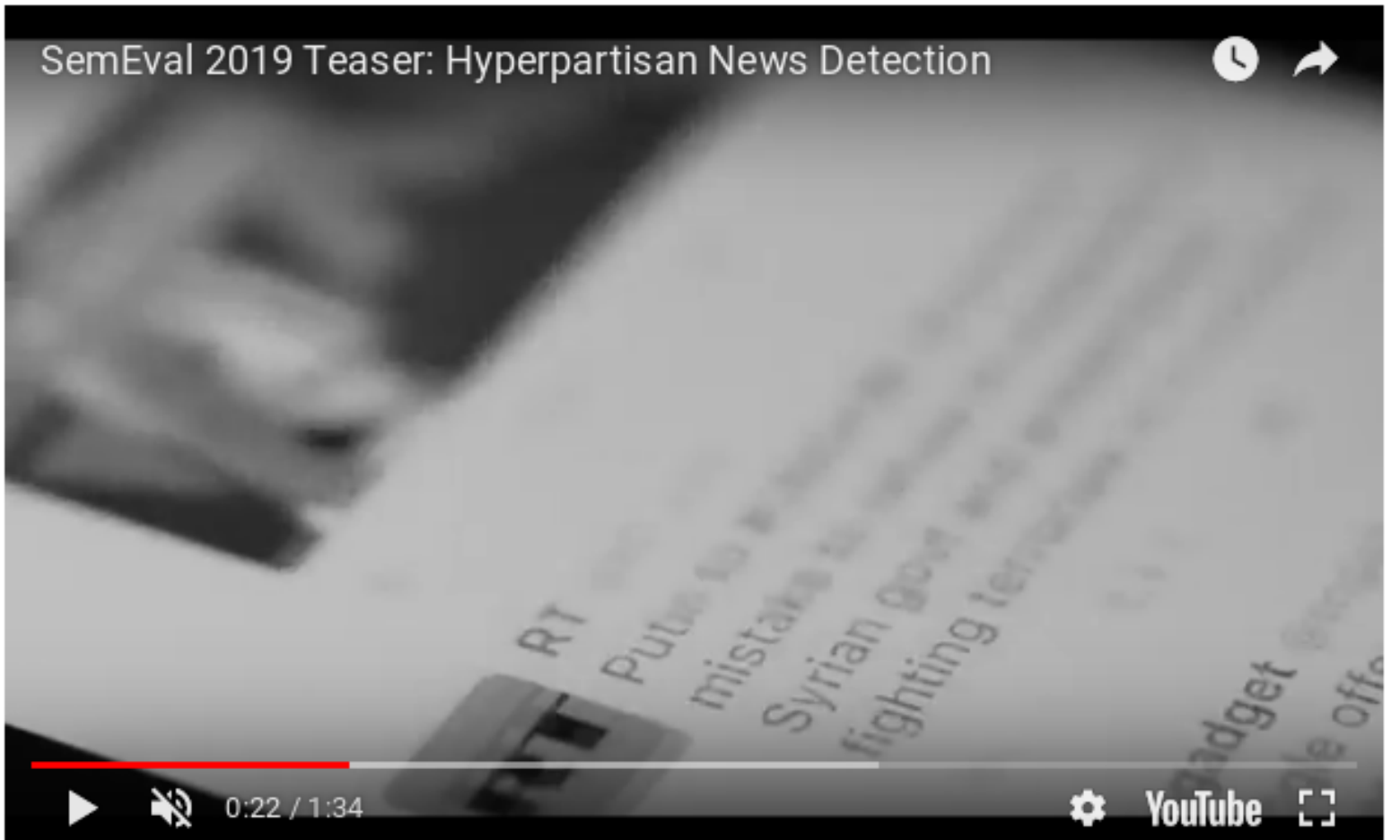
- The style of alt left and alt right news is very similar
- Linguistic evidence for the horseshoe theory of the political spectrum?

➜ Large-scale analysis required

SemEval 2019 Task 4

Hyperpartisan News Detection

Register now!



SemEval 2019 Task 4

Hyperpartisan News Detection

Register now!

webis.de/events/semeval-19

Style Model

Features

- n-grams with n $\in [1, 3]$ of characters, stop words, parts-of-speech
- 10 readability scores
- Dictionary features based on General Inquirer
- Ratios of quoted words, external links, number of paragraphs, and their average length

Feature selection

- Discard word features (n-gram features) occurring in less than 2.5% (10%) of documents

Training set

- Balancing using oversampling
- Publishers are not represented in both training and test set

Learning algorithm

- WEKA's random forest with default parameters