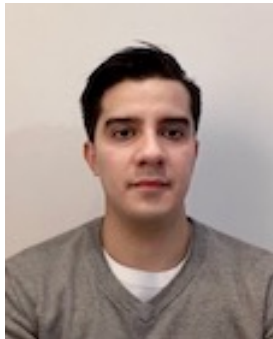# Differential Bias:

On the Perceptibility of Stance Imbalance in Argumentation
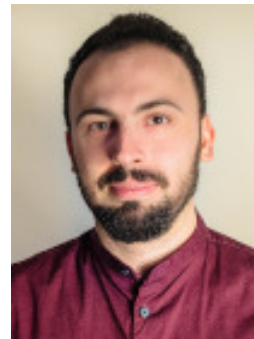
Alonso Palomino   Martin Potthast   Khalid Al-Khatib   Benno Stein

Leipzig University

University of Groningen

Bauhaus-Universität Weimar

Assessing Bias

Is Text X Biased?

Text X on "Tidal Energy"
Many forms of energy have potential environmental impacts. The focus should be on designing regulations that minimize or eliminate these impacts. Regulations can minimize the environmental impact of tidal energy.
Bays and estuaries are always naturally "flushed" or cleansed and replenished by tide-waters flowing in and out. To the extent that tidal energy impairs the natural flows of these tides, it impairs this natural "flushing" mechanism. This can alter and even destroy an ecosystem.
Few studies that have been undertaken to date to identify the environmental impacts of a tidal power scheme, which have determined that each specific site is different and the impacts depend greatly upon local geography. It is thus crucial to judge the potential for tidal energy on a case by case basis.
Offshore turbines do not alter the flow of tides as much as barrages can, so their environmental impact is smaller. Offshore turbines don't hamper the flow of tides.

Assessing Bias

Is Text X Biased?

What is bias? [Merriam Webster]

❑ Deviation from the expectation

❑ Prejudice against so. or sth.

Text X on "Tidal Energy"
Many forms of energy have potential environmental impacts. The focus should be on designing regulations that minimize or eliminate these impacts. Regulations can minimize the environmental impact of tidal energy.
Bays and estuaries are always naturally "flushed" or cleansed and replenished by tide-waters flowing in and out. To the extent that tidal energy impairs the natural flows of these tides, it impairs this natural "flushing" mechanism. This can alter and even destroy an ecosystem.
Few studies that have been undertaken to date to identify the environmental impacts of a tidal power scheme, which have determined that each specific site is different and the impacts depend greatly upon local geography. It is thus crucial to judge the potential for tidal energy on a case by case basis.
Offshore turbines do not alter the flow of tides as much as barrages can, so their environmental impact is smaller. Offshore turbines don't hamper the flow of tides.

Assessing Bias

Is Text X Biased?

What is bias? [Merriam Webster]

- ❏ Deviation from the expectation
- ❏ Prejudice against so. or sth.

What is Text X? [Dutilh Novaes 2021]

- ❏ Argumentation:
  Communication involving arguments
- ❏ Argument: Conclusion supported by premises
- ❏ Goal: Deliberate, convince, persuade

Text *X* on "Tidal Energy"
Many forms of energy have potential environmental impacts. The focus should be on designing regulations that minimize or eliminate these impacts. Regulations can minimize the environmental impact of tidal energy.
Bays and estuaries are always naturally "flushed" or cleansed and replenished by tide-waters flowing in and out. To the extent that tidal energy impairs the natural flows of these tides, it impairs this natural "flushing" mechanism. This can alter and even destroy an ecosystem.
Few studies that have been undertaken to date to identify the environmental impacts of a tidal power scheme, which have determined that each specific site is different and the impacts depend greatly upon local geography. It is thus crucial to judge the potential for tidal energy on a case by case basis.
Offshore turbines do not alter the flow of tides as much as barrages can, so their environmental impact is smaller. Offshore turbines don't hamper the flow of tides.

Assessing Bias

Is Text X Biased?

What is bias? [Merriam Webster]

- ❑ Deviation from the expectation
- ❑ Prejudice against so. or sth.

What is Text X? [Dutilh Novaes 2021]

- ❑ Argumentation:
  Communication involving arguments
- ❑ Argument: Conclusion supported by premises
- ❑ Goal: Deliberate, convince, persuade

How can argumentation be biased?

- ❑ Omission or one-sidedness: [Walton 1999]
  - – Positions or stances (pro, con)
  - – Frames (e.g., economics)
- ❑ Inclusion or entailment [Recasens 2013]
  - – Contradicting accepted truths / untruths.
  - – Framing (e.g., implict stereotypes)

Text X on "Tidal Energy"
Many forms of energy have potential environmental impacts. The focus should be on designing regulations that minimize or eliminate these impacts. Regulations can minimize the environmental impact of tidal energy.
Bays and estuaries are always naturally "flushed" or cleansed and replenished by tide-waters flowing in and out. To the extent that tidal energy impairs the natural flows of these tides, it impairs this natural "flushing" mechanism. This can alter and even destroy an ecosystem.
Few studies that have been undertaken to date to identify the environmental impacts of a tidal power scheme, which have determined that each specific site is different and the impacts depend greatly upon local geography. It is thus crucial to judge the potential for tidal energy on a case by case basis.
Offshore turbines do not alter the flow of tides as much as barrages can, so their environmental impact is smaller. Offshore turbines don't hamper the flow of tides.

Assessing Bias

Is Text X Biased?

Text X on "Tidal Energy"
Many forms of energy have potential environmental impacts. The focus should be on designing regulations that **pro** or eliminate these impacts. Regu can minimize the environmental impact of tidal energy.
Bays and estuaries are always naturally "flushed" or cleansed and replenished by tide-waters flowing in and out. To the extent that tidal energy impairs th **con** natural flows of these tides, it impairs this natural "flushing" mechanism. This can alter and even destroy an ecosystem.
Few studies that have been undertaken to date to identify the environmental impacts of a tidal power scheme, which have determined that each speci **pro** different and the impacts depend greatly upon local geography. It is thus crucial to judge the potential for tidal energy on a case by case basis.
Offshore turbines do not alter the flow of tides as much as barrages can, s **pro** environmental impact is smaller. turbines don't hamper the flow of tides.

Assessing Bias

Is Text Y Biased?

Text Y on "Tidal Energy"
Anchors of tidal energy systems can damage ecosystems. They might damage salt marshes, estuaries, and near-shore reefs, or alter the pro... that maintain ocean and coastal ecosystems, such as the movement of sand, silt, animals, and larvae. **con**
Barrages can protect coastal ecosystems from a storm. This more than ma... for minor damage tidal energy ma... marine ecosystems. **pro**
While there may still be some concerns for the safety and local environmental impact of nuclear energy, nuclear energy releases 0-emissions and so is a... important part of the fight against global warming. It should be pushed forward; tidal energy should not "replace" it. **con**
There is only one major tidal station in operation. This is a 240 megawatt station at the mouth of the La Rance river estuary on the northern coast of ... (a large coal or nuclear power plant generates about 1,000 MW of electricity). Tidal energy generates very little energy. **con**

Assessing Bias

Is Text Y Biased?

Absolute bias classification:

- ❑ Identification of complex language phenomena

- ❑ Handling subjectivity of assessors

Text Y on "Tidal Energy"
Anchors of tidal energy systems can damage ecosystems. They might damage salt marshes, estuaries, and near-shore reefs, or alter the pro... that maintain ocean and coastal ecosystems, such as the movement of sand, silt, animals, and larvae. **con**
Barrages can protect coastal ecosystems from a storm. This more than ma... for minor damage tidal energy ma... marine ecosystems. **pro**
While there may still be some concerns for the safety and local environmental impact of nuclear energy, nuclear energy releases 0-emissions and so is a... important part of the fight against global warming. It should be pushed forward; tidal energy should not "replace" it. **con**
There is only one major tidal station in operation. This is a 240 megawatt station at the mouth of the La Rance river estuary on the northern coast of... (a large coal or nuclear power plant generates about 1,000 MW of electricity). Tidal energy generates very little energy. **con**

Assessing Bias

Is Text Y Biased?

Absolute bias classification:

- ❑ Identification of complex language phenomena
- ❑ Handling subjectivity of assessors

Sources of subjectivity:

- ❑ Sociocultural background
- ❑ Environmental conditioning
- ❑ Educational attainment
- ❑ Partisanship

Text Y on "Tidal Energy"
Anchors of tidal energy systems can damage ecosystems. They might damage salt marshes, estuaries, and near-shore reefs, or alter the proc that maintain ocean and coastal ecosystems, such as the movement of sand, silt, animals, and larvae. **con**
Barrages can protect coastal ecosystems from a storm. This more than ma for minor damage tidal energy ma marine ecosystems. **pro**
While there may still be some concerns for the safety and local environmental impact of nuclear energy, nuclear energy releases 0-emissions and so is a important part of the fight against global warming. It should be pushed forward; tidal energy should not "replace" it. **con**
There is only one major tidal station in operation. This is a 240 megawatt station at the mouth of the La Rance river estuary on the northern coast of (a large coal or nuclear power plant generates about 1,000 MW of electricity). Tidal energy generates very little energy. **con**

Assessing Bias

Is Text Y Biased?

Absolute bias classification:

- ❏ Identification of complex language phenomena
- ❏ Handling subjectivity of assessors

Sources of subjectivity:

- ❏ Sociocultural background
- ❏ Environmental conditioning
- ❏ Educational attainment
- ❏ Partisanship

Is a relative bias assessment feasible?

➜ Differential Bias

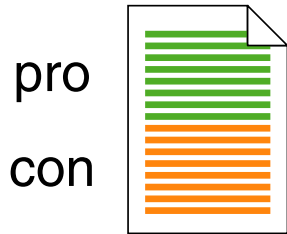| Text Y on "Tidal Energy" |
|---|
| Anchors of tidal energy systems can damage ecosystems. They might damage salt marshes, estuaries, and near-shore reefs, or alter the proc... that maintain ocean and coastal ecosystems, such as the movement of sand, silt, animals, and larvae. **con** |
| Barrages can protect coastal ecosystems from a storm. This more than ma... for minor damage tidal energy ma... marine ecosystems. **pro** |
| While there may still be some concerns for the safety and local environmental impact of nuclear energy, nuclear energy releases 0-emissions and so is a... important part of the fight against global warming. It should be pushed forward; tidal energy should not "replace" it. **con** |
| There is only one major tidal station in operation. This is a 240 megawatt station at the mouth of the La Rance river estuary on the northern coast of... (a large coal or nuclear power plant generates about 1,000 MW of electricity). Tidal energy generates very little energy. **con** |

Differential Bias

Is Text X More Biased?

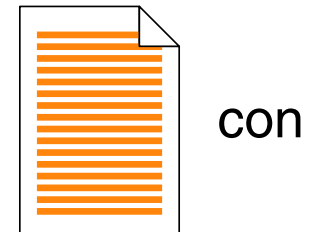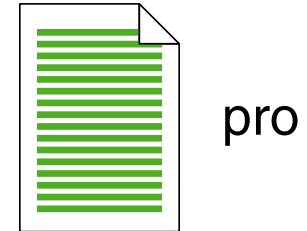| Text X on "Tidal Energy"	Text Y on "Tidal Energy"
Many forms of energy have potential environmental impacts. The focus should be on designing regulations that [...] or eliminate these impacts. Regu[...] can minimize the environmental impact of tidal energy. **pro**	Anchors of tidal energy systems can damage ecosystems. They might damage salt marshes, estuaries, and near-shore reefs, or alter the proc[...] that maintain ocean and coastal ecosystems, such as the movement of sand, silt, animals, and larvae. **con**
Bays and estuaries are always naturally "flushed" or cleansed and replenished by tide-waters flowing in and out. To the extent that tidal energy impairs th[...] natural flows of these tides, it impairs this natural "flushing" mechanism. This can alter and even destroy an ecosystem. **con**	Barrages can protect coastal ecosystems from a storm. This more than ma[...] for minor damage tidal energy ma[...] marine ecosystems. **pro**
Few studies that have been undertaken to date to identify the environmental impacts of a tidal power scheme, which have determined that each speci[...] different and the impacts depend greatly upon local geography. It is thus crucial to judge the potential for tidal energy on a case by case basis. **pro**	While there may still be some concerns for the safety and local environmental impact of nuclear energy, nuclear energy releases 0-emissions and so is a[...] important part of the fight against global warming. It should be pushed forward; tidal energy should not "replace" it. **con**
Offshore turbines do not alter the flow of tides as much as barrages can, s[...] environmental impact is smaller. [...] turbines don't hamper the flow of tides. **pro**	There is only one major tidal station in operation. This is a 240 megawatt station at the mouth of the La Rance river[...] estuary on the northern coast of [...] (a large coal or nuclear power plant generates about 1,000 MW of electricity). Tidal energy generates very little energy. **con**

Differential Bias

Is Text X More Biased?

Text X on "Tidal Energy"	Text Y on "Tidal Energy"
Many forms of energy have potential environmental impacts. The focus should be on designing regulations that or eliminate these impacts. Regu can minimize the environmental impact of tidal energy. **pro**	Anchors of tidal energy systems can damage ecosystems. They might damage salt marshes, estuaries, and near-shore reefs, or alter the pro that maintain ocean and coastal ecosystems, such as the movement of sand, silt, animals, and larvae. **con**
Bays and estuaries are always naturally "flushed" or cleansed and replenished by tide-waters flowing in and out. To the extent that tidal energy impairs th natural flows of these tides, it impairs this natural "flushing" mechanism. This can alter and even destroy an ecosystem. **con**	Some fishing communities worry that their fishing grounds could be dis Toxic hydraulic fluids might leak. energy may disrupt local fishing industry. **con**
Tidal energy can slow the movement of water in a bay or estuary, which reduces the amount of kinetic energy and causes the body of water to freeze-over often or for longer periods of time has consequences for marine eco-systems.Tidal energy can lead to undesirable prolonged winter icing. **con**	While there may still be some concerns for the safety and local environmental impact of nuclear energy, nuclear energy releases 0-emissions and so is a important part of the fight against global warming. It should be pushed forward; tidal energy should not "replace" it. **con**
Offshore turbines do not alter the flow of tides as much as barrages can, s environmental impact is smaller. turbines don't hamper the flow of tides. **pro**	There is only one major tidal station in operation. This is a 240 megawatt station at the mouth of the La Rance river estuary on the northern coast of (a large coal or nuclear power plant generates about 1,000 MW of electricity). Tidal energy generates very little energy. **con**

Differential Bias
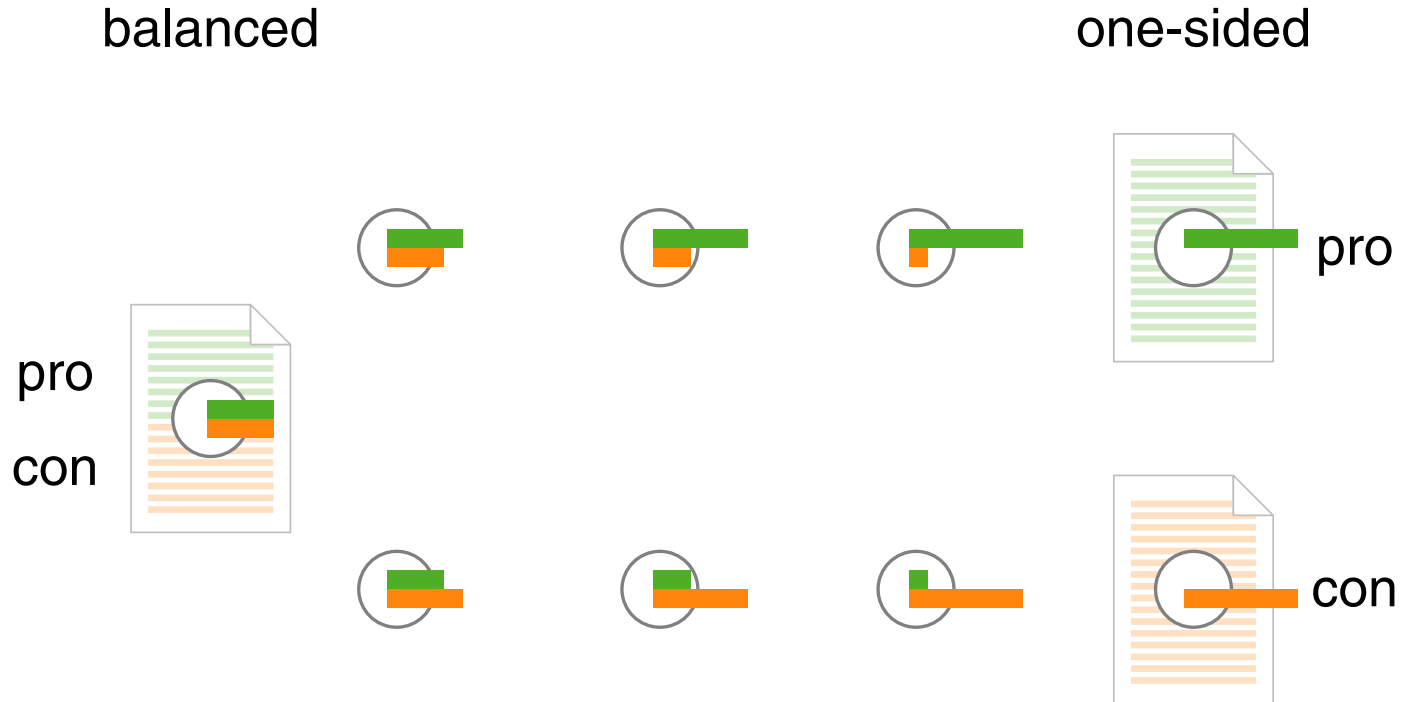## Example: Stance Bias

balanced   one-sided

pro

con

pro

con

❑ Perfect stance balance and maximal stance one-sidedness are extremes.

Differential Bias
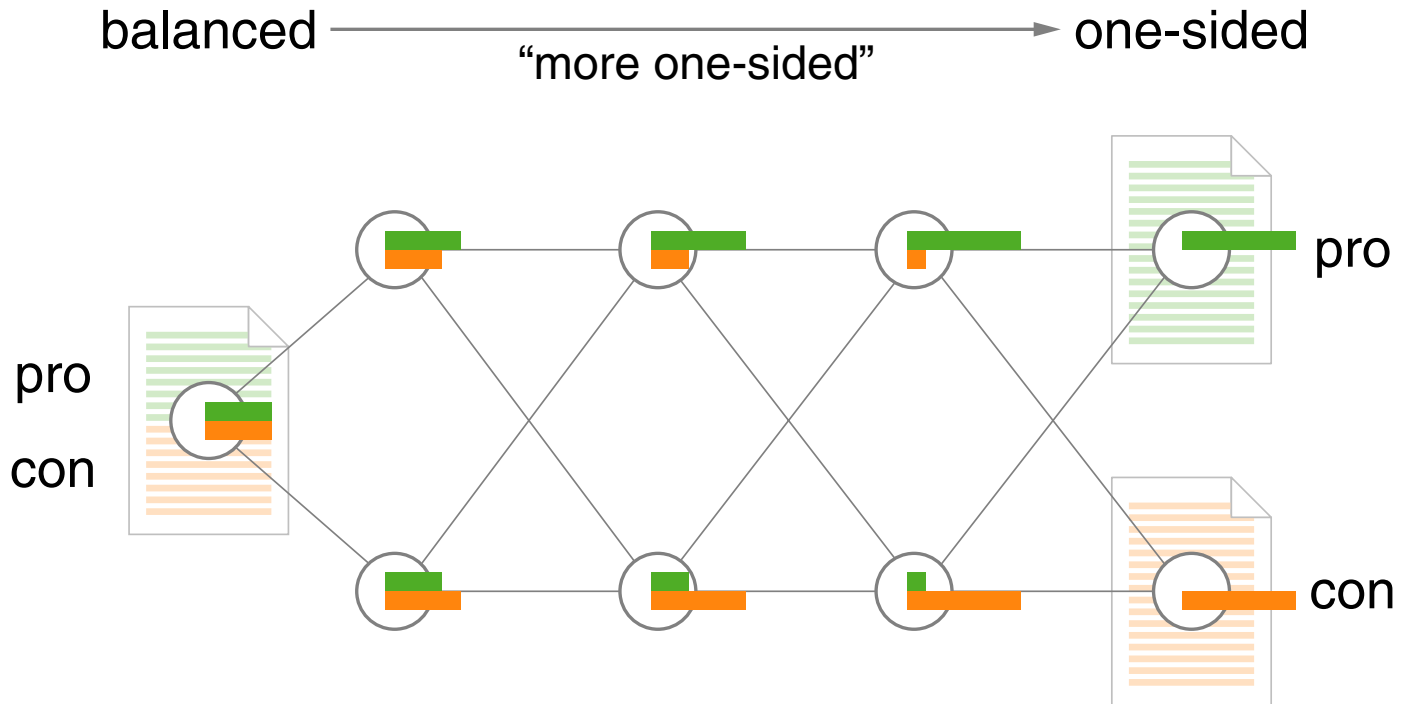
Example: Stance Bias

balanced   one-sided

pro
con

pro

con

❏ Perfect stance balance and maximal stance one-sidedness are extremes.

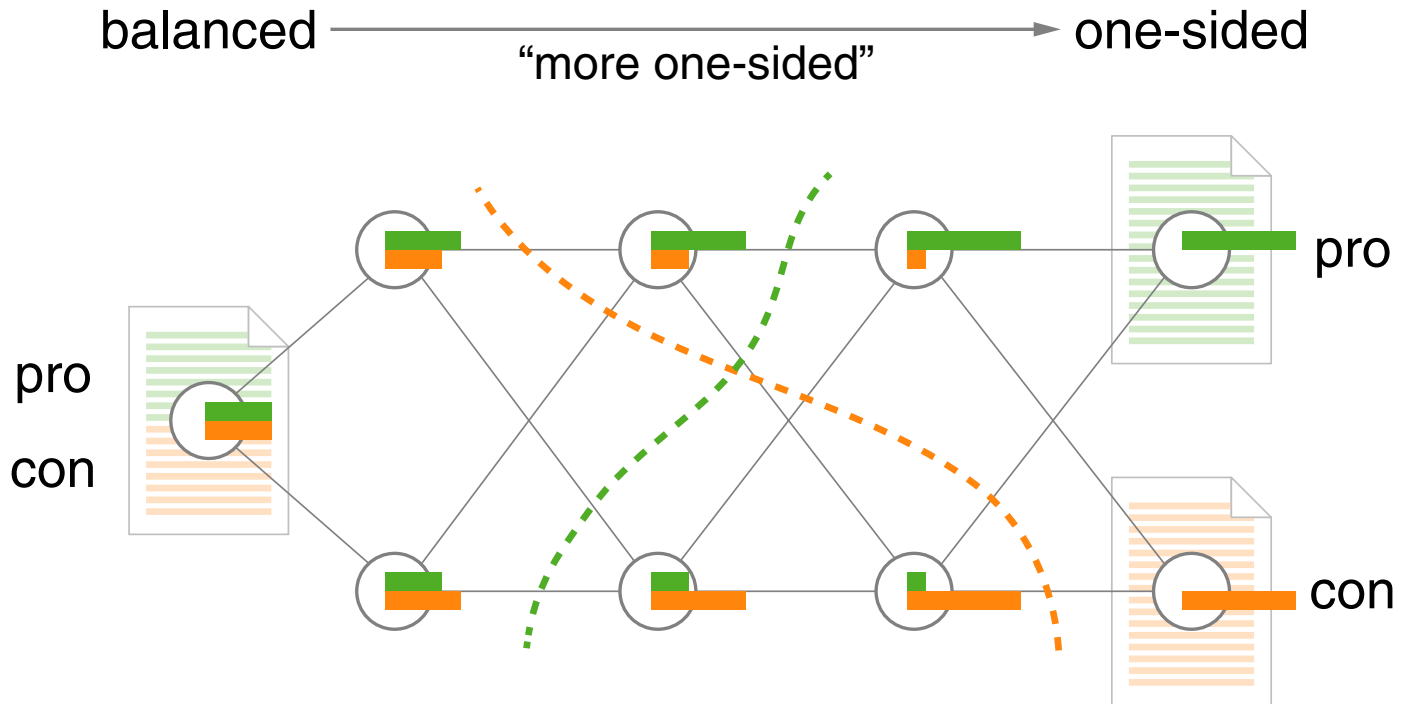❏ A spectrum of stance distributions are encountered in argumentations.

Differential Bias

Example: Stance Bias

balanced ———————————————→ one-sided
"more one-sided"

pro
con

pro

con

- Perfect stance balance and maximal stance one-sidedness are extremes.
- A spectrum of stance distributions are encountered in argumentations.
- A measure of stance balance induces a partial order (preference relation).

Differential Bias
Example: Stance Bias

balanced ⟶ one-sided
"more one-sided"

pro con

pro

con

❏ Perfect stance balance and maximal stance one-sidedness are extremes.

❏ A spectrum of stance distributions are encountered in argumentations.

❏ A measure of stance balance induces a partial order (preference relation).

❏ Subjective decision rules capture personal bias acceptance thresholds.

Differential Bias
Measuring Stance Balance

Definition 1 **(Stance Balance, Differential Stance Bias)**

Let X denote an argumentation, where $X = X^+ \cup X^-$ is the union of a set of pro arguments X^+ and con arguments X^-, so that $X^+ \cap X^- = \emptyset$.

The stance balance δ between pro and con arguments in X is measured as the absolute size difference between X^+ and X^-:

$$\delta(X) = \left| |X^+| - |X^-| \right|.$$

The closer δ is to 0, the more balanced the two sides are in X.

Given a second argumentation Y, $|X| = |Y|$, the differential stance bias of X and Y can then be quantified as

$$\Delta(X, Y) = |\delta(X) - \delta(Y)|.$$

The closer Δ is to 0, the less strong the differential bias is between X and Y.

Differential Bias

Measuring Stance Balance

Definition 1 **(Stance Balance, Differential Stance Bias)**

Let X denote an argumentation, where $X = X^+ \cup X^-$ is the union of a set of pro arguments X^+ and con arguments X^-, so that $X^+ \cap X^- = \emptyset$.

The stance balance δ between pro and con arguments in X is measured as the absolute size difference between X^+ and X^-:

$$\delta(X) = \big||X^+| - |X^-|\big|.$$

The closer δ is to 0, the more balanced the two sides are in X.

Given a second argumentation Y, $|X| = |Y|$, the differential stance bias of X and Y can then be quantified as

$$\Delta(X, Y) = |\delta(X) - \delta(Y)|.$$

The closer Δ is to 0, the less strong the differential bias is between X and Y.

Differential Bias
Measuring Stance Balance

Definition 1 (**Stance Balance, Differential Stance Bias**)

Let X denote an argumentation, where $X = X^+ \cup X^-$ is the union of a set of pro arguments X^+ and con arguments X^-, so that $X^+ \cap X^- = \emptyset$.

The stance balance δ between pro and con arguments in X is measured as the absolute size difference between X^+ and X^-:

$$\delta(X) = \left| |X^+| - |X^-| \right|.$$

The closer δ is to 0, the more balanced the two sides are in X.

Given a second argumentation Y, $|X| = |Y|$, the differential stance bias of X and Y can then be quantified as

$$\Delta(X, Y) = |\delta(X) - \delta(Y)|.$$

The closer Δ is to 0, the less strong the differential bias is between X and Y.

Differential Bias
Measuring Stance Balance: Example

- Given an argumentation length of $|X| = |Y| = 4$, the image of both δ and Δ is $\{0, 2, 4\}$.

- Top: $\delta(X) = 0$ and $\delta(Y) = 4$ and thus $\Delta(X, Y) = 4$, the maximum possible differential bias in this setting. It follows that Y is more biased than X.

- Middle: $\delta(X) = 2$ and $\delta(Y) = 2$ and thus $\Delta(X, Y) = 0$, the minimum possible differential bias in this setting. Neither text is more biased than the other.

- Bottom: $\delta(X) = 0$ and $\delta(Y) = 2$ and thus $\Delta(X, Y) = 2$, the medium possible differential bias in this setting. It follows that Y is more biased than X.

Perceptibility of Stance Imbalance

Experiment: Design

```
┌─────────────────────────────────────────────┐
│ Subtasks 1/2 (high difficulty)              │
│                                              │
│   ┌──────────────┐  ┌──────────────┐        │
│   │    Text X     │  │    Text Y     │        │
│   ├──────────────┤  ├──────────────┤        │
│   │ Argument 1   │  │ Argument 1   │        │
│   │ Argument 2   │  │ Argument 2   │        │
│   │ Argument 3   │  │ Argument 3   │        │
│   │ Argument 4   │  │ Argument 4   │        │
│   └──────────────┘  └──────────────┘        │
│                                              │
│  (1) Which text is more biased?             │
│  (2) How sure are you?                      │
│  (3) Explain your decision.                 │
└─────────────────────────────────────────────┘
```

Task difficulty: (All: Explanation of bias as one-sidedness.)

❑   High: No mention of stance.

Perceptibility of Stance Imbalance
Experiment: Design

Subtasks 1/2 (high difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Which text is more biased?

(2) How sure are you?

(3) Explain your decision.

Subtask 3 (medium difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Which text has more pro stances?

(2) Which text has more con stances?

(3) Which text is more biased?

(4) How sure are you?

(5) Explain your decision

Task difficulty: (All: Explanation of bias as one-sidedness.)

❑ High: No mention of stance.

❑ Medium: First questions hint at stance.

Perceptibility of Stance Imbalance

Experiment: Design

Subtasks 1/2 (high difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Which text is more biased?

(2) How sure are you?

(3) Explain your decision.

Subtask 3 (medium difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Which text has more pro stances?

(2) Which text has more con stances?

(3) Which text is more biased?

(4) How sure are you?

(5) Explain your decision

Subtask 4 (low difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Label the stances of each argument.

(2) Which text is more biased?

(3) How sure are you?

(4) Explain your decision

Task difficulty: (All: Explanation of bias as one-sidedness.)

- ❑ High: No mention of stance.
- ❑ Medium: First questions hint at stance.
- ❑ Low: First question asks for stance labeling; colors correspond to label chosen.

Perceptibility of Stance Imbalance

Experiment: Design

Subtasks 1/2 (high difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Which text is more biased?

(2) How sure are you?

(3) Explain your decision.

Subtask 3 (medium difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Which text has more pro stances?

(2) Which text has more con stances?

(3) Which text is more biased?

(4) How sure are you?

(5) Explain your decision

Subtask 4 (low difficulty)

Text *X*	Text *Y*
Argument 1	Argument 1
Argument 2	Argument 2
Argument 3	Argument 3
Argument 4	Argument 4

(1) Label the stances of each argument.

(2) Which text is more biased?

(3) How sure are you?

(4) Explain your decision

Task 1 ⟶

⟵ **Task 2**

Task difficulty: (All: Explanation of bias as one-sidedness.)

- ❑ High: No mention of stance.
- ❑ Medium: First questions hint at stance.
- ❑ Low: First question asks for stance labeling; colors correspond to label chosen.

- ❑ Training: high-to-low (Task 1); low-to-high (Task 2)

Differential Bias

ID	Variable	Range
D	perceptibility of differential stance bias	
I_1	differential stance bias	3 levels
I_2	difficulty	low, medium, high
I_3	participant training	autodidact vs. trained
I_4	participant expertise	stance labeling accuracy
I_5	participant confidence	3-level self-assessment
C_1	argumentation length	4 arguments each
C_2	argument length	2-3 sentences each
C_3	argument frames	all arguments from 1 frame
C_4	order effects I	random order of arguments
C_5	order effects II	no textual coherence
C_6	opinion diversity	9 participants per topic
U_1	stance perceptibility	e.g., explicit vs. implicit
U_2	other language biases	e.g., at the lexical level

Differential Bias
Experiment: Setup

Argument Corpus:

❑ Webis Argument Framing Corpus 2019 [Ajjour 2019]
  12,000 arguments, 465 topics, stance and frame labels, source: debaptepedia.org

❑ Sampling: per subtask 8 arguments of the same topic, frame, and length

❑ Review: dictionary-based and manual check for language biases

Differential Bias
Experiment: Setup

Argument Corpus:

❑ Webis Argument Framing Corpus 2019 [Ajjour 2019]
   12,000 arguments, 465 topics, stance and frame labels, source: debaptepedia.org

❑ Sampling: per subtask 8 arguments of the same topic, frame, and length

❑ Review: dictionary-based and manual check for language biases

Crowdsourcing Study:

❑ Subjects: Workers recruited on Amazon's Mechanical Turk
   20 HITS, 10 per task, 571 workers, 9 per HIT, one-time participation, no revision of subtasks

❑ Composition: HIT with $(\Delta(X,Y)=0)$ and without differential bias $(\Delta(X,Y)>0)$

❑ Quality control:
   – Work time: exclusion when reading times <3m instructions and <8m subtask
   – MACE score [Hovy 2013]: exclusion when worker score <80%
   – Approval rate: exclusion of rates <90%
   – Language proficiency: exclusion when explanations are ungrammatical / nonsensical
   – Admission test (Task 2): admission on successful completion of a toy example
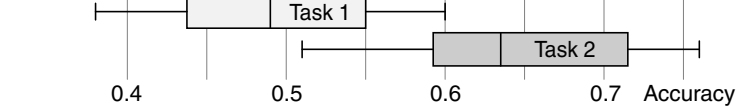
Differential Bias

Experiment: Results

Differential Bias		Accuracy per Subtask and Difficulty							
$\Delta(X,Y)$	$\delta(X){:}\delta(Y)$	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4
0	0:0	0.27	0.33	0.58	0.66	0.60	0.55	0.42	0.50
0	2:2	0.33	0.28	0.56	0.57	0.62	0.67	0.56	0.54
0	4:4	0.11	0.36	0.44	0.64	0.70	0.64	0.52	0.50
2	2:0 / 0:2	0.41	0.60	0.55	0.68	0.88	0.86	0.62	0.70
2	4:2 / 2:4	0.44	0.44	0.55	0.65	0.77	0.69	0.66	0.59
4	4:0 / 0:4	0.45	0.55	0.68	0.70	0.88	0.66	0.75	0.66

(b)

(c)

Average Accuracy per Task and Subtask Difficulty

Sub.	Diffic.	Acc.	Δ Acc.	Sub.	Diffic.	Acc.	Δ Acc.
1.1 / 1.2	high	0.38	0.00	2.1	low	0.74	0.00
1.3	med.	0.55	0.17	2.2	med.	0.67	−0.07
1.4	low	0.64	0.09	2.3 / 2.4	high	0.58	−0.09

(d)

Differential Bias		Accuracy per Worker Competence Level															
		Task 1								Task 2							
		Low				High				Low				High			
$\Delta(X,Y)$	$\delta(X){:}\delta(Y)$	1.1	1.2	1.3	1.4	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.1	2.2	2.3	2.4
0	0:0	0.11	0.00	0.18	0.28	0.37	0.55	0.66	0.53	0.55	0.44	0.33	0.24	0.61	0.63	0.55	0.66
0	2:2	0.14	0.33	0.42	0.11	0.44	0.22	0.52	0.55	0.43	0.25	0.22	0.12	0.70	0.65	0.50	0.55
0	4:4	0.00	0.26	0.33	0.24	0.68	0.44	0.66	0.48	0.50	0.33	0.48	0.44	0.62	0.55	0.66	0.58
2	2:0 / 0:2	0.20	0.44	0.55	0.35	0.72	0.66	0.63	0.68	0.66	0.60	0.50	0.54	0.80	0.75	0.58	0.72
2	4:2 / 2:4	0.48	0.45	0.42	0.28	0.55	0.77	0.60	0.66	0.52	0.57	0.55	0.33	1.0	0.88	0.70	0.66
4	4:0 / 0:4	0.55	0.36	0.48	0.60	0.66	0.62	0.74	0.76	0.61	0.46	0.56	0.20	0.75	0.68	0.68	0.64
Average		0.25	0.31	0.40	0.31	0.57	0.54	0.63	0.61	0.54	0.44	0.40	0.29	0.74	0.69	0.61	0.63

Differential Bias
Experiment: Results

Overall Insight:

- ❑ overall, subjects ($a\bar{c}c = 0.32$) are slightly better than random ($a\bar{c}c = 0.27$)
- ❑ stance labeling itself is a hard task: only one in for subjects has high expertise level
- ❑ subject demographics (self-reported) have an influence

  - – education and language skill are significantly associated with expertise ($p = 0.017$, $p = 0.002$)
  - – But: not associated with perceptibility ($p = 0.683$, $p = 0.878$)

Differential Bias

Experiment: Results

I_1 Differential Bias Level

- bias differential of arguments displayed to the subject

- 4 premises in each argument, three possible levels of bias

Bias Level Differential	Bias Perception Accuracy
0	0.21
2	0.42
4	0.44

Differential bias level and perceptibility are significantly associated ($p = 0.000$)

Subjects perceive bias better in cases of high bias differential.

Differential Bias

I_2 **Training Level**

   – training operationalized by
      subtask progression

   – Trained: increasing difficulty;
      Untrained: decreasing difficulty

Training Task Progression	Bias Perception Accuracy
Untrained	0.32
Trained	0.33

Training did not significantly improve
subjects' perceptibility ($p = 0.842$)

Task progression was not an effective training tool
to teach subjects to perceive bias by stance differential.

Differential Bias

Experiment: Results

I_3 Subtask Difficulty

- each subtask gives decreasing amounts of aid to the subject

- e.x. visual highlighting, multi-stage decisions

Subtask Difficulty Decision Support	Bias Perception Accuracy
Low	0.36
Medium	0.31
High	0.31

Subtask difficulty and perceptibility
are not significantly associated ($p = 0.325$)

Decreasing task difficulty with decision support seems
to increase perceptibility, but not significantly.

Differential Bias

Experiment: Results

I_4 Worker Expertise Level

– operationalized with stance labeling accuracy (Subtask 4)

– < 0.7 low, ≥ 0.7 high

Expertise Level Stance Labeling	Bias Perception Accuracy
Low	0.30
High	0.38

Stance labeling expertise and perceptibility are significantly associated ($p = 0.043$)

Subjects' with higher expertise in stance labeling perceive bias better.

Differential Bias

Experiment: Results

I_5 Worker Confidence

- self reported on 3-point scale

- aggregated over all subtasks

Worker Confidence Self-judged	Bias Perception Accuracy
Low	0.32
Medium	0.33
High	0.31

Subjects' confidence is not significantly associated with perceptibility ($p = 0.089$).

Strong evidence that subjects cannot self-judge their ability to spot bias.

Conclusion

Take-away messages:

❑ Bias different can be recognized, even without much outside help.

❑ The higher the difficulty, and the less the prior training, the more difficult it becomes to recognize biases

❑ Workers were able to justify their judgments well, albeit wrong choices were made not due to stance differences.

Conclusion

Take-away messages:

- ❏ Bias different can be recognized, even without much outside help.

- ❏ The higher the difficulty, and the less the prior training, the more difficult it becomes to recognize biases

- ❏ Workers were able to justify their judgments well, albeit wrong choices were made not due to stance differences.

Future work:

- ❏ Application of differential bias analysis to other kinds of biases

- ❏ Application of statistical models to derive bias scores from pairwise preferences

- ❏ Development of UIs that support the identification of (stance) bias

- ❏ Tools to support formal debate training curricula

Conclusion

Take-away messages:

- ❑ Bias different can be recognized, even without much outside help.

- ❑ The higher the difficulty, and the less the prior training, the more difficult it becomes to recognize biases

- ❑ Workers were able to justify their judgments well, albeit wrong choices were made not due to stance differences.

Future work:

- ❑ Application of differential bias analysis to other kinds of biases

- ❑ Application of statistical models to derive bias scores from pairwise preferences

- ❑ Development of UIs that support the identification of (stance) bias

- ❑ Tools to support formal debate training curricula

Thank You!