

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



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



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

FIRSTDRAFT

7 TYPES OF MIS- AND DISINFORMATION



SATIRE OR PARODY

No intention to cause harm but has potential to fool



MISLEADING CONTENT

Misleading use of information to frame an issue or individual



IMPOSTER CONTENT

When genuine sources are impersonated



FABRICATED CONTENT

New content is 100% false, designed to deceive and do harm



FALSE CONNECTION

When headlines, visuals or captions don't support the content



FALSE CONTEXT

When genuine content is shared with false contextual information



MANIPULATED CONTENT

When genuine information or imagery is manipulated to deceive

Image: Claire Wardle, First Draft

What are Fake News?

Disinformation displayed as news articles

FIRSTDRAFT

7 TYPES OF MIS- AND DISINFORMATION

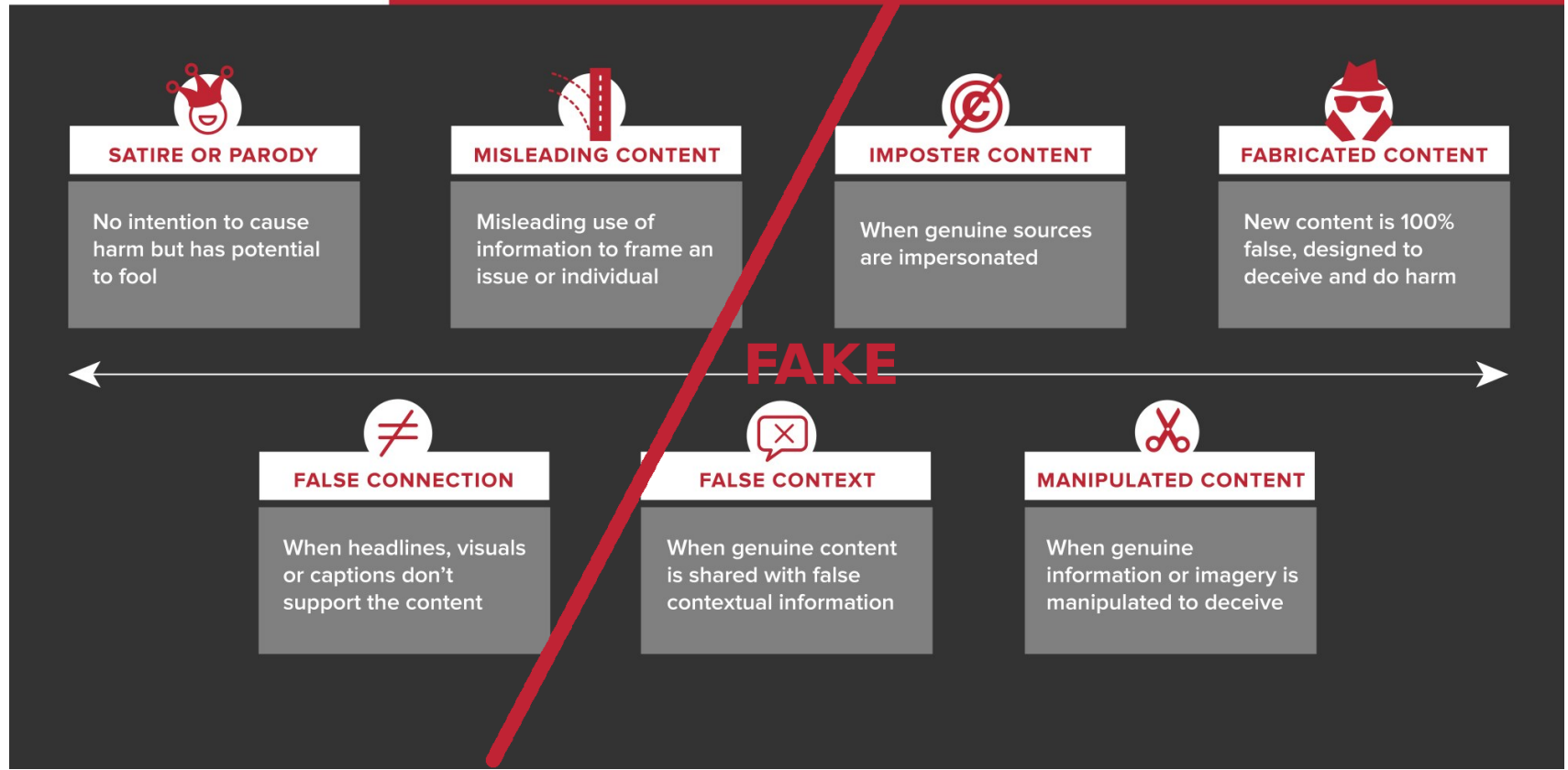



Image: Claire Wardle, First Draft

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. [\[Wikipedia\]](#)

Alt-left

Left

Center

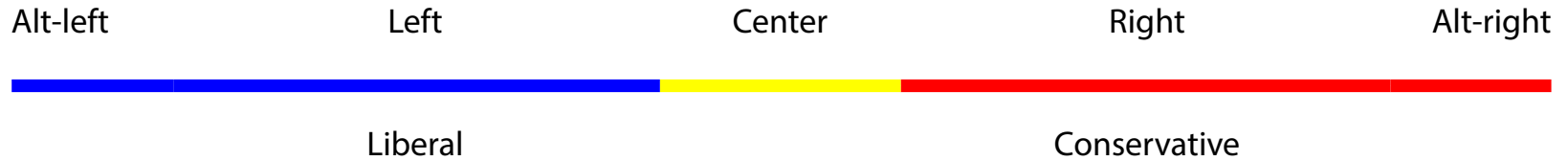
Right

Alt-right



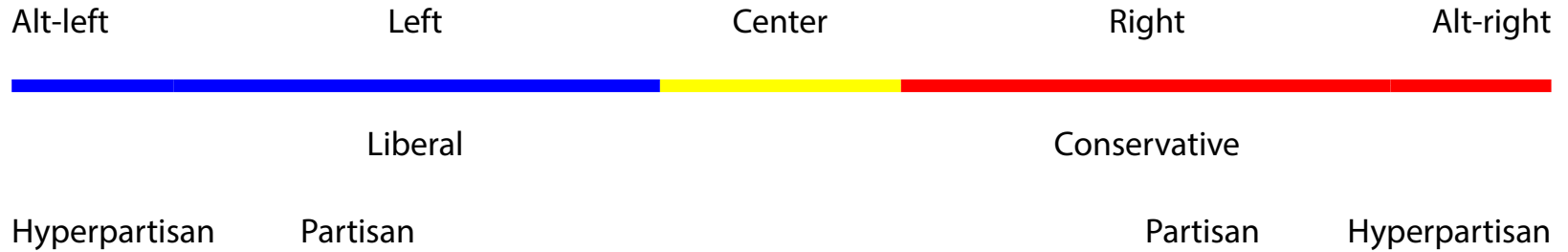
The Political Spectrum

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

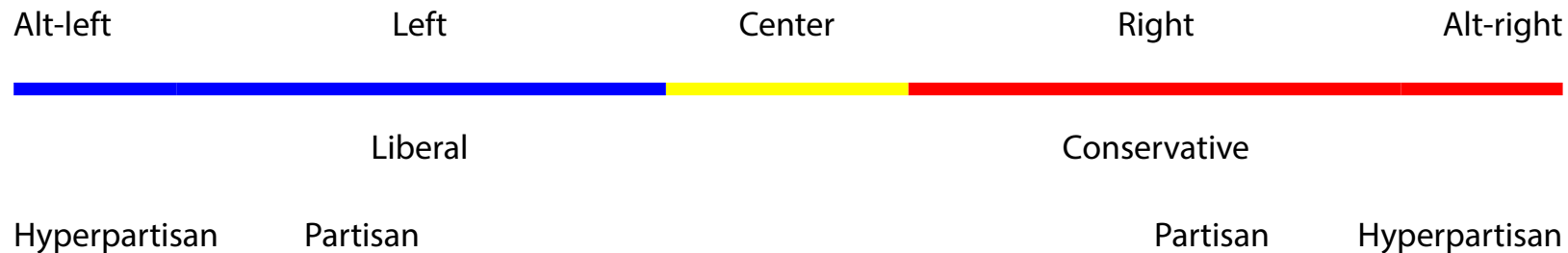
The left-right political spectrum is a system of classifying political positions, ideologies and parties. Left-wing politics and right-wing politics are often presented as opposed, although either may adopt stances from the other side. [\[Wikipedia\]](#)



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.

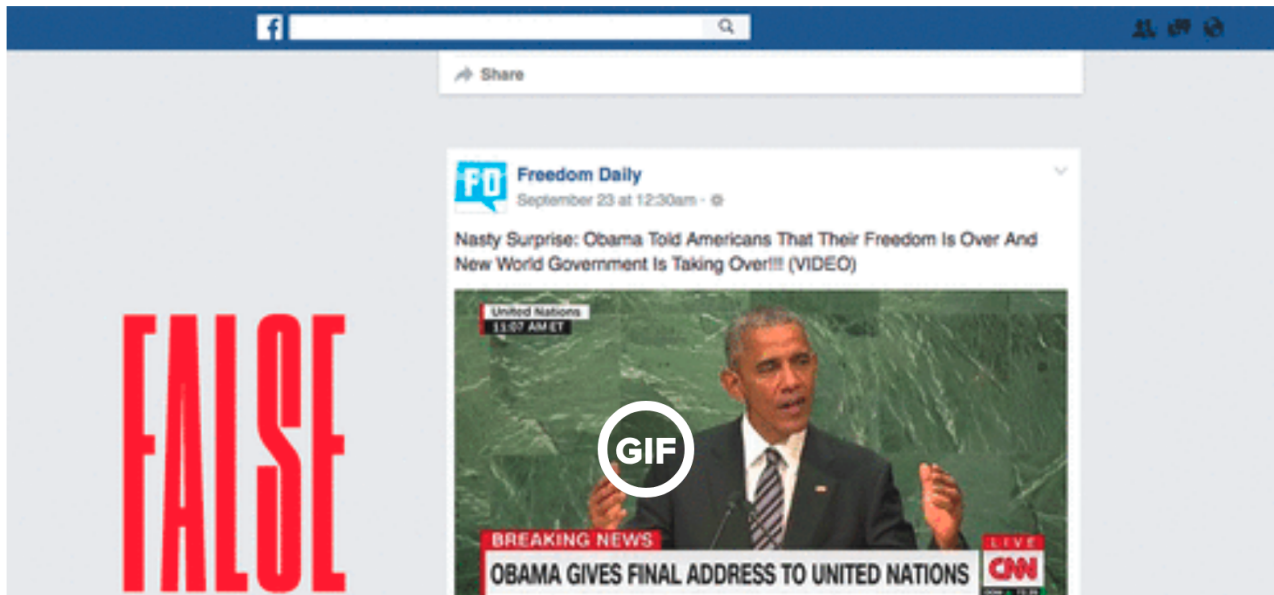
Fake News and Hyperpartisan News

BuzzFeedNEWS / 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?

FIRSTDRAFT

MISINFORMATION MATRIX








	 SATIRE OR PARODY	 FALSE CONNECTION	 MISLEADING CONTENT	 FALSE CONTEXT	 IMPOSTER CONTENT	 MANIPULATED CONTENT	 FABRICATED CONTENT
POOR JOURNALISM		✓	✓	✓			
TO PARODY	✓				✓		✓
TO PROVOKE OR TO 'PUNK'					✓	✓	✓
PASSION				✓			
PARTISANSHIP			✓	✓			
PROFIT		✓			✓		✓
POLITICAL INFLUENCE			✓	✓		✓	✓
PROPAGANDA			✓	✓	✓	✓	✓

Image: Claire Wardle, First Draft

Why are Fake News Published by Hyperpartisan Pages?

FIRSTDRAFT

MISINFORMATION MATRIX








	 SATIRE OR PARODY	 FALSE CONNECTION	 MISLEADING CONTENT	 FALSE CONTEXT	 IMPOSTER CONTENT	 MANIPULATED CONTENT	 FABRICATED CONTENT
POOR JOURNALISM		✓	✓	✓			
TO PARODY	✓				✓		✓
TO PROVOKE OR TO 'PUNK'					✓	✓	✓
PASSION				✓			
PARTISANSHIP			✓	✓			
PROFIT		✓			✓		✓
POLITICAL INFLUENCE			✓	✓		✓	✓
PROPAGANDA			✓	✓	✓	✓	✓

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 detection

Knowledge-based (also called fact checking)

Information retrieval

Etzioni et al., 2018
Magdy and Wanas, 2010
Ginsca et al., 2015

Semantic web / LOD

Wu et al., 2014
Ciampaglia et al, 2015
Shi and Weninger, 2016

Context-based

Social network analysis

Long et al., 2017
Mocanu et al., 2015
Acemoglu et al., 2010
Kwon et al., 2013
Ma et al., 2017
Volkova et al., 2017
Budak et al., 2011
Nguyen et al. 2012
Derczynski et al., 2017
Tambuscio et al., 2015

Style-based

Deception detection

Wei et al., 2013
Chen et al., 2015
Rubin et al., 2015
Wang et al., 2017
Bourgonje et al., 2017

Text categorization

Afroz et al., 2012
Badaskar et al., 2008
Rubin et al., 2016
Yang et al., 2017
Rashkin et al., 2017
Horne and Adali, 2017
Pérez-Rosas et al., 2017

Fake News and Hyperpartisan News Corpus Construction



Fake News and Hyperpartisan News Corpus Construction

<i>Orientation</i> Publisher	Fact-checking results				Σ
	true	mix	false	n/a	
<i>Center</i>	806	8	0	12	826
ABC News	90	2	0	3	95
CNN	295	4	0	8	307
Politico	421	2	0	1	424
<i>Left-wing</i>	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
<i>Right-wing</i>	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

<i>Orientation</i>	Fact-checking results				Σ
	true	mix	false	n/a	
Publisher					
<i>Center</i>	806	8	0	12	826
ABC News	2	1	1	3	95
CNN					307
Politico					424
<i>Left-wing</i>				8	256
Addicting Info				7	135
Occupy Democr				0	91
The Other 98%				1	30
<i>Right-wing</i>	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 \approx 42%

Recall \approx 41%

DOES NOT WORK!

Fake News and Hyperpartisan News Selected Results

<i>Orientation</i> Publisher	Fact-checking results				Σ
	true	mix	false	n/a	
<i>Center</i>	806	8	0	12	826
<i>ABC News</i>	22	2	1	2	95
<i>CNN</i>					307
<i>Politico</i>					424
<i>Left-wing</i>					256
<i>Addict</i>					135
<i>Occupy</i>					91
<i>The O</i>					30
<i>Right-wing</i>	276	153	72	44	545
<i>Eagle Rising</i>	106	47	25	36	214
<i>Freedom Daily</i>	49	24	22	4	99
<i>Right Wing News</i>	121	82	25	4	232
Σ	1264	212	87	64	1627

Orientation Detection

Precision \approx 21% Precision \approx 56%

Recall \approx 20% Recall \approx 59%



Annotations provided by journalists at BuzzFeed

Fake News and Hyperpartisan News Selected Results

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

Hyperpartisanship Detection

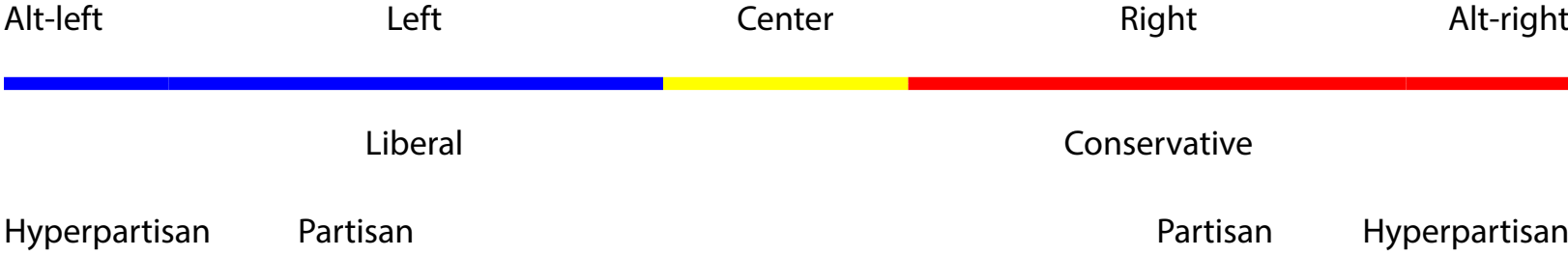
Precision \approx 69%

Recall \approx 89%



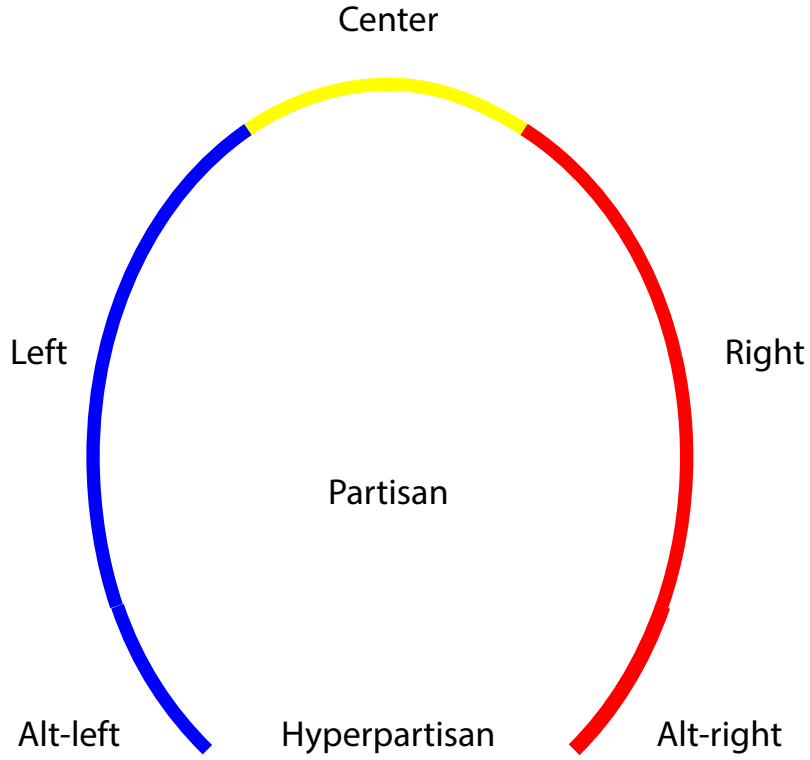
Fake News and Hyperpartisan News

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?



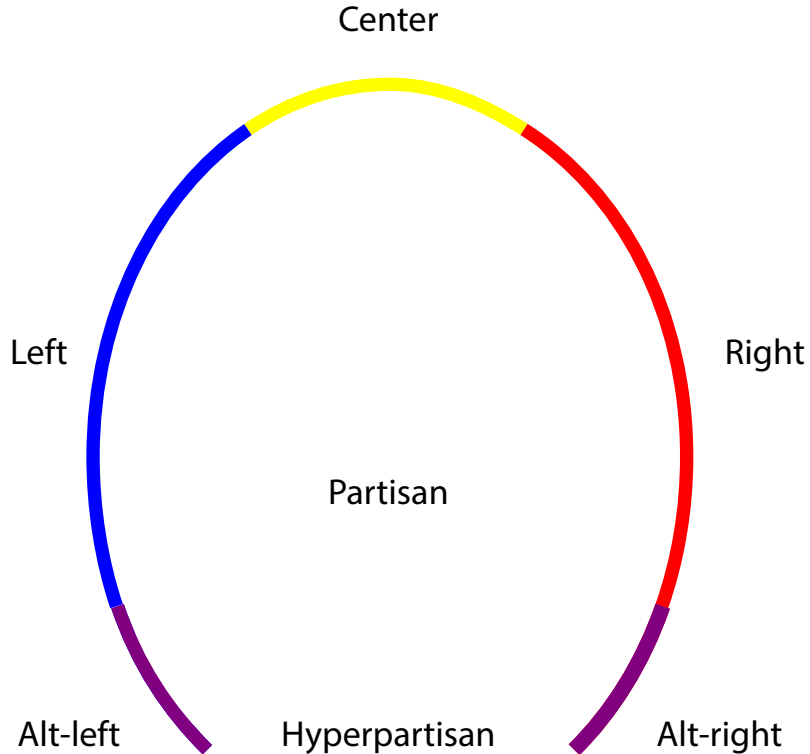
Fake News and Hyperpartisan News

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%**



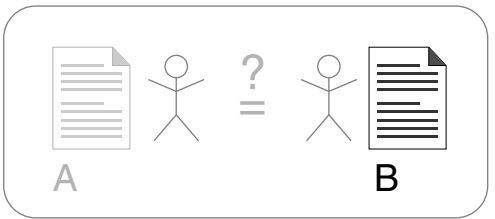
center



right-wing

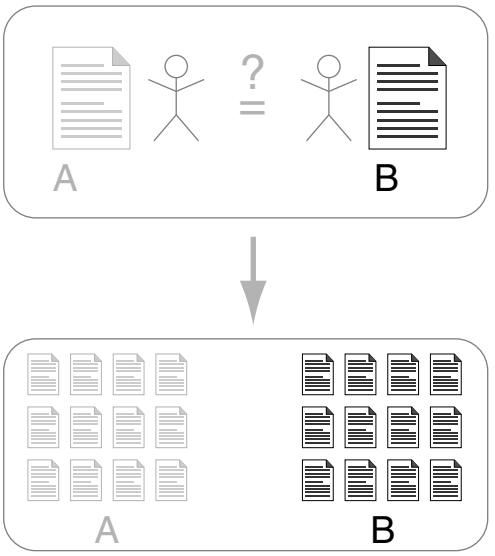
Horseshoe Validation Experiment II

Unmasking [Koppel/Schler 2004]

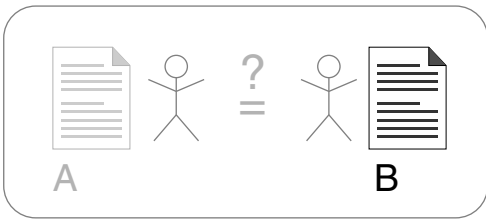


Horseshoe Validation Experiment II

Unmasking [Koppel/Schler 2004]

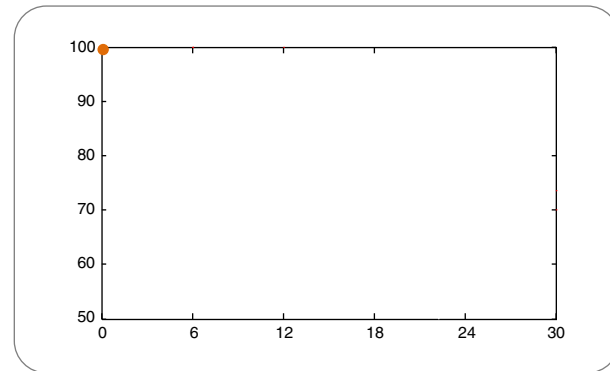
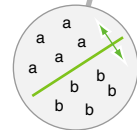
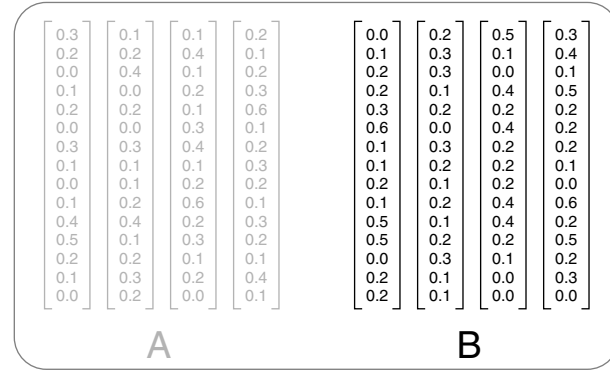
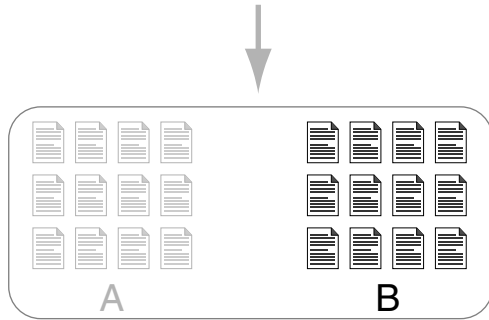
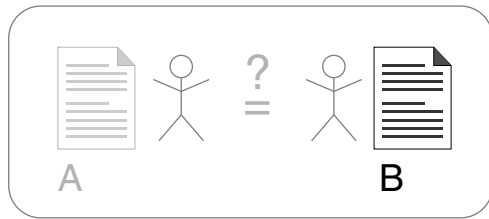


Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]

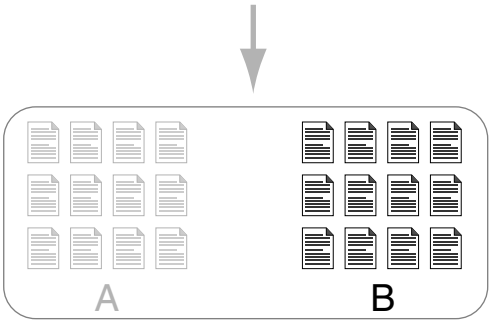
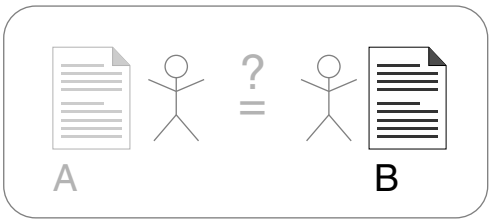


0.3	0.1	0.1	0.2	0.0	0.2	0.5	0.3
0.2	0.2	0.4	0.1	0.1	0.3	0.1	0.4
0.0	0.4	0.1	0.2	0.2	0.3	0.0	0.1
0.1	0.0	0.2	0.3	0.2	0.1	0.4	0.5
0.2	0.2	0.1	0.6	0.3	0.2	0.2	0.2
0.0	0.0	0.3	0.1	0.6	0.0	0.4	0.2
0.3	0.3	0.4	0.2	0.1	0.3	0.2	0.2
0.1	0.1	0.1	0.3	0.1	0.2	0.2	0.1
0.0	0.1	0.2	0.2	0.2	0.1	0.2	0.0
0.1	0.2	0.6	0.1	0.1	0.2	0.4	0.6
0.4	0.4	0.2	0.3	0.5	0.1	0.4	0.2
0.5	0.1	0.3	0.2	0.5	0.2	0.2	0.5
0.2	0.2	0.1	0.1	0.0	0.3	0.1	0.2
0.1	0.3	0.2	0.4	0.2	0.1	0.0	0.3
0.0	0.2	0.0	0.1	0.2	0.1	0.0	0.0
A				B			

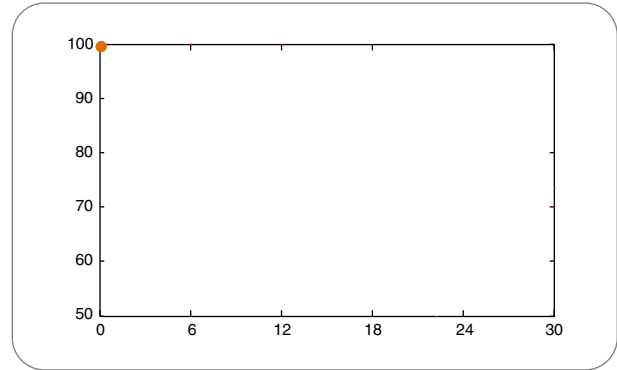
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



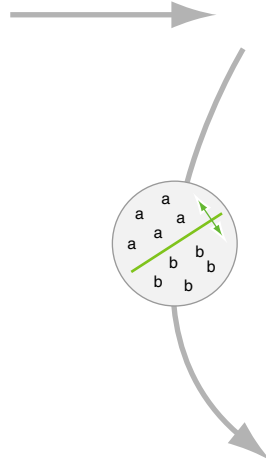
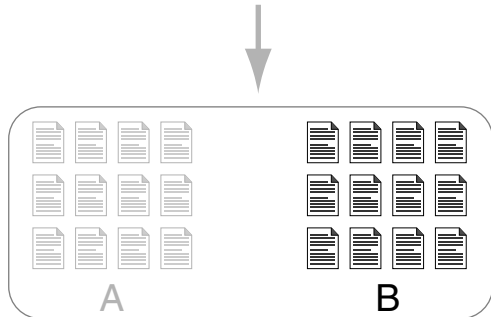
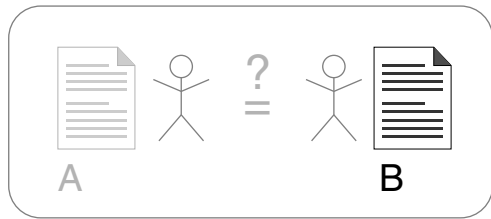
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



0.3	0.1	0.1	0.2	0.0	0.2	0.5	0.3
0.2	0.2	0.4	0.1	0.1	0.3	0.1	0.4
0.0	0.4	0.1	0.2	0.2	0.3	0.0	0.1
0.1	0.0	0.2	0.3	0.2	0.1	0.4	0.5
0.2	0.2	0.1	0.6	0.3	0.2	0.2	0.2
0.0	0.0	0.0	0.1	0.0	0.0	0.4	0.2
0.3	0.3	0.4	0.2	0.1	0.3	0.2	0.2
0.1	0.1	0.1	0.3	0.1	0.2	0.2	0.1
0.0	0.1	0.2	0.2	0.2	0.1	0.2	0.0
0.1	0.2	0.6	0.1	0.1	0.2	0.4	0.6
0.4	0.4	0.2	0.3	0.5	0.1	0.4	0.2
0.5	0.1	0.3	0.2	0.5	0.2	0.2	0.5
0.2	0.2	0.1	0.1	0.0	0.3	0.1	0.2
0.1	0.3	0.2	0.4	0.2	0.1	0.0	0.3
0.0	0.2	0.0	0.1	0.2	0.1	0.0	0.0

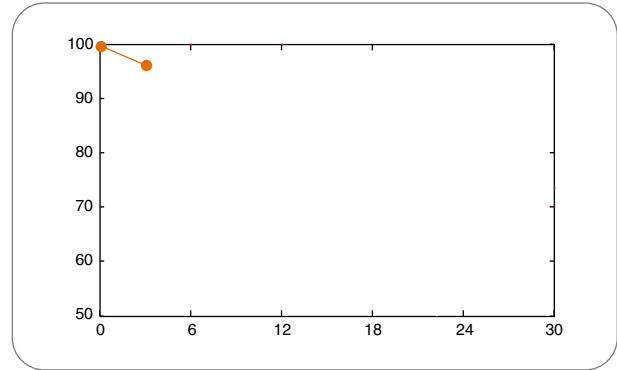


Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]

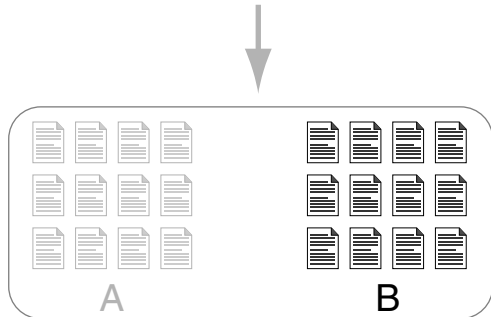
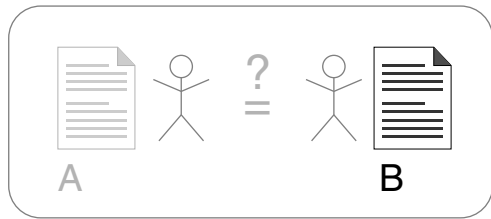


0.3	0.1	0.1	0.2	0.0	0.2	0.5	0.3
0.2	0.2	0.4	0.1	0.1	0.3	0.1	0.4
0.0	0.4	0.1	0.2	0.2	0.3	0.0	0.1
0.1	0.0	0.2	0.3	0.2	0.1	0.4	0.5
0.2	0.2	0.1	0.6	0.3	0.2	0.2	0.2
0.0	0.0	0.0	0.1	0.0	0.0	0.4	0.2
0.3	0.3	0.4	0.2	0.1	0.3	0.2	0.2
0.1	0.1	0.1	0.3	0.1	0.2	0.2	0.1
0.0	0.1	0.2	0.2	0.2	0.1	0.2	0.0
0.1	0.2	0.6	0.1	0.1	0.2	0.4	0.6
0.4	0.4	0.2	0.3	0.5	0.1	0.4	0.2
0.5	0.1	0.3	0.2	0.5	0.2	0.2	0.5
0.2	0.2	0.1	0.1	0.0	0.0	0.1	0.2
0.1	0.3	0.2	0.4	0.2	0.1	0.0	0.3
0.0	0.2	0.0	0.1	0.2	0.1	0.0	0.0

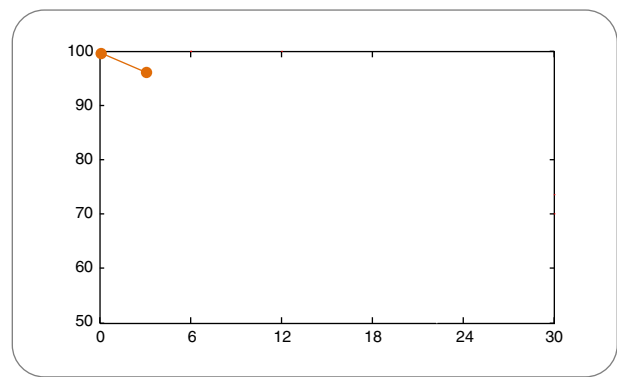
A B



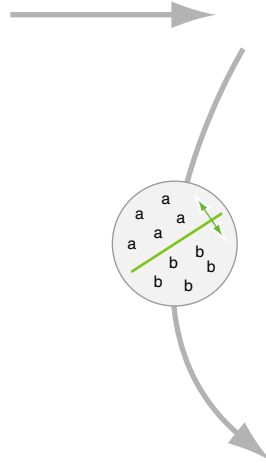
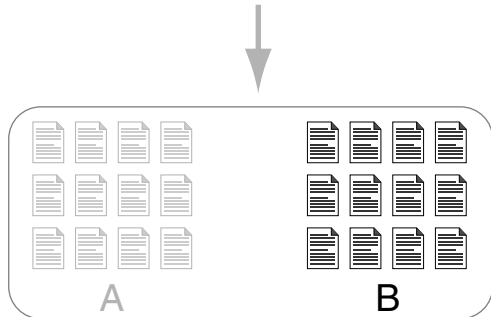
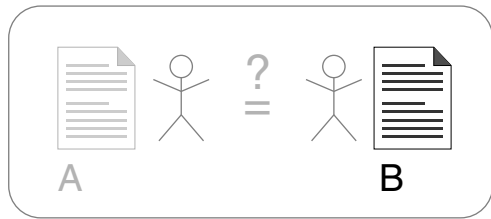
Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]



0.3	0.1	0.1	0.2	0.0	0.2	0.5	0.3
0.2	0.2	0.4	0.1	0.1	0.3	0.1	0.4
0.0	0.4	0.1	0.2	0.2	0.1	0.0	0.1
0.1	0.0	0.2	0.3	0.2	0.1	0.4	0.5
0.2	0.2	0.1	0.6	0.3	0.2	0.2	0.2
0.0	0.0	0.0	0.1	0.0	0.0	0.4	0.2
0.0	0.0	0.4	0.2	0.1	0.0	0.2	0.2
0.1	0.1	0.1	0.3	0.1	0.2	0.2	0.1
0.0	0.1	0.2	0.2	0.2	0.1	0.2	0.0
0.1	0.2	0.6	0.1	0.1	0.2	0.4	0.6
0.4	0.4	0.2	0.3	0.5	0.1	0.4	0.2
0.5	0.1	0.3	0.2	0.5	0.2	0.2	0.5
0.2	0.2	0.1	0.1	0.0	0.0	0.1	0.2
0.1	0.3	0.2	0.4	0.2	0.1	0.0	0.3
0.0	0.2	0.0	0.1	0.2	0.1	0.0	0.0

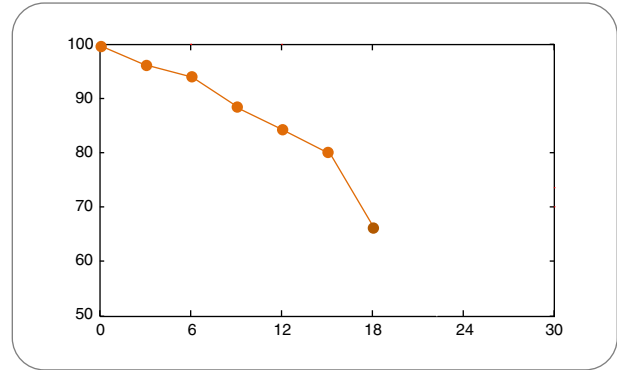


Horseshoe Validation Experiment II Unmasking [Koppel/Schler 2004]

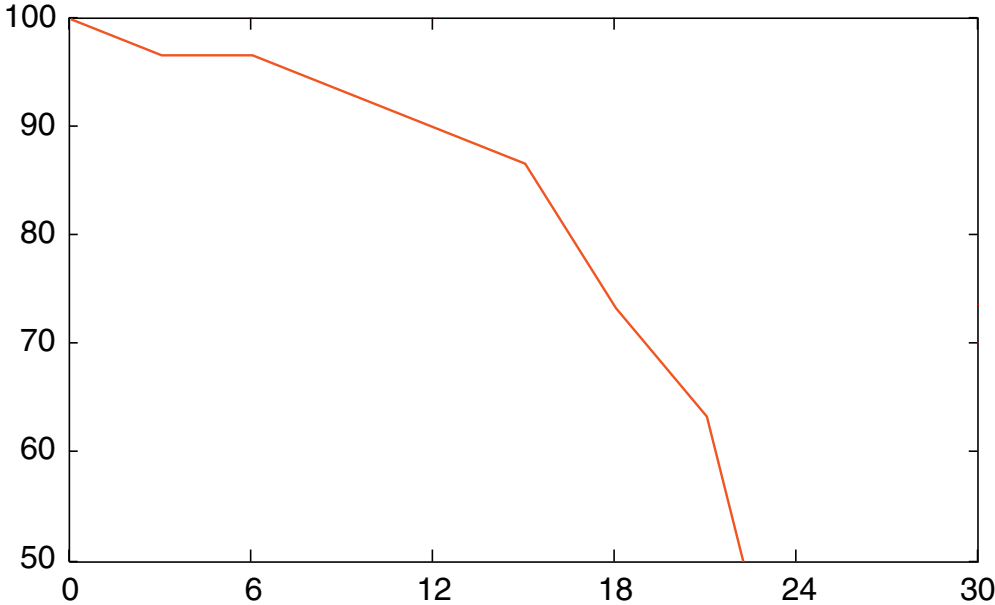


0.0	0.1	0.1	0.2	0.0	0.2	0.5	0.3
0.2	0.2	0.4	0.1	0.1	0.3	0.1	0.4
0.0	0.4	0.1	0.2	0.2	0.3	0.0	0.1
0.1	0.0	0.2	0.3	0.2	0.1	0.4	0.5
0.2	0.2	0.1	0.0	0.3	0.2	0.2	0.2
0.0	0.0	0.0	0.1	0.0	0.0	0.4	0.2
0.3	0.3	0.4	0.2	0.1	0.3	0.2	0.2
0.1	0.1	0.1	0.3	0.1	0.2	0.2	0.1
0.0	0.1	0.2	0.2	0.2	0.1	0.2	0.0
0.1	0.2	0.6	0.1	0.1	0.2	0.4	0.6
0.4	0.4	0.2	0.3	0.5	0.1	0.4	0.2
0.5	0.1	0.3	0.2	0.5	0.2	0.2	0.5
0.2	0.2	0.4	0.1	0.0	0.3	0.1	0.2
0.1	0.3	0.2	0.4	0.2	0.1	0.0	0.3
0.0	0.2	0.0	0.1	0.2	0.1	0.0	0.0

A B

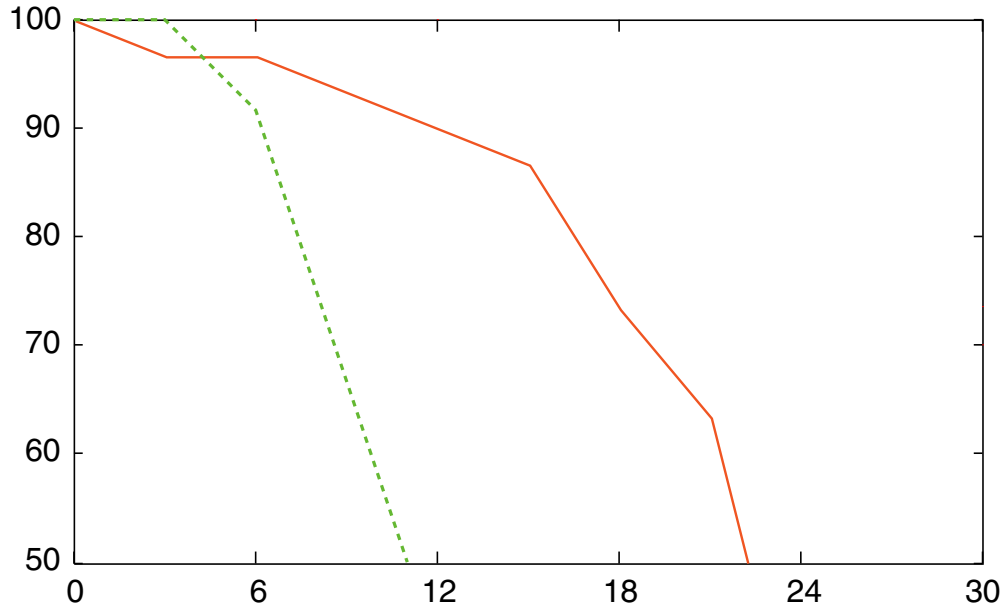


Typical learning characteristic for ...



different authors ($A \neq B$)

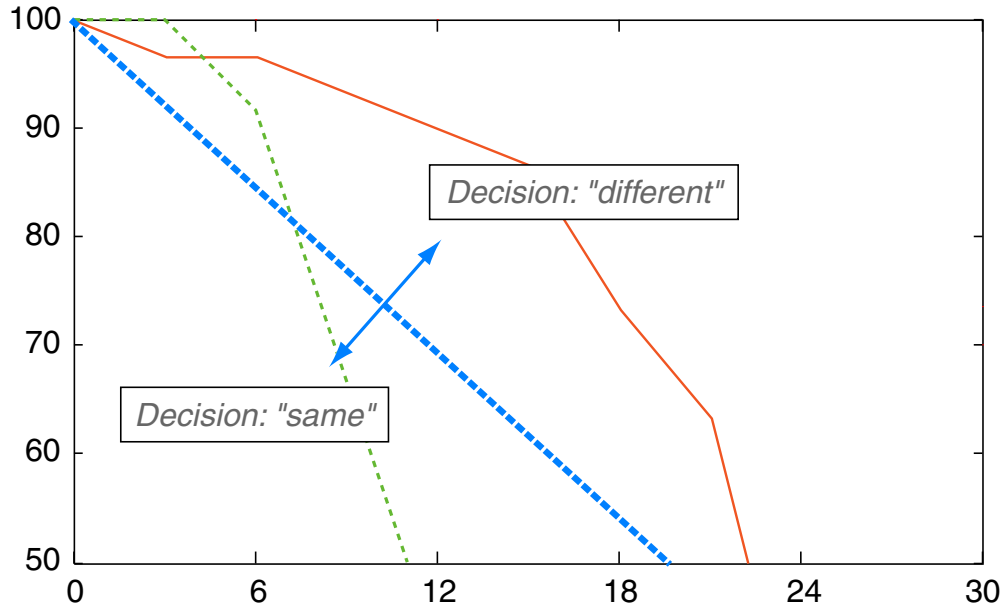
Typical learning characteristic for ...



different authors ($A \neq B$)

same author ($A = B$)

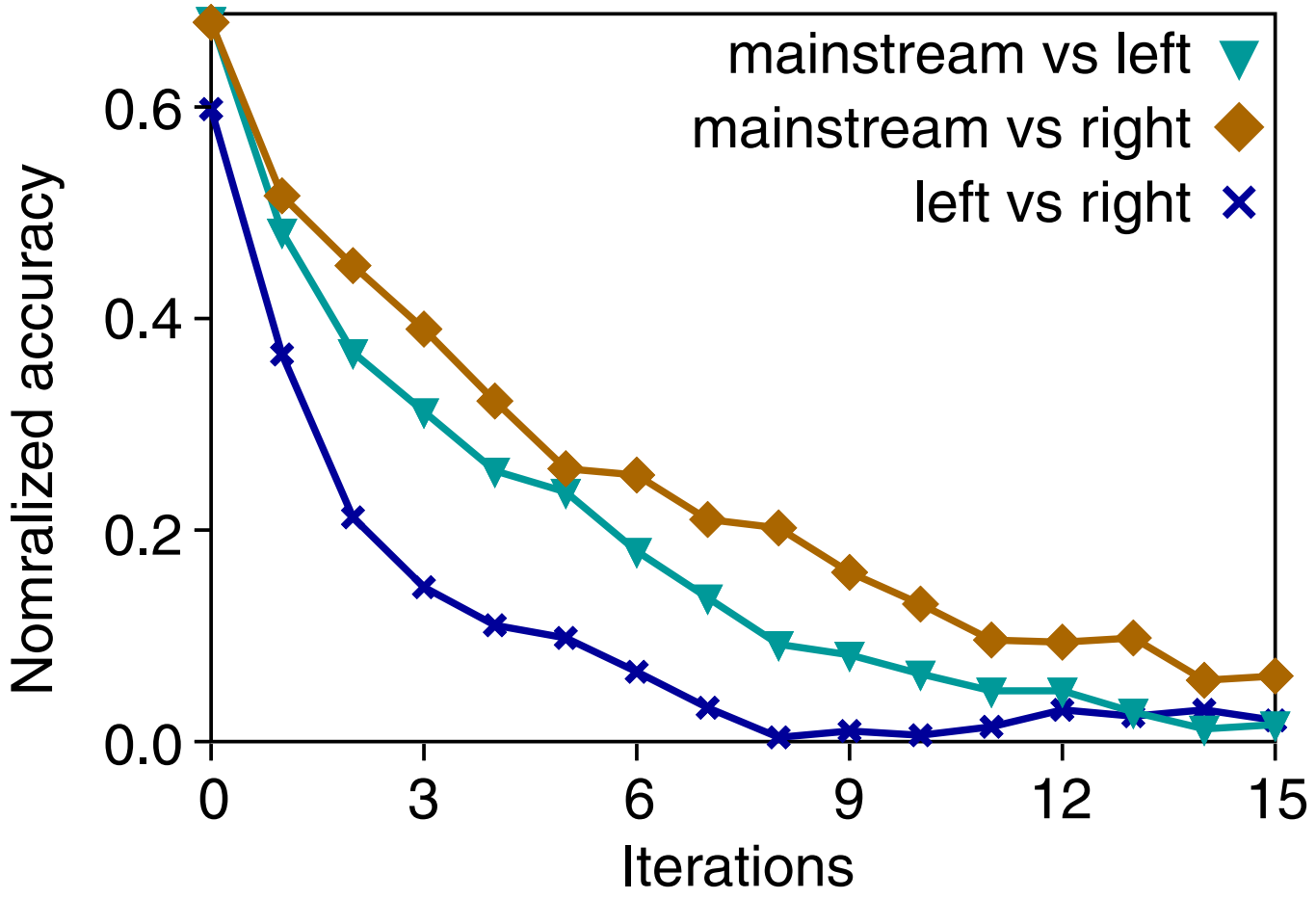
Typical learning characteristic for ...



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.



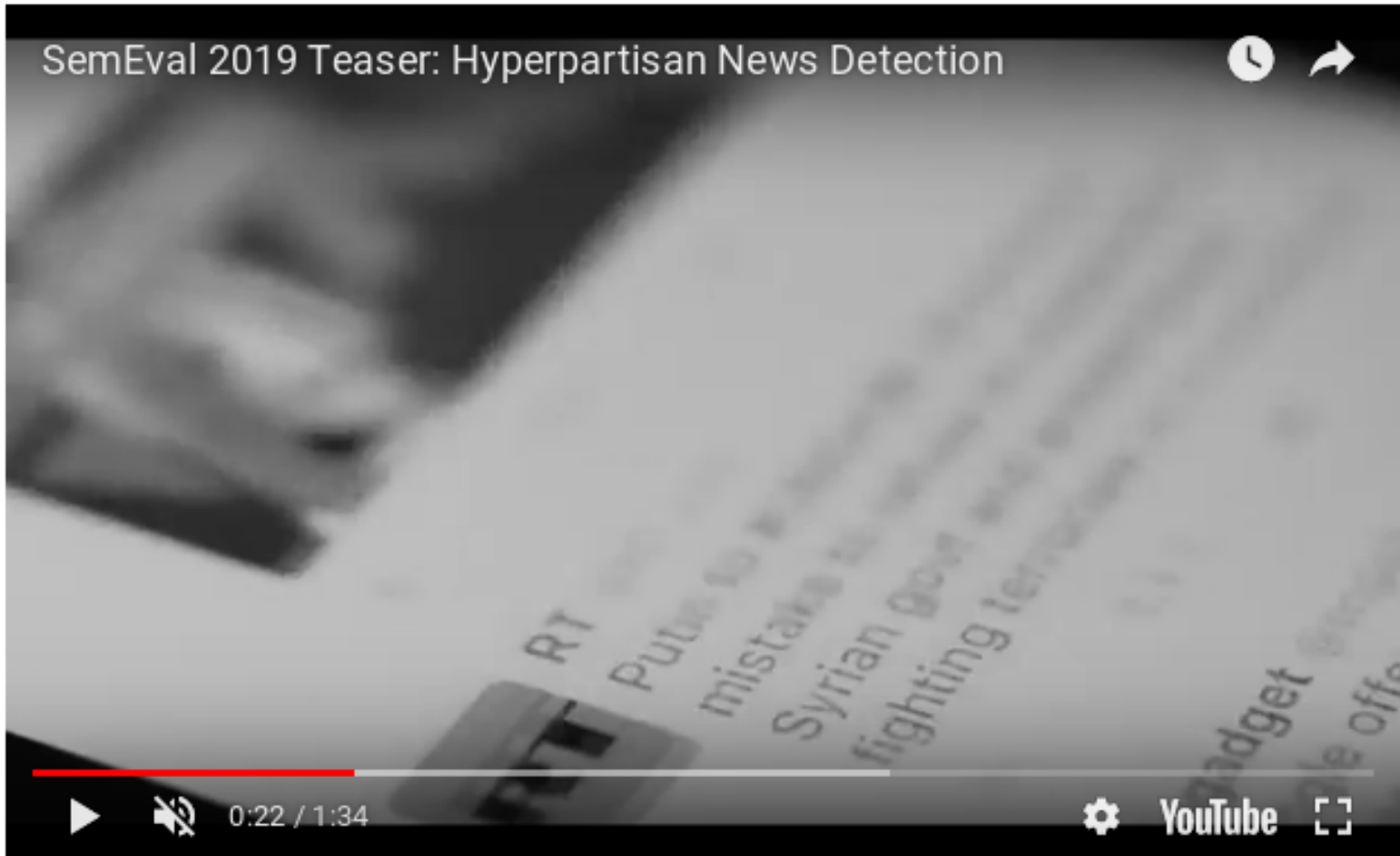
Summary and Outlook

- ❑ Hyperpartisan news pages produce relatively many fake news articles
- ❑ Hyperpartisan news can be distinguished quite 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

Register now!

Hyperpartisan News Detection



Register now!

Hyperpartisan News Detection

SemEval 2019 Teaser: Hyperpartisan News Detection



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