Deception and Trust in Spoken Dialogue

Sarah Ita Levitan & Julia Hirschberg

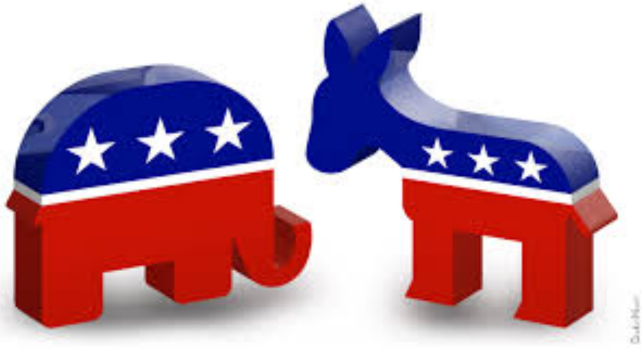
COMS 6998

April 12, 2019

Today

- Deception
  - Detection from text and speech
- Trust
  - Trust and mistrust in deceptive dialogue
  - LieCatcher game for trust annotation
  - Trust in news articles

Motivation



Human performance at deception detection
(Aamodt & Mitchell, 2004)

Group	# Studies	# Subjects	Accuracy %
Criminals	1	52	65.40
Secret service	**1**	**34**	**64.12**
Psychologists	4	508	61.56
Judges	**2**	**194**	**59.01**
Cops	**8**	**511**	**55.16**
Federal officers	**4**	**341**	**54.54**
Students	122	8,876	54.20
Detectives	**5**	**341**	**51.16**
Parole officers	**1**	**32**	**40.42**

Prior work

Body posture and gestures (Burgoon et al, ‘94)

Facial expressions (Ekman, ‘76; Frank, ‘03)

Biometric factors (Horvath, ‘73)

Brain imaging technologies (Bles & Haynes, ‘08)

Language-based

- Text (Adams, ’96, Pennebaker et al., ’01, Choudhury, ‘14)
- Speech (Enos, ‘09)

Language-based deception detection

Practitioners

Statement analysis (Adams, 1996)

SCAN (Smith, 2001)

Reid & Associates (Buckley, 2000)

Forensic linguists

Text

Bachenko et al. (2008)

Ott et al. (2011)

Perez-Rosas & Mihalcea (2015)

Speech

Voice Stress Analysis (Horvath, 1982)

Streeter et al. (1977)

Ekman et al. (1991)

Enos (2009)

Challenges

Data

Ground truth annotation

Laboratory vs. real-world deception

Individual and cultural differences

Related work

Case studies

Intuition

Partial automation

Few speakers

Small amounts of data

Almost all text-based

Limited domains

Some pseudo-science

No stakes

Current work

Large scale speech: 120 hrs, 340 speakers

Automatically extracted features

Machine learning

Statistical methods

Speech + text

Gender, culture, personality differences

Dialogue

Fake resume paradigm

Financial incentive

Research Goals

Increase scientific understanding of deceptive behavior

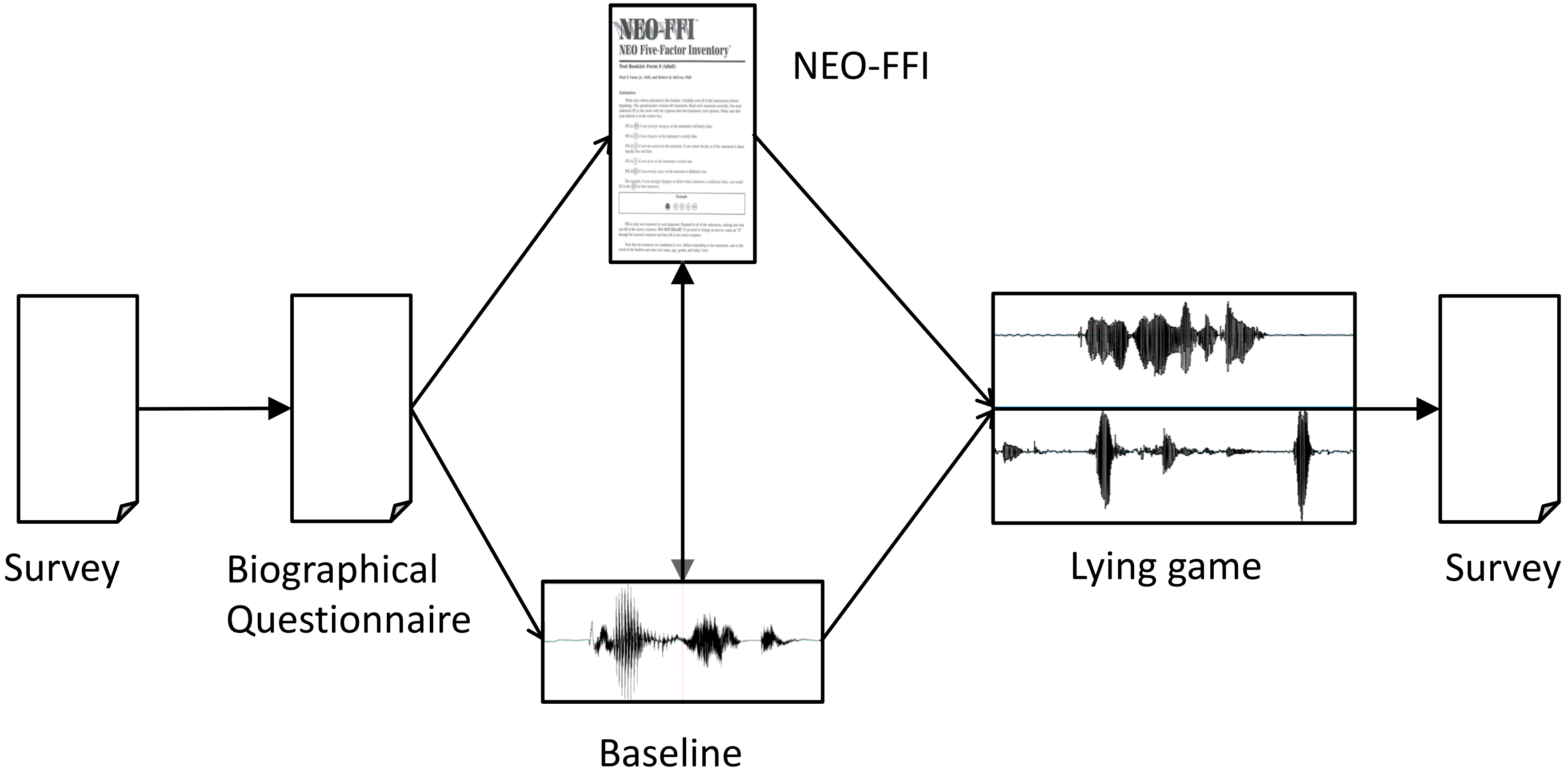
- What are the acoustic-prosodic and linguistic characteristics of deceptive speech?
- How do cues to deception differ across gender, culture, and personality types?

Develop automated methods to detect deceptive language

Contributions

- Large-scale corpus of deceptive dialogues
- Acoustic-prosodic and linguistic cues to deception
- Automatic deception classification
- Study of entrainment in deceptive dialogue
- Individual differences in cues to deception
- Deception classification leveraging speaker differences

Columbia X-Cultural Deception Corpus



Columbia X-Cultural Deception Corpus

>120 hours of subject speech

340 subjects

Native speakers of MC and SAE

Fake resume paradigm

NEO-FFI personality scores

Baseline sample

Financial incentive

Deception production and perception

Global and local deception labels

Units of analysis

IPU Pause-free segment of speech from a single speaker

Turn Sequence of speech from one speaker without intervening speech from the other speaker

Question response Interviewee turn following an interviewer biographical question

Question chunk Set of interviewee turns responding to an interviewer biographical question and subsequent follow-up questions

Units of analysis

Unit	Interviewer	Interviewee	Total
IPU	81536	111428	192964
Turn	41768	43673	85459
Question Response	8092	8092	16184
Question Chunk	8092	8092	16184

“Have you ever tweeted?”

TRUE or FALSE?

“Have you ever tweeted?”

Deception Detection from Text and Speech

Research questions:

1. What are the acoustic-prosodic and linguistic characteristics of deceptive and truthful speech?
2. Can we train machine learning classifiers to automatically distinguish between truthful and deceptive speech?

Acoustic-prosodic and linguistic characteristics of deception and truth

Four feature sets

- Acoustic-prosodic (Praat; Boersma et al., 2002)
- Linguistic Deception Indicators (LDI)
- Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015)
- Complexity (Lu, 2010)

Two units of analysis

- Question response
- Question chunk

Paired t-tests; FDR correction, α=0.05

152 features

Acoustic-prosodic (8) pitch {max, mean}, intensity {max, mean}, speaking rate, jitter, shimmer NHR

LDI (28) hedge words, filled pauses, contractions, denials, laughter, DAL (Dictionary of Affect in Language; Whissel et al., 1986), specificity (Li & Nenkova, 2015)

LIWC (93) word counts for semantic classes – linguistic, markers of psychological processes, punctuation, formality

Complexity (23) measures of syntactic complexity (e.g. clauses per sentence, coordinate phrases per clause)

Acoustic-prosodic characteristics

Pitch max

Pitch mean

Intensity max

Intensity mean

Speaking rate

Jitter

Shimmer

NHR

Acoustic-prosodic characteristics

Pitch max

Pitch mean

Intensity max

Intensity mean

Speaking rate

Jitter

Shimmer

NHR

Acoustic-prosodic characteristics

Pitch max

Pitch mean

Intensity max

Intensity mean

Speaking rate

Jitter

Shimmer

NHR

Increased pitch

Ekman et al. (1976)

Streeter et al. (1977)

DePaulo et al. (2003)

Increased intensity

DePaulo et al. (2003) – no effect

Linguistic Deception Indicators (LDI)

hasAbsolutelyReally	isJustNo	hasHedgePhrase
hasContraction	noYesORNo	numHedgePhrases
hasI	specificDenial	hasLaugh
hasWe	thirdPersonPronouns	numLaugh
hasYes	hasFalseStart	DAL.wc
hasNAposT	hasFilledPause	DAL.pleasant
hasNo	numFilledPauses	DAL.activate
hasNot	hasCuePhrase	DAL.imagery
isJustYes	numCuePhrases	specScores

Linguistic Deception Indicators (LDI)

hasAbsolutelyReally	isJustNo	hasHedgePhrase
hasContraction	noYesOrNo	numHedgePhrases
hasI	specificDenial	hasLaugh
hasWe	thirdPersonPronouns	numLaugh
hasYes	hasFalseStart	DAL.wc
hasNAposT	hasFilledPause	DAL.pleasant
hasNo	numFilledPauses	DAL.activate
hasNot	hasCuePhrase	DAL.imagery
isJustYes	numCuePhrases	specScores

Linguistic Deception Indicators (LDI)

hasAbsolutelyReally
hasContraction
hasI
hasWe
hasYes
hasNAposT
hasNo
hasNot
isJustYes
isJustNo
noYesOrNo
specificDenial
thirdPersonPronouns
hasFalseStart
hasFilledPause
numFilledPauses
hasCuePhrase
numCuePhrases
hasHedgePhrase
numHedgePhrases
hasLaugh
numLaugh
DAL.wc
DAL.pleasant
DAL.activate
DAL.imagery
specScores

Linguistic Deception Indicators (LDI)

hasAbsolutelyReally
hasContraction
hasI
hasWe
hasYes
hasNApostT
hasNo
hasNot
isJustYes

isJustNo
noYesOrNo
specificDenial
thirdPersonPronouns
hasFalseStart
hasFilledPause
numFilledPauses
hasCuePhrase
numCuePhrases

hasHedgePhrase
numHedgePhrases
hasLaugh
numLaugh
DAL.wc
DAL.pleasant
DAL.activate
DAL.imagery
specScores

Linguistic Deception Indicators (LDI)

hasAbsolutelyReally
hasContraction
hasI
hasWe
hasYes
hasNApostT
hasNo
hasNot
isJustYes

isJustNo
noYesOrNo
specificDenial
thirdPersonPronouns
hasFalseStart
hasFilledPause
numFilledPauses
hasCuePhrase
numCuePhrases

hasHedgePhrase
numHedgePhrases
hasLaugh
numLaugh
DAL.wc
DAL.pleasant
DAL.activate
DAL.imagery
specScores

Linguistic Inquiry and Word Count (LIWC)

Adj	Auxverb	Family	Netspeak	Social
Adverb	Bio	Focuspast	Nonflu	Space
Affect	Clout	Focuspres	Number	tentat
Affiliation	Cogproc	Function	Posemo	Time
Analytic	Compare	I	Ppron	Tone
Apostro	Conj	Informal	Prep	Verb
Article	Dic	Insight	Pronoun	Wc
Assent	Differ	Ipron	Relative	Work
Authentic	Drives	Negate	Sixltr	WPS

Linguistic Inquiry and Word Count (LIWC)

Adj	Auxverb	Family	Netspeak	Social
Adverb	Bio	Focuspast	Nonflu	Space
Affect	Clout	Focuspres	Number	tentat
Affiliation	Cogproc	Function	Posemo	Time
Analytic	Compare	I	Ppron	Tone
Apostro	Conj	Informal	Prep	Verb
Article	Dic	Insight	Pronoun	WC
Assent	Differ	Ipron	Relative	Work
Authentic	Drives	Negate	Sixltr	WPS

Linguistic Inquiry and Word Count (LIWC)

Adj	Auxverb	Family	Netspeak	Social
Adverb	Bio	Focuspast	Nonflu	Space
Affect	Clout	Focuspres	Number	tentat
Affiliation	Cogproc	Function	Posemo	Time
Analytic	Compare	I	Ppron	Tone
Apostro	Conj	Informal	Prep	Verb
Article	Dic	Insight	Pronoun	WC
Assent	Differ	Ipron	Relative	Work
Authentic	Drives	Negate	Sixltr	WPS

Linguistic Inquiry and Word Count (LIWC)

Adj	Auxverb	Family	Netspeak	Social
Adverb	Bio	Focuspast	Nonflu	Space
Affect	Clout	Focuspres	Number	tentat
Affiliation	**Cogproc**	Function	Posemo	Time
Analytic	Compare	I	Ppron	Tone
Apostro	Conj	Informal	Prep	Verb
Article	Dic	Insight	Pronoun	WC
Assent	Differ	Ipron	Relative	Work
Authentic	Drives	Negate	Sixltr	WPS

Complexity

W words

VP verb phrase

C clauses

T t-units

DC dep. clause

CT complex t-unit

CP coordinate phrase

CN complex nominal

MLS mean length sentence

MLT mean length t-unit

MLC mean length clause

C.S clauses/sentence

VP.T verb phrases/t-unit

C.T clauses/t-unit

DC.C dep clauses/clause

DC.T dep clauses/t-unit

T.S t-units/sentence

CT.T complex t-units/t-unit

CP.T coord phrases/t-unit

CP.C coord phrases/clause

CN.T complex nom/t-unit

CN.C complex nom/clause

Complexity

W words

VP verb phrase

C clauses

T t-units

DC dep. clause

CT complex t-unit

CP coordinate phrase

CN complex nominal

MLS mean length sentence

MLT mean length t-unit

MLC mean length clause

C.S clauses/sentence

VP.T verb phrases/t-unit

C.T clauses/t-unit

DC.C dep clauses/clause

DC.T dep clauses/t-unit

T.S t-units/sentence

CT.T complex t-units/t-unit

CP.T coord phrases/t-unit

CP.C coord phrases/clause

CN.T complex nom/t-unit

CN.C complex nom/clause

Summary: acoustic-prosodic and linguistic characteristics of deception and truth

Deception

Increased pitch & intensity max

Poor speech planning

Descriptive, detailed

Complex

Hedge

Entrainment

Truth

Negation

Cue phrases

Cognitive process

Function words

Automatic deception detection

Four units of analysis:

- IPU, turn, question response, question chunk

Four statistical classifiers: Random Forest, Logistic Regression, SVM, Naïve Bayes

Three neural network classifiers: DNN, LSTM, Hybrid

Six feature sets (all segmentations):

- Praat, ISO9, LDI, LIWC, complexity, n-grams

Six syntactic feature sets (question response, question chunk):

- POS, word+POS, PR-unlex, PR-lex, GPR-unlex, GPR-lex (Feng et al. 2010)

Evaluation:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Baselines:

- Random: 50% accuracy
- Human: 56.75% accuracy (question chunk units)

Combined features

Acoustic

- Praat, IS09

Lexical

- LIWC, LDI, n-grams

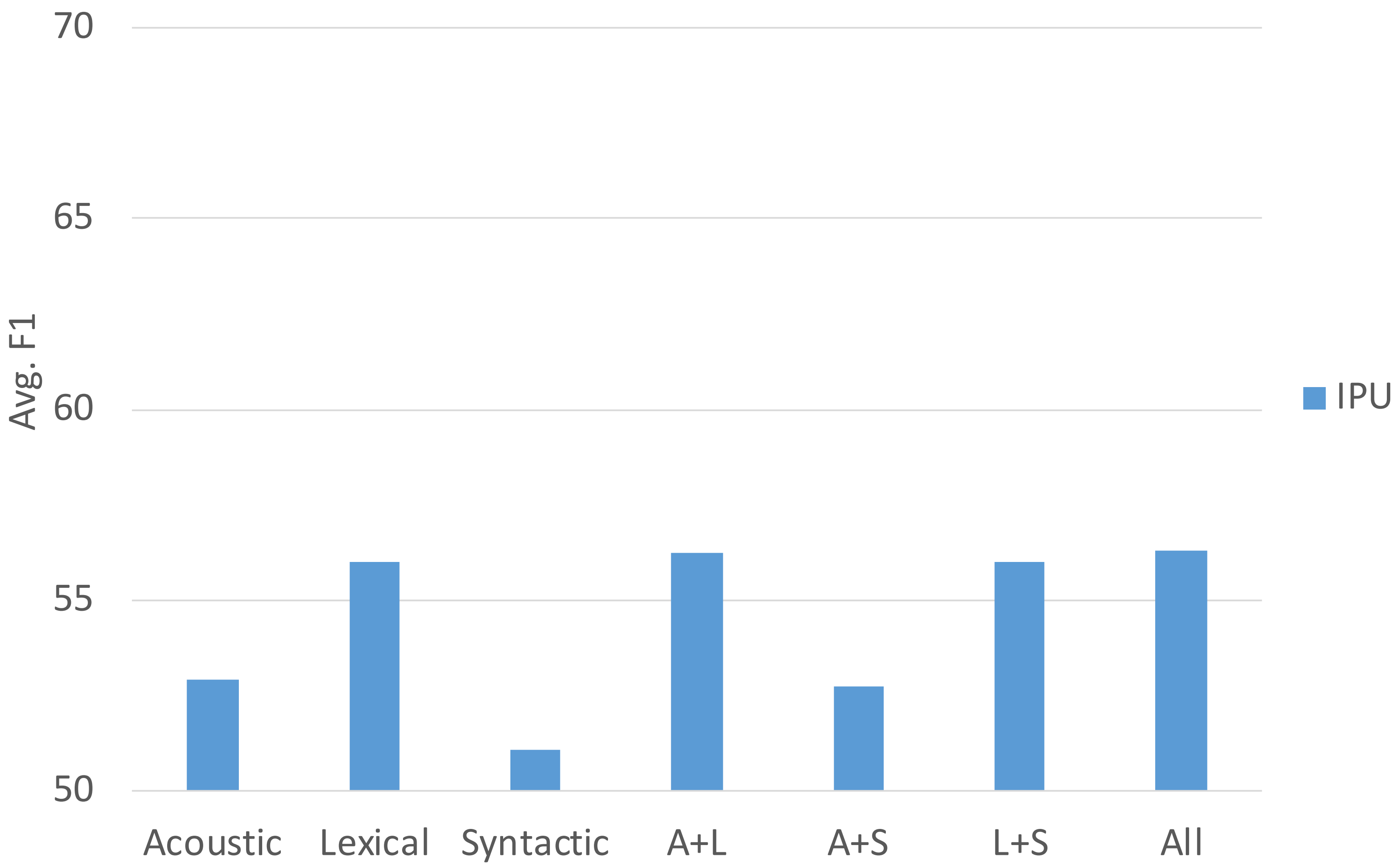
Syntactic

- IPU, turn: complexity
- Question response, question chunk: complexity, POS, word+POS, prod rules

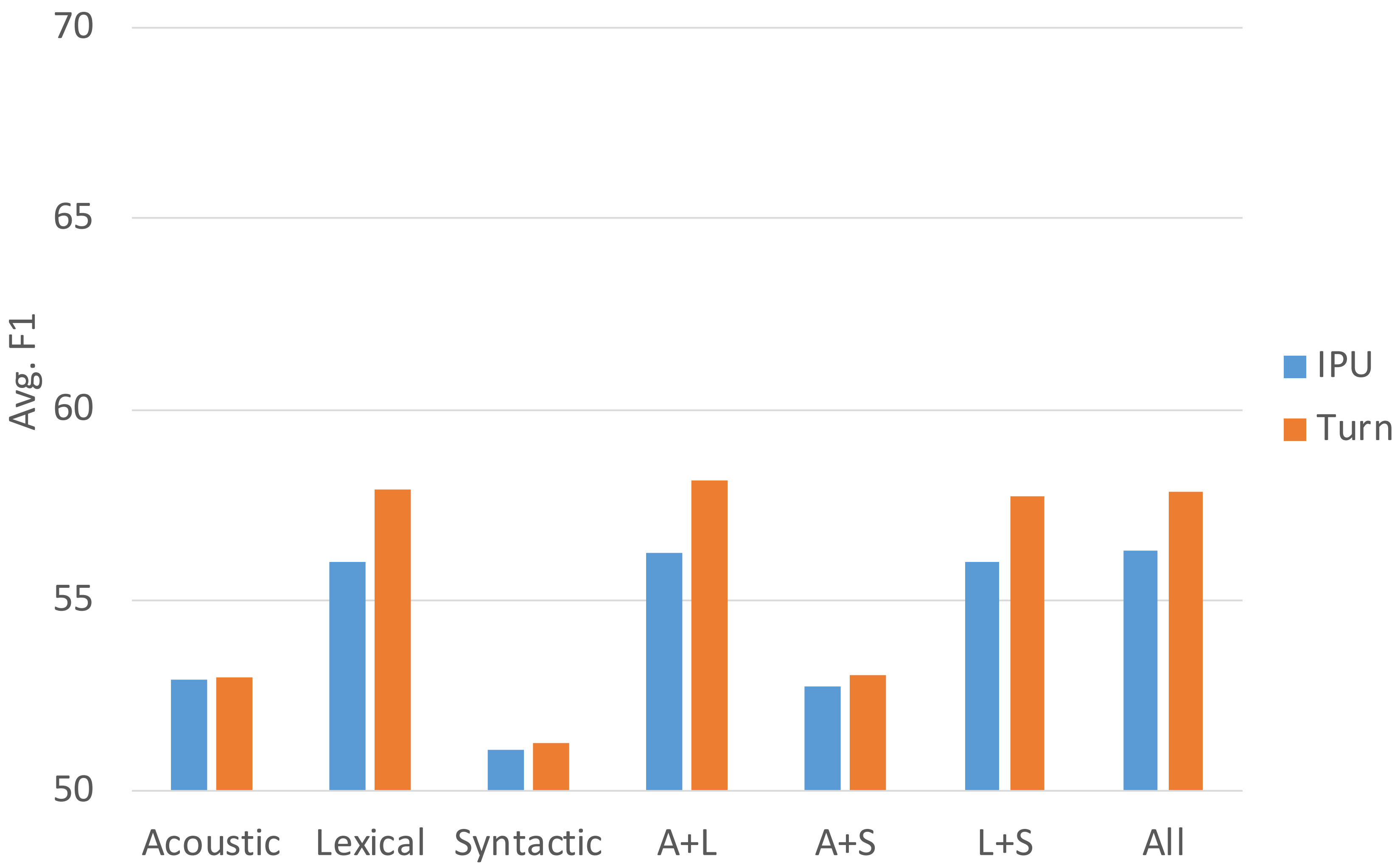
Feature selection

- selectKBest – ANOVA F-value

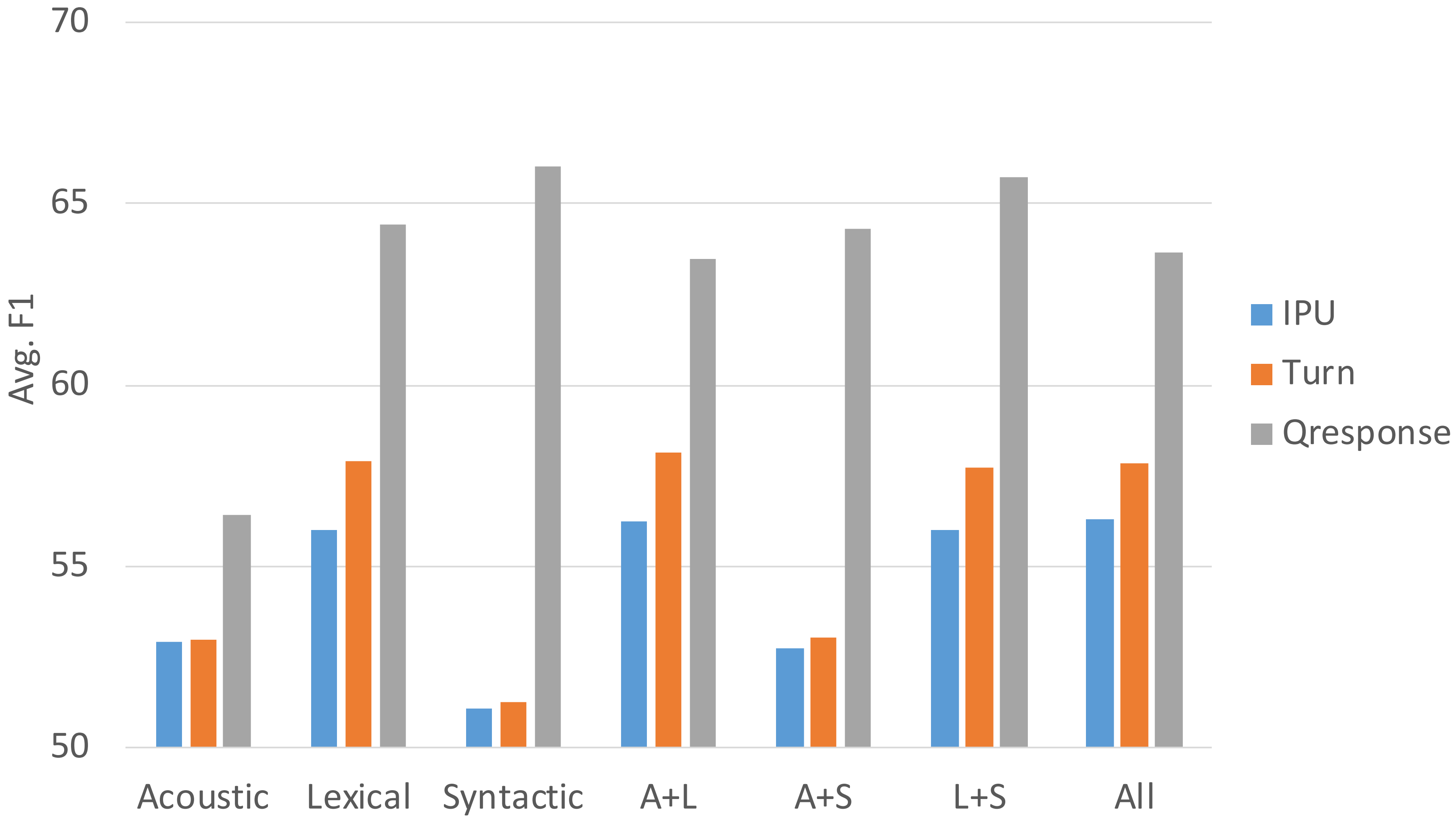
Combined features



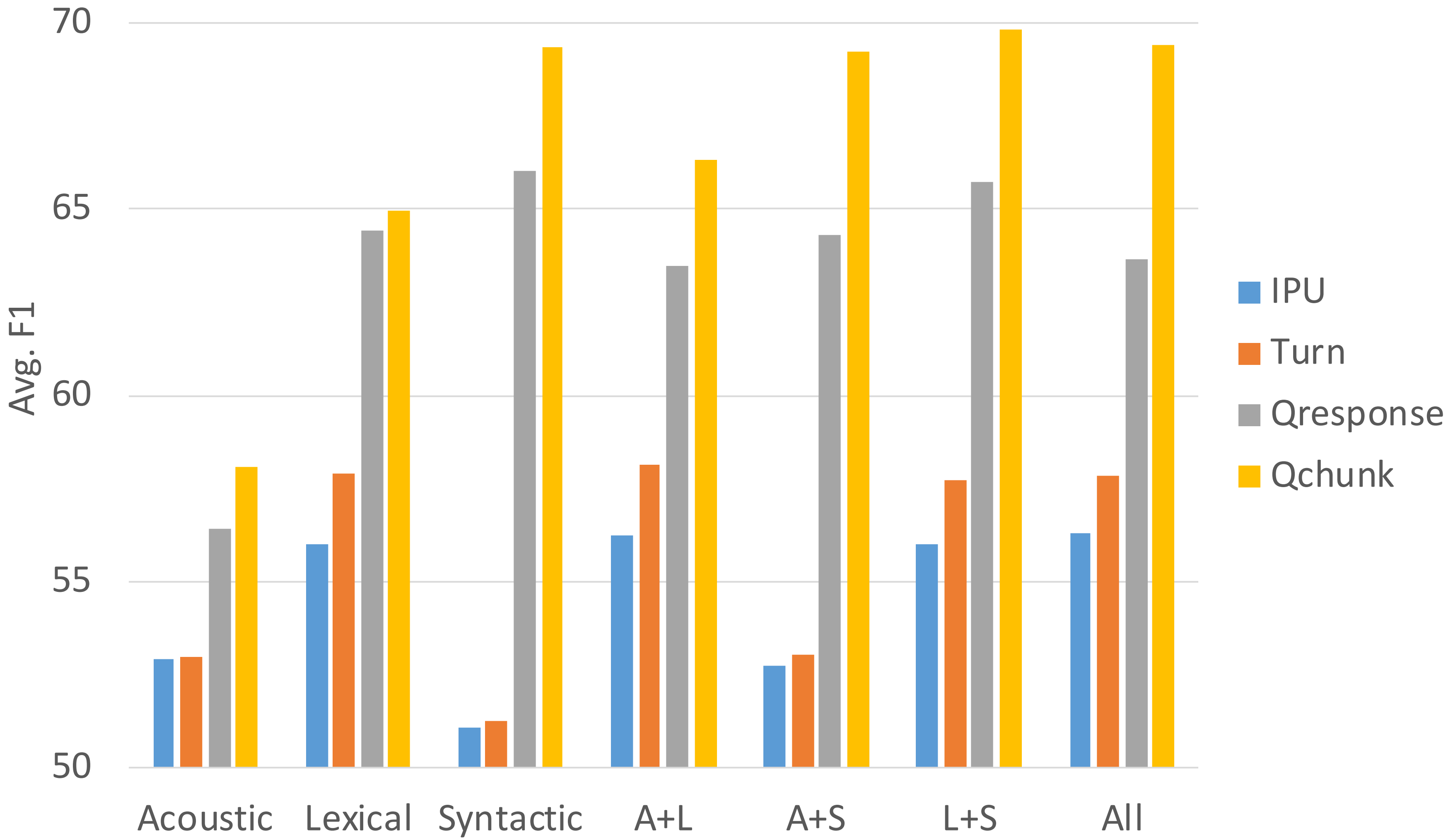
Combined features



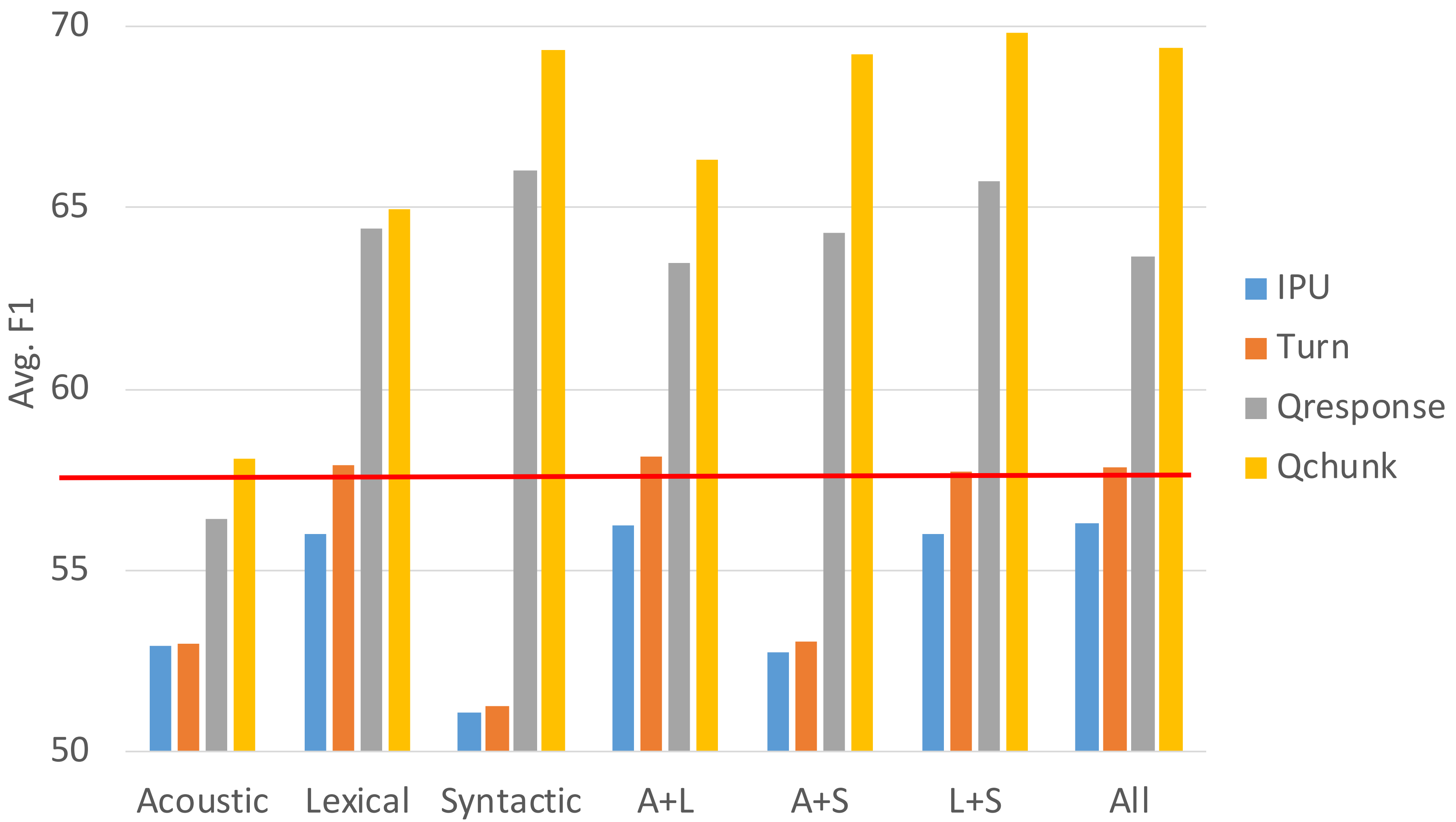
Combined features



Combined features



Combined features



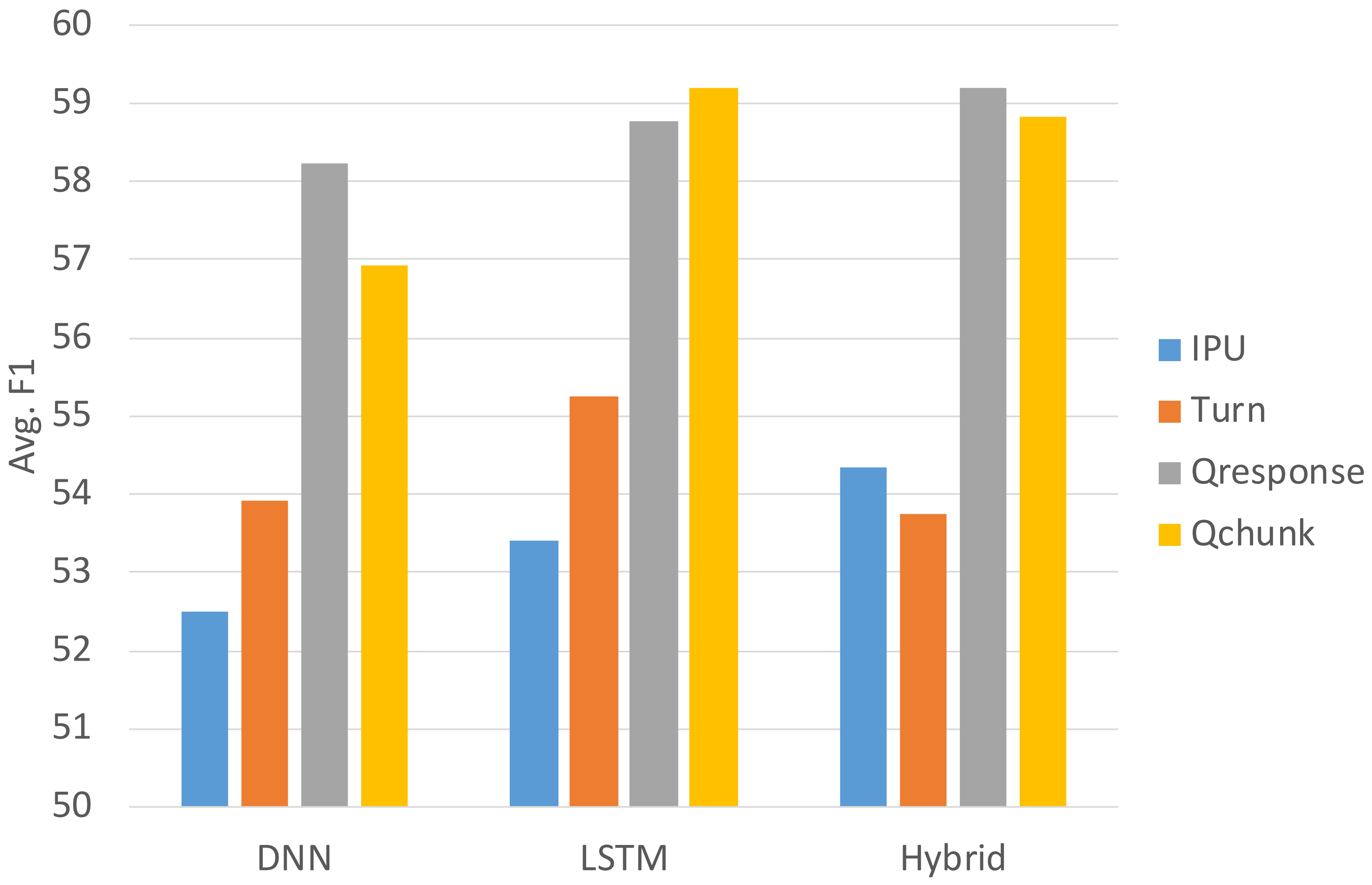
Neural network models

DNN – IS09 openSMILE features

LSTM – GloVe word embeddings

Hybrid – DNN+LSTM

Neural network models



Human vs. machine performance

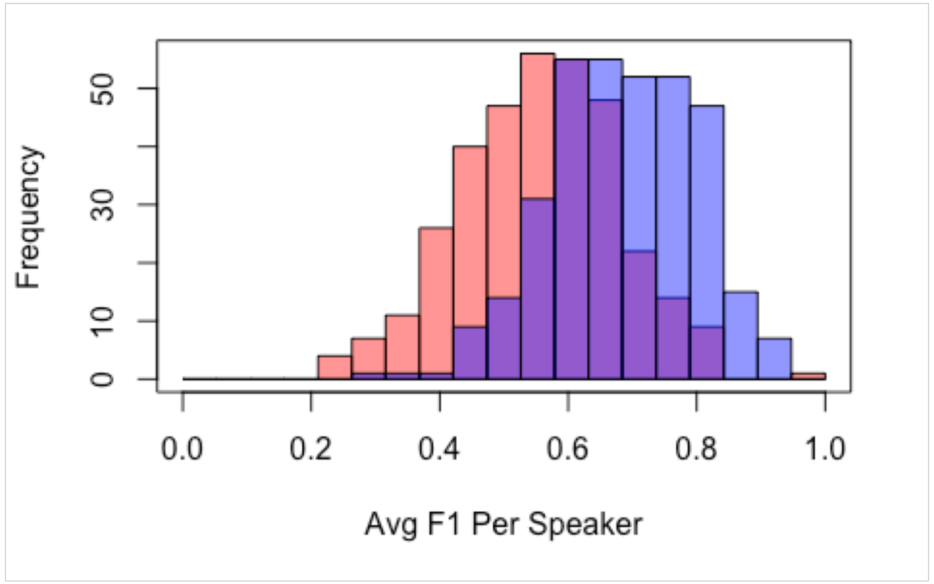
Are there particular groups of **speakers** that are easier/harder to judge?

- Gender, native language, personality

Are there particular kinds of **segments** that are easier/harder to judge?

- Duration, question type

Human vs. machine performance



Judge	Mean	SD	Min	25%	50%	75%	Max
CLF	68.48	11.17	31.25	60.79	69.42	77.22	91.66
Human	55.33	12.50	22.57	46.67	54.17	64.26	100.00

Are classifier and human judgments related?

Speaker-level – not related

$r(340)=-0.02$, $p=0.73$

Segment-level – strongly related

Human vs. machine **judgments** $X^2(1, N=7772) = 94.65$, $p \approx 0$

Human vs. machine **performance** $X^2(1, N=7772) = 32.17$, $p \approx 0$

What kinds of **speakers** are easy/hard to classify?

No effect of **gender** or **native language**

Significant effect of **Conscientiousness** on classifier performance:

$F(2,337)=3.99$, $p=0.02$

Classifier performed better at detecting deception for speakers who were **low** in Conscientiousness

What kinds of **segments** are easy/hard to classify?

Response characteristics:

duration, follow-up questions

question number, question type

Segment characteristics

Judge	Feature	Judgments				Performance			
		t	df	p	Sig.	t	df	p	Sig.
CLF	Duration Follow-up	35.48 27.48	7772 7772	0 0	*** ***	1.09 1.31	7772 7772	0.28 0.18	NS NS
Human	Duration Follow-up	6.19 5.19	7772 7772	0 0	*** ***	0.43 0.09	7772 7772	0.67 0.93	NS NS

Segment characteristics

Judge	Feature	Judgments				Performance			
		t	**df**	**p**	**Sig.**	**t**	**df**	**p**	**Sig.**
CLF	Duration Follow-up	35.48 27.48	7772 7772	0 0	*** ***	1.09 1.31	7772 7772	0.28 0.18	NS NS
Human	Duration Follow-up	6.19 5.19	7772 7772	0 0	*** ***	0.43 0.09	7772 7772	0.67 0.93	NS NS

Segment characteristics

Judge	Feature	Judgments				Performance			
		t	df	p	Sig.	t	df	p	Sig.
CLF	Duration Follow-up	35.48 27.48	7772 7772	0 0	*** ***	1.09 1.31	7772 7772	0.28 0.18	NS NS
Human	Duration Follow-up	6.19 5.19	7772 7772	0 0	*** ***	0.43 0.09	7772 7772	0.67 0.93	NS NS

Human vs. machine performance per question

Strong correlation: $r(24) = 0.69$, $p=0.0002$

Easy

- Question #5: Have your parents divorced?
- Question #13: Have you ever gotten into trouble with the police?
- Question # 16: What is the most you have ever spent on a pair of shoes?

Hard

- Question #8: Have you ever stayed overnight in the hospital as a patient?

Easy for classifier, **hard** for humans

- Question #6: Have you ever broken a bone?

Summary: deception detection

Trained automatic classifiers for deception detection

- **IPU** 56.3 F1 (LR acoustic+lexical+syntactic)
- **Turn** 58.1 F1 (LR acoustic+lexical)
- **Question response** 66 F1 (SVM syntactic)
- **Question chunk** 69.8 F1 (NB lexical+syntactic)

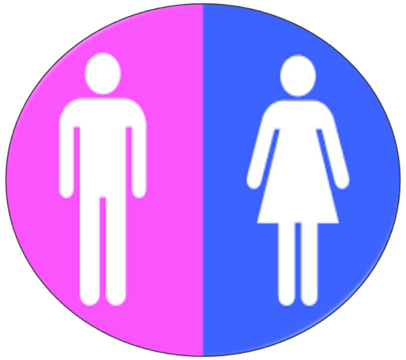
Human vs. machine performance

- Significant variation across speakers and segments
- Human and CLF judgments correlated at the segment level, not speaker level
- Speaker traits: slight effect of personality (C-score) on CLF performance
- Segment characteristics: **duration**, **question number**, **question type** affect human and classifier judgments and performance

Individual differences in deceptive behavior

Research goals:

1. Identify differences in cues to deception across **gender**, **native language,** and **personality**
2. Leverage speaker differences in deception classification



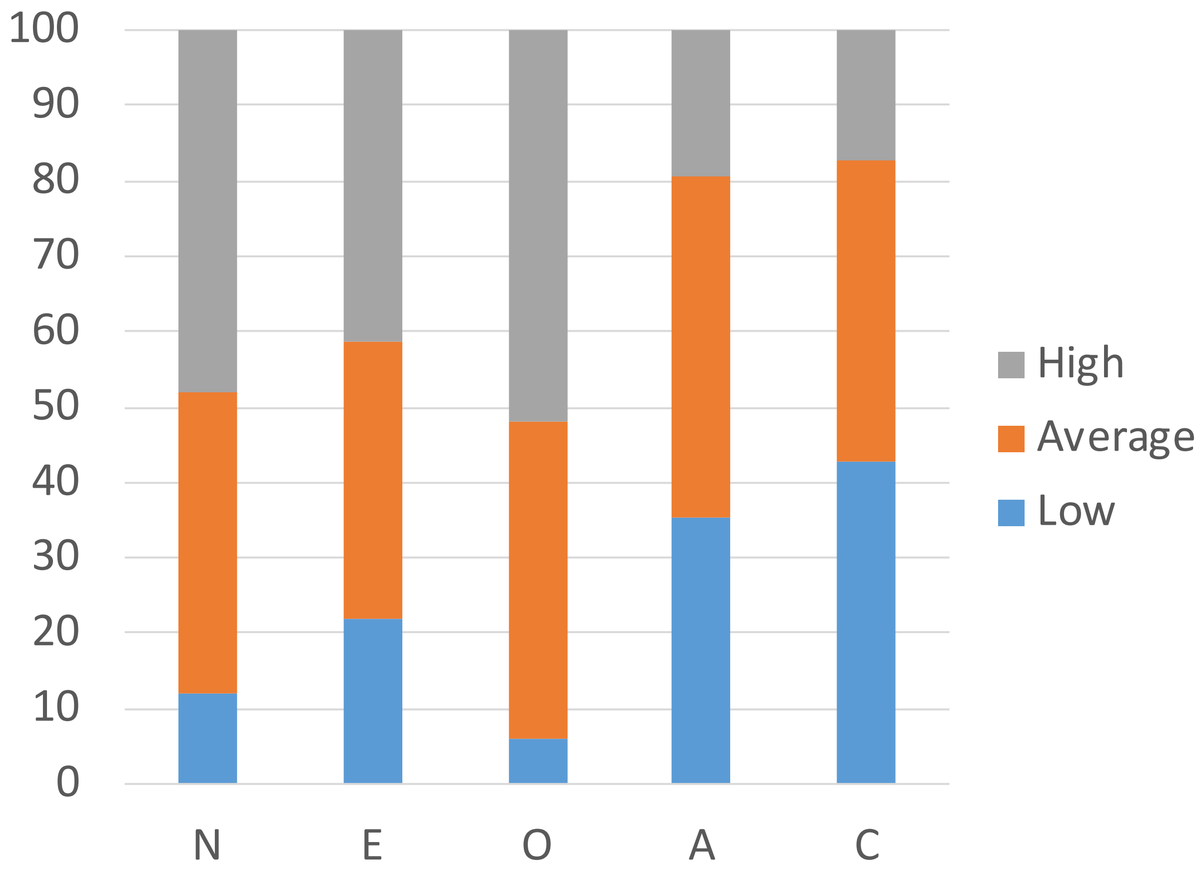
Male vs. female cues to deception

	Male	Female
Acoustic	Pitch mean	Intensity mean
LDI	**DAL.imagery** hasWe **hasNot** hasAbsReally numHedgePhrases **hasFalseStart** hasHedgePhrase	**DAL.wc** **hasContraction**
LIWC	Conj Pronoun **Focuspast** **Relativ** nonflu **space** **Prep**	netspeak Adj **allPunc** **Apostro**
Complexity	**W** **CN** **CT.T** **DC** **MLS** **CP.T** **CT** **DC.C** **CP.C** **CP** **DC.T** **CN.T**	CN.C

Native English vs. Mandarin cues to deception

	English			Mandarin	
Acoustic	**Intensity mean**		**jitter** shimmer	Pitch mean	**Speaking rate**
LDI	**hasHedgePhrase** **hasI** hasLaugh numLaugh **thirdPersonPronouns**		**DAL.wc** **hasCuePhrase** **hasNot**	**hasFalseStart** **hasYes**	
LIWC	Adverb **Conj** I **netspeak** **Nonflu**	**Posemo** **Ppron** **Pronoun** **Social** **Tone**	**Focuspresent**	**space**	**Cogproc**
Complexity	**DC** **CT** **CP**	**CN** **DC.C** **DC.T**	**CT.T** **CN.T** **CN.C**		

Personality bin distribution



$$\mathrm{TFdiff}_f = \frac{\sum_{s_i \in F} f(s_i)}{size_F} - \frac{\sum_{s_i \in T} (s_i)}{size_T}$$

Personality differences in cues to deception

	N Neuroticism	E Extroversion	O Openness	A Agreeableness	C Conscientiousness
Acoustic	Intensity max	shimmer			NHR
LDI	specScores	hasWe	hasYes isJustYes numFilledPauses specScores	specificDenial	
LIWC	Authentic relativ space	focuspast	work	informal	
Complexity	**VP** **C.S** **C** VP.T DC **C.T** MLT DC.T				

Classification leveraging speaker differences

Three approaches:

1. Classification with individual traits as features
2. Classification with homogenous data
3. Classification with speaker dependent features

Classification experiments

Generic: session features

Speaker-dependent: session – baseline features

Combined: generic + speaker-dependent features

Summary: speaker differences

Identified group differences in cues to deception

- **Gender** male speakers increased pitch mean when lying, female speakers increased intensity mean when lying
- **Native language** signs of increased cognitive load for native Chinese speakers; complexity features only useful for native English speakers
- **Personality** most differences for Neuroticism

Classification leveraging speaker differences

- Speaker-dependent features may improve performance

Trust


Mail
Phone
What can I help you with?

NISSAN
#AutonomousDrive
Autonomous Drive
NISSAN


Acoustic-Prosodic Indicators of Deception and Trust in Interview Dialogues

What are the acoustic-prosodic characteristics of **truthful** and **deceptive** speech?

What are the acoustic-prosodic characteristics of **trusted** and **mistrusted** speech?

Are there universal characteristics and/or **individual differences** in production and perception of deception?

Can we **automatically classify** deceptive speech using acoustic-prosodic features?

Truthful vs. Deceptive Interviewee Responses

Deception/Mistrust Truth/Trust

Feature	Deception	Trust
Pitch Max		
Pitch Mean		
Intensity Max		
Intensity Mean		
Speaking Rate		
Jitter		
Shimmer		
NHR		

Gender and Native Language Analysis

Deception **Truth**

Feature	Male	Female	English	Chinese	All
Pitch Max					
Pitch Mean					
Intensity Max					
Intensity Mean					
Speaking Rate					
Jitter					
Shimmer					
NHR					

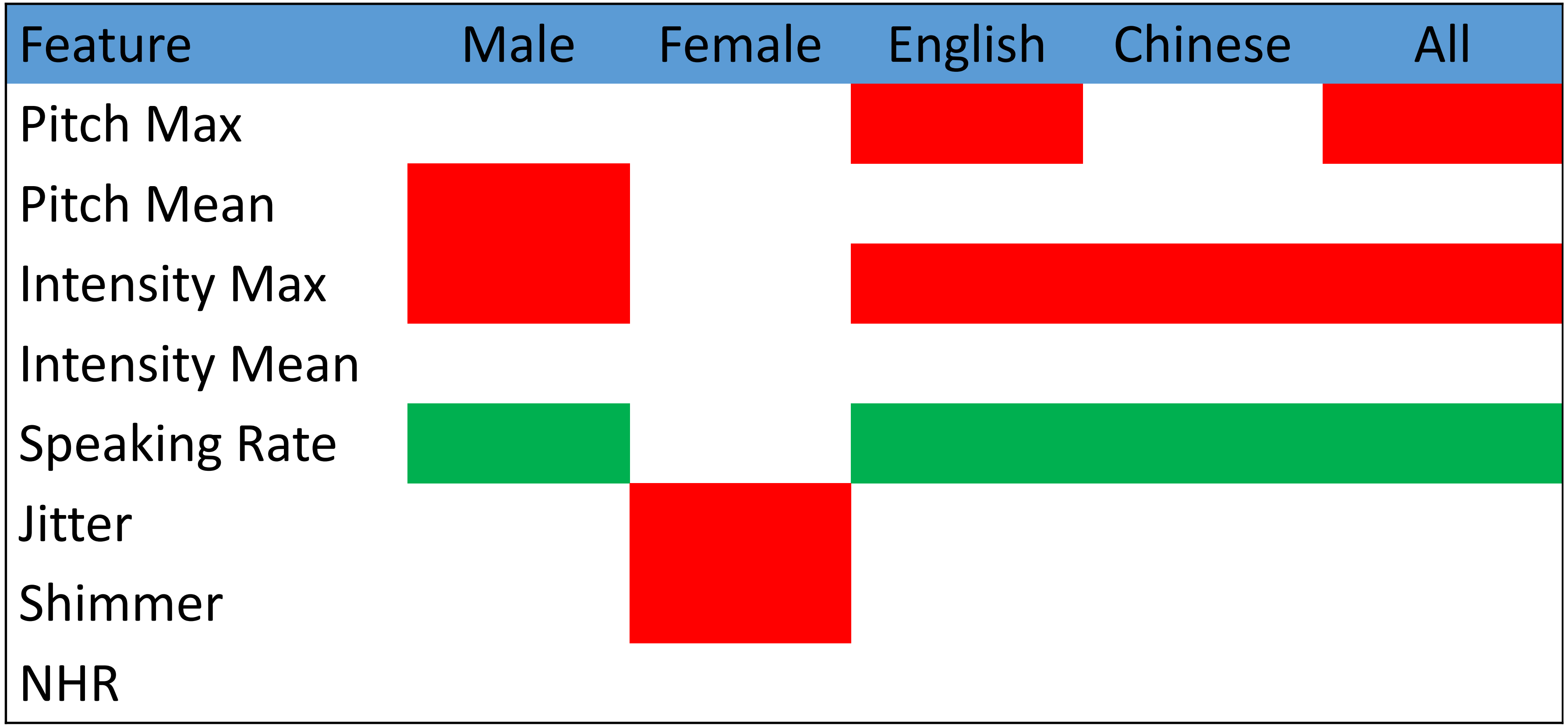
Gender and Native Language: Analysis of **Interviewee** Traits

Mistrusted **Trusted**

Feature	Male	Female	English	Chinese	All
Pitch Max					
Pitch Mean					
Intensity Max					
Intensity Mean					
Speaking Rate					
Jitter					
Shimmer					
NHR					

Gender and Native Language: Analysis of **Interviewer** Traits

Mistrusted **Trusted**



Questions

- Can we use this information to:
  - Automatically detect trustworthy speech?
  - Create trustworthy synthesized speech?

Speech Corpus Annotation

- Experts
- Crowdsourcing
- Games with a purpose (GWAP)

Games With A Purpose (GWAP)

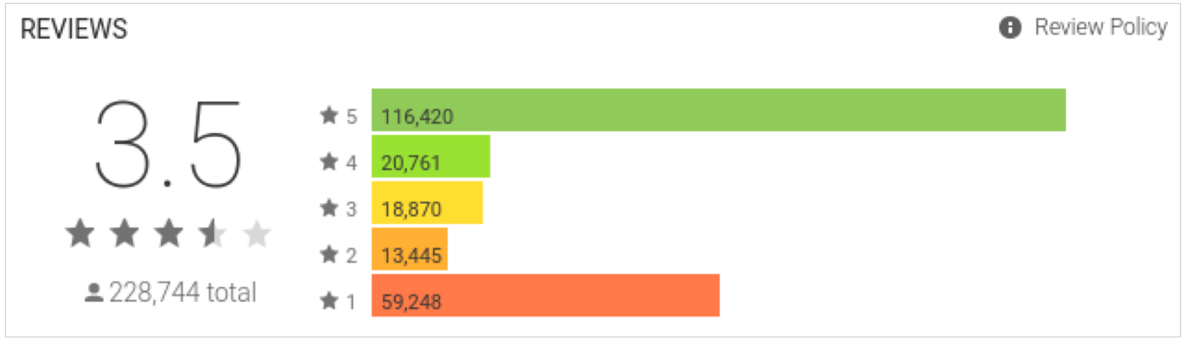
Idea: Motivate people to solve computational problems by presenting the problem as a series of simple steps in an enjoyable game format.

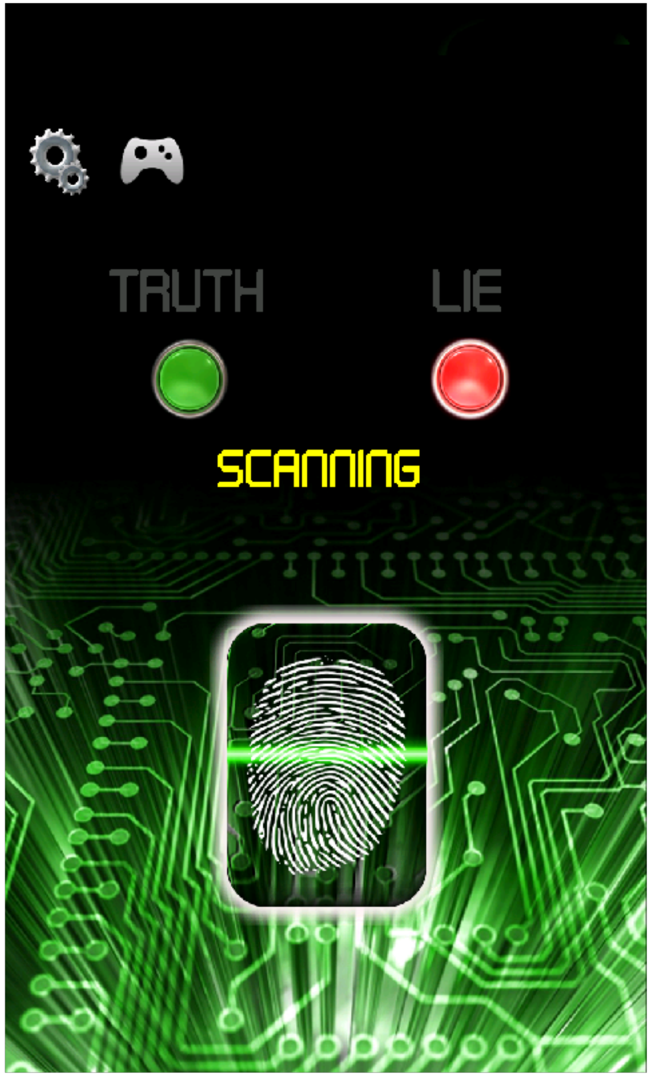
GWAP Advantages

- More engaging format than monotonous annotation tasks
- Built-in incentives
- Affordable/free annotation
- Easy to distribute, accessible

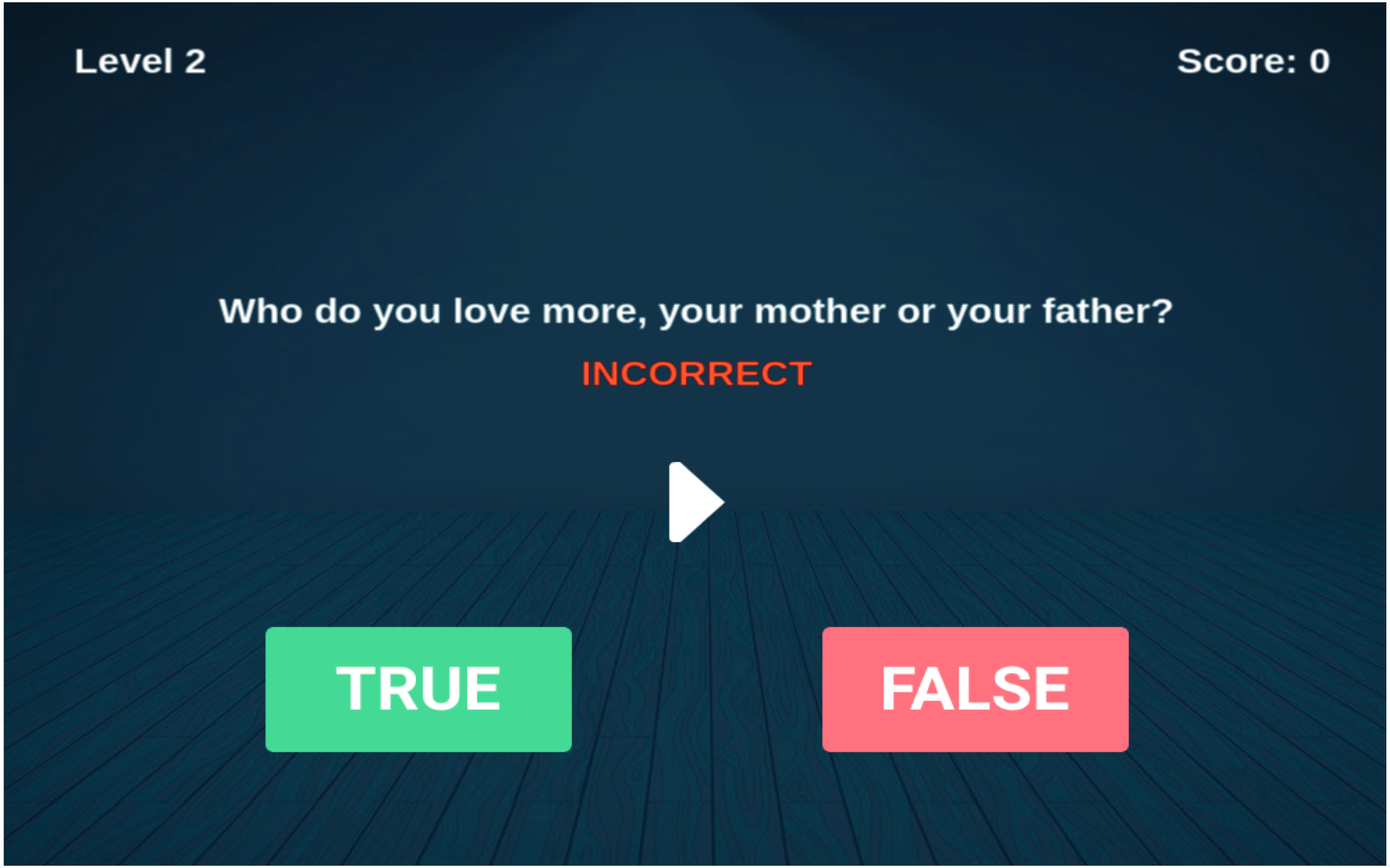
Lie Detection Games

- Amusing to assess lie detection ability
- Popular lie detection app:
  “Lie Detector Simulator Fun”
  - 10,000,000+ installs on Google Play alone
  - 3.5/5 rating, 228,744 reviews






LieCatcher
Start
Rules


Level 2
Score: 0
Who do you love more, your mother or your father?
INCORRECT
TRUE
FALSE

Pilot Study

- Early feedback about game design
- 40 student participants
- Pre and post game surveys
- 2 levels – with and without instant feedback
- Quality control questions

Survey Responses

- Positive feedback!
  - 85% found game easy to use
  - 75% might or would definitely recommend to friend
  - 73% preferred level 2 – with instant feedback
  - 70% liked the premise of the game

Player Behavior

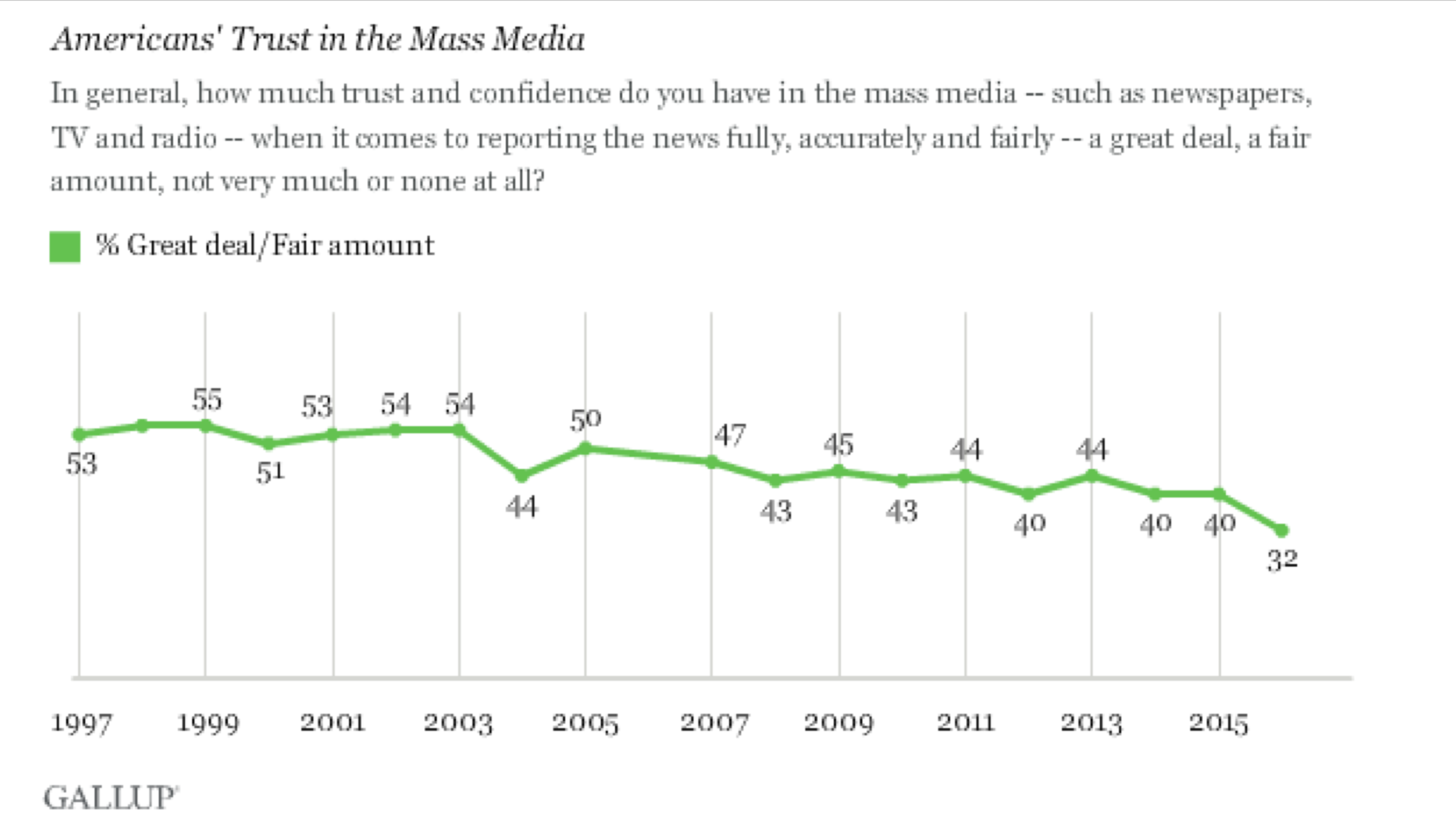
- Player accuracy: 49.86%
  - Level 1: 45.66%
  - Level 2: 54.44%
- 100% correct answer for quality control questions
- Some questions were “easier” than others
- Some samples were more “trusted” than others
  - But no clear consensus on “mistrusted” segments
- Gender differences

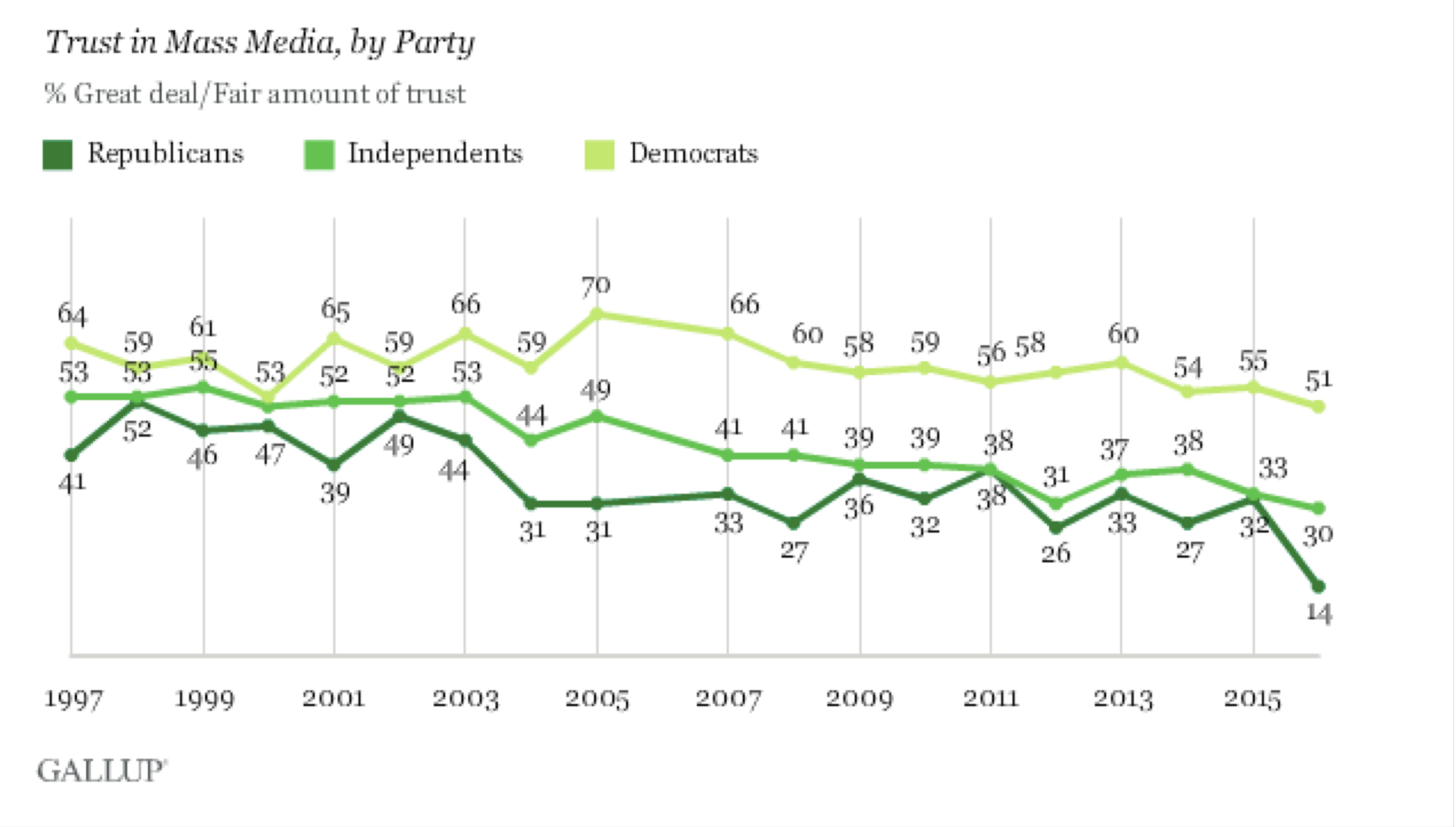
Ongoing work

- Incorporate feedback from pilot study
- Distribute game to wide audience (initial study with Amazon Mechanical Turk)
- Study acoustic-prosodic characteristics of trustworthy speech

Linguistic Indicators of Trust in Media

Americans’ trust in media is at an all-time low



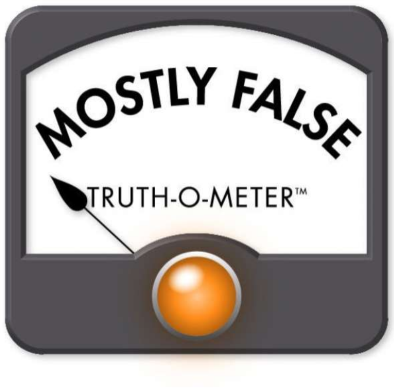

Trust in Mass Media, by Party
% Great deal/Fair amount of trust
Republicans
Independents
Democrats
64 59 61 53 65 59 66 59 70 66 60 58 59 56 58 60 54 55 51
53 53 55 52 52 53 44 49 41 41 39 39 38 31 37 38 33 30
41 52 46 47 39 49 44 31 31 33 27 36 32 38 26 33 27 32 14
1997 1999 2001 2003 2005 2007 2009 2011 2013 2015
GALLUP

It is difficult to distinguish between real and fake news stories

- People often believe and spread fake news
- People question and mistrust accurate reports

Fake news detection

- Fact-checking data (Potthast et al., 2017; Wang, 2017)
- Crowd-sourced data (Perez-Rosas et al., 2017)
- Satirical vs. legitimate news (Rubin et al., 2016)



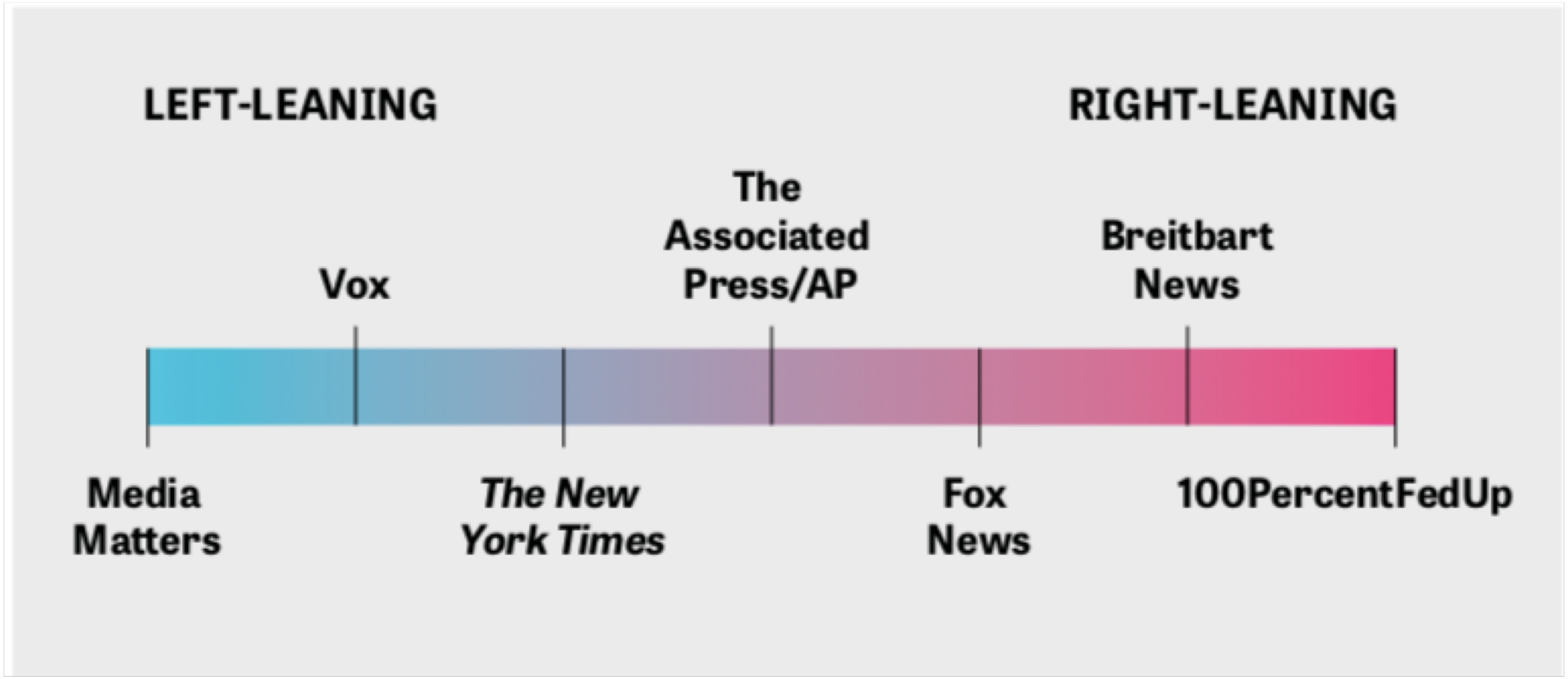
Our Goal: Trusted News Detection

- What are the linguistic characteristics of **trustworthy** and **untrustworthy** news?
- Are there differences in perception of trust across **demographic** groups?

Trusted News Corpus

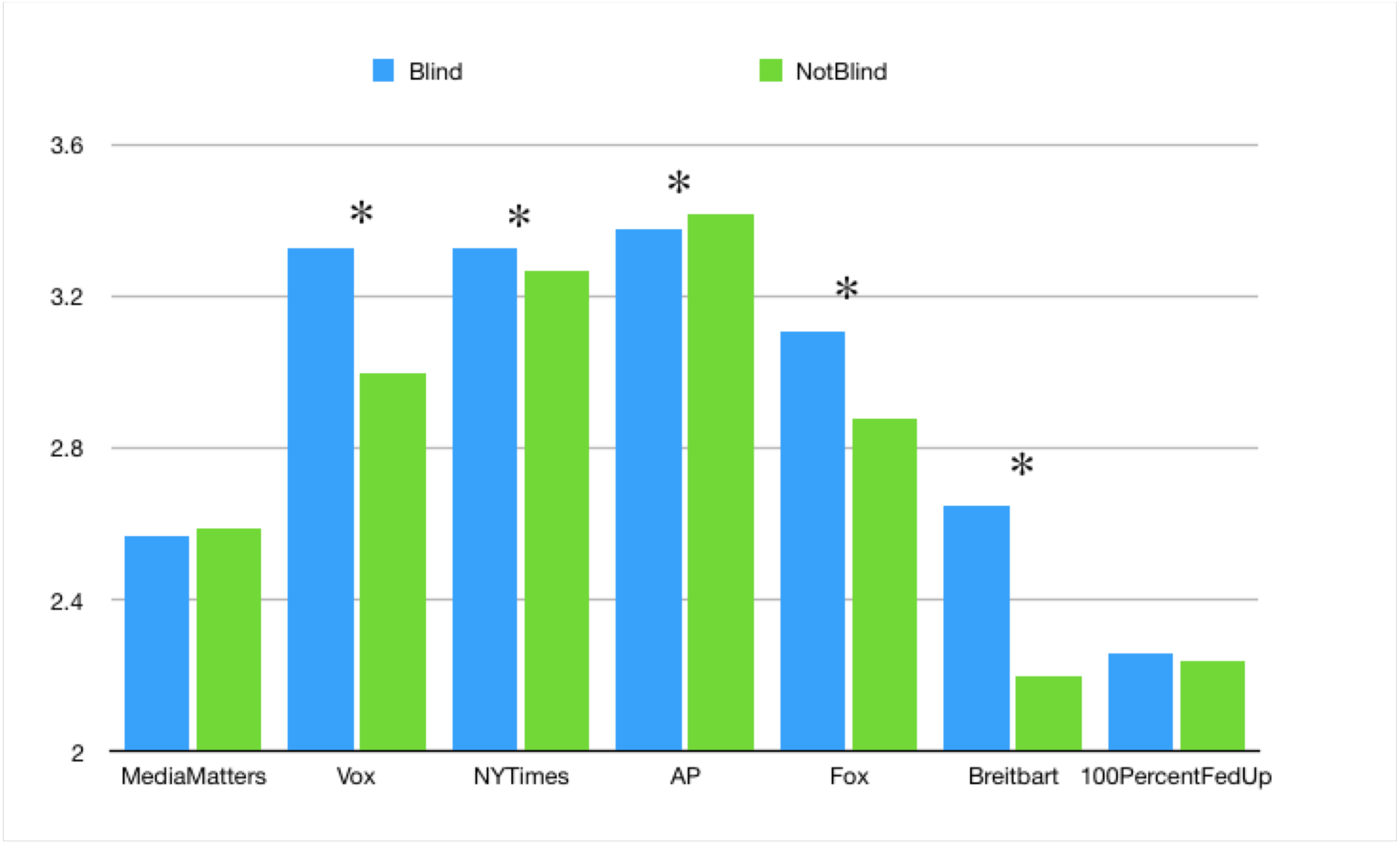
- Built by Gallup an Knight Foundation
- Online experimental platform, participants recruited from the Gallup Panel
- Trust ratings on a scale from 1-5
- Blind (B) vs. Not-blind (NB) conditions

Trusted News Corpus



Trusted News Corpus

- 1,914 news articles
- Categories: politics, economics, science
- 3,420 readers
- 66,597 judgments


Blind
NotBlind
3.6
3.2
2.8
2.4
2
*
*
*
*
*
MediaMatters
Vox
NYTimes
AP
Fox
Breitbart
100PercentFedUp

Features

- LIWC
  - Custom lexicons: hedge, bias (Recasens et al., 2013)
- Syntactic complexity
- Dictionary of Affect in Language
- N-grams

Characteristics of Trusted News

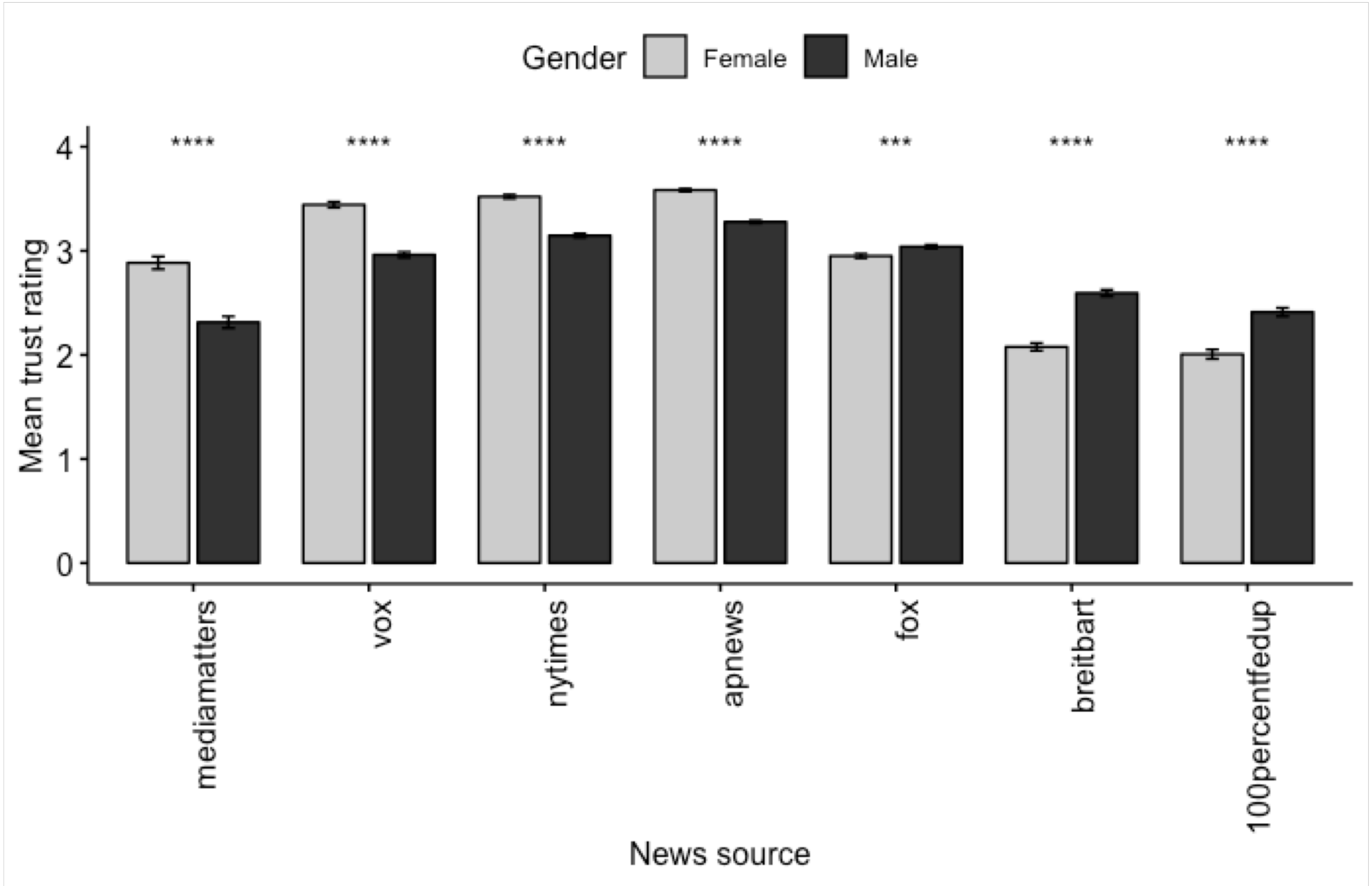
- Correlate linguistic features with trust ratings
- Analyze blind ratings only
- Headline vs. body

- BREAKING NEWS: NFL Reporter Says Colin Kaepernick Will STAND For National Anthem If NFL Team Will Give Him A Job
- Nations to work on curbing climate change despite Trump
- Canadian Professor Says 'Romanticized' Concept of Debate Breeds Anti-Abortion Activists
- Commander of 1st flight of space shuttle Challenger dies
- Paul Singer-Funded Washington Free Beacon Behind Initial Fusion GPS Trump Effort
- Domestic Abusers Are Barred From Gun Ownership, but Often Escape the Law

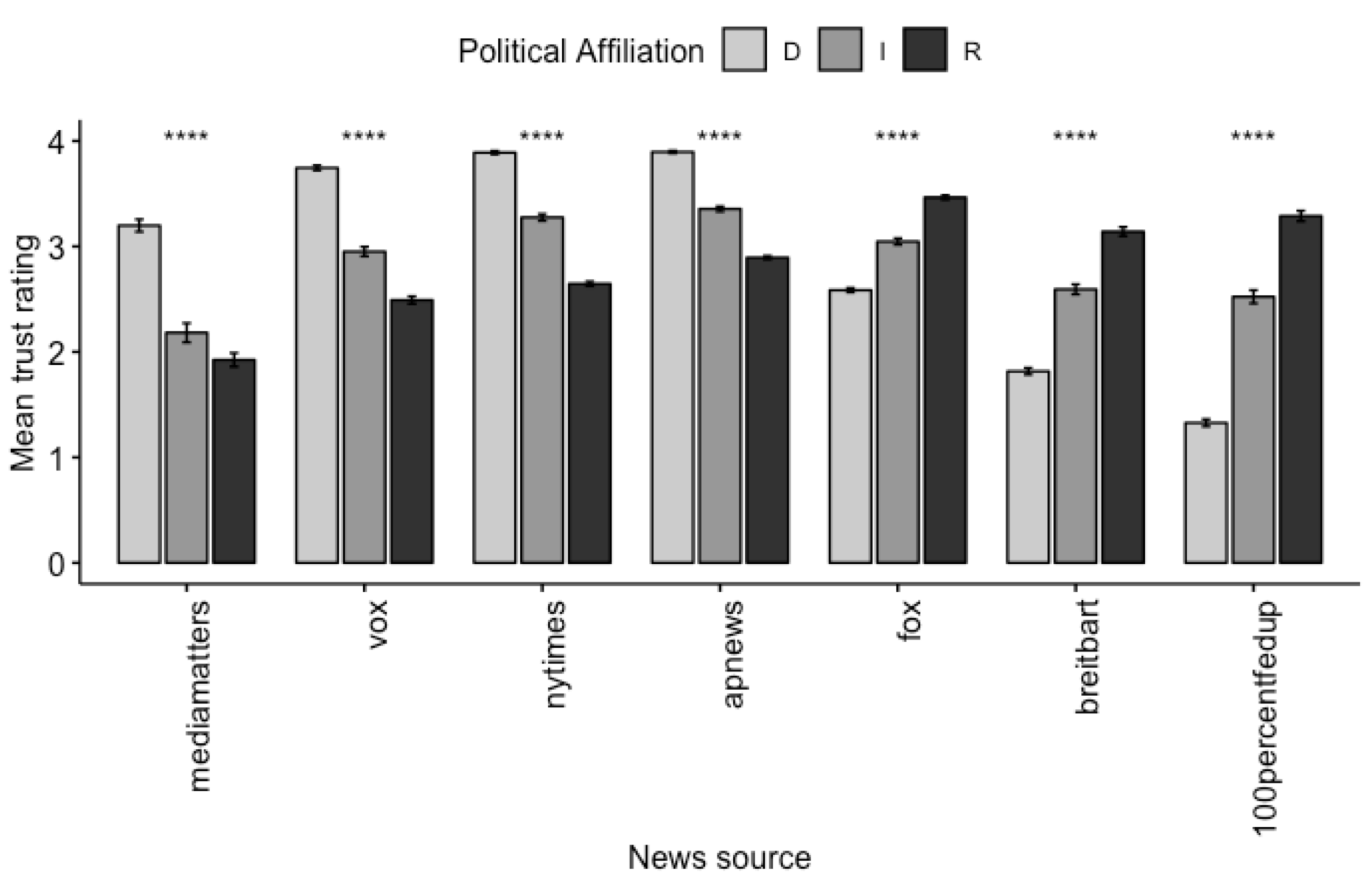
- BREAKING NEWS: NFL Reporter Says Colin Kaepernick Will STAND For National Anthem If NFL Team Will Give Him A Job
- Nations to work on curbing climate change despite Trump
- Canadian Professor Says ‘Romanticized’ Concept of Debate Breeds Anti-Abortion Activists
- Commander of 1st flight of space shuttle Challenger dies
- Paul Singer-Funded Washington Free Beacon Behind Initial Fusion GPS Trump Effort
- Domestic Abusers Are Barred From Gun Ownership, but Often Escape the Law

Headline analysis

Features	Trusted	Mistrusted
LIWC (*custom)	Dic, we, ipron, prep, interrog, sad, insight, bio, body, health, ingest, achieve, reward, relativ, motion, space, time, home, death, Comma, SemiC	WC, WPS, ppron, shehe, they, auxverb, verb, bias*, social, friend, female, male, discrep, certain, differ, percept, see, hear, feel, sexual, focuspresent, leisure, money, relig, informal, netspeak, AllPunc, Period, Colon, Exclam, Quote, Apostro, Parenth, OtherP
Complexity		W, S, VP, C, T, DC, CT, CP, CN, MLS, MLT, MLC, C.S, VP.T, C.T, DC.C, DC.T, CT.T, CP.T, CP.C, CN.T, CN.C
DAL	pleasant, activate, imagery	
N-gram	puerto rico, white house, wants to, in puerto, we know, know about, to know, paris climate, need to, public health, on trump, here s what, in august, things to, opioid epidemic, end of, a new	tax reform, tax cuts, steve bannon, tax cut, hillary clinton, donald trump, fox news, fake news, tax bill, social media, republican tax, president trump, in virginia, gun control, to keep, against trump, sexual misconduct, fox friends, tried to, breaking news, new jersey, on tax, senate candidate



Gender
Female
Male
Mean trust rating
0
1
2
3
4
****
****
****
****
***
****
****
mediamatters
vox
nytimes
apnews
fox
breitbart
100percentfedup
News source


Political Affiliation
D
I
R
Mean trust rating
0
1
2
3
4
****
mediamatters
vox
nytimes
apnews
fox
breitbart
100percentfedup
News source

	Democrat	Republican
Trusted	**LIWC** WC, WPS, percept, see, risk, focuspast, time, home **Complexity** W, VP, C, DC, CT, CP, CN, MLS, MLT, MLC, C.S, VP.T, C.T, DC.C, DC.T, CT.T, CP.T, CN.T **DAL** activate **N-gram** the white house, in a statement, the associated press, the university of, part of the, going to be, at the university, is going to, the first time, in puerto rico, said in a statement, at the university of	**LIWC** Sixltr, We, They, posemo **Complexity** CP.C **DAL** pleasant **N-gram** one of the, a lot of, in the us, the end of, the washington post, at the time, the department of, in order to, president of the, is expected to
Mistrusted	**LIWC** they, conj, sad, cogproc, cause, discrep, sexual, power, reward, focusfuture, money, nonflu **N-gram** one of the, according to the, be able to, at the time, in order to, the end of the	**LIWC** WPS, tentat, differ, percept, hear, risk **Complexity** MLS, MLT, C.S, VP.T, C.T, DC.C, DC.T, CT.T, CN.T **DAL** activate **N-gram** president donald trump, the trump administration, the new york, the federal government, the first time, the new york times

Summary

- Deception detection from text and speech
- Characteristics of trustworthy speech
- Game framework for annotating trust
- Trust in news media

Thank you!

Julia Hirschberg
Michelle Levine
Andrew Rosenberg
Guozhen An
Bingyan Hu
Gideon Mendels
Angel Maredia
Jessica Xiang
Kai-Zhan Lee
Zoe Baker-Peng

James Shin
Ivy Chen
Meredith Cox
Gauri Narayan
Mandi Wang
Leighanne Hsu
Yocheved Levitan
Grace Ulinski
Arthur Shen
Nishi Cestero

Jennifer Senior
Yvonne Missry
Molly Scott
William Wang