

Detecting Biased Arguments: A Site for Argumentative Injustice

Asma Tajuddin

tajuddi@uwindsor.ca

University of Windsor

Research Article

Keywords: Argumentative injustice, biased argumentation, context sensitivity, epistemic resources

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11462541/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Biased argumentation is described as one-sided, containing selective reasoning, myside bias, etc. This paper addresses how detecting biased argumentation can become a site of argumentative injustice. Bondy's (2010) account of argumentative injustice can help situate how different accounts of biased argumentation address context sensitivity. I argue that even though these accounts help identify important patterns in reasoning, they fail to adequately address the social identity of the arguer and the alternative argumentative practices that do not conform to dominant logocentric forms. When non-conforming arguments such as narratives, testimonials, or emotional arguments are discounted, bias detection can reproduce argumentative injustice. The paper identifies this point of exclusion and demonstrates how biased argument detection needs to account for context sensitivity.

1. Introduction

Hundleby's foundational critique of dominant argumentation structures (2013) helps situate bias at the level of argument. According to Walton (2008), an argument is biased if it pushes only for one side. In his book on the *Pragmatic Theory of Fallacy* (1995), Walton explains that an argument is biased when it unfairly dismisses or ignores opposing viewpoints. Biased argumentation accounts for one-sided arguments, failures of critical doubt, motivated reasoning, myside bias and social bias (Correia, 2011, 2014; Kienpointner & Kindt, 1997; Walton, 1991, 1999; Wolfe & Britt, 2008; Spliethöver & Wachsmuth, 2020).

Walton (1991, 1999) defines biased arguments as one-sided, failing to address appropriate critical doubt. Kienpointner and Kindt (1997) conceptualize bias in political argumentation as a lack of balance in reporting relevant dimensions of an issue or lacking relevant argument structures. Correia (2011) links biased argumentation to motivational distortion using selective evidence, selective premises, biased interpretation, hasty generalization, one-sidedness, and unintended straw-man reasoning. Wolfe and Britt (2008) define myside bias as the failure to consider information supporting the opposing side. Wolfe et al. (2009) distinguished between concessions, rebuttals and denial of counter positions. Tomasetto et al. (2009) show that majority and minority social positions can shape argumentative strategies, affecting how arguers produce arguments and respond to opposing positions. For example, testimonial arguments may be disregarded as they may not conform to dominant logocentric models (Al Tamimi, 2017). Pimenova (2022, 2025) demonstrates how institutions treat opposing arguments. Her work on Argument Continuity shows how dominant arguers may repeatedly reproduce the same counterarguments to dismiss opposing arguments without engaging. These studies raise questions about the standards by which biased arguments are detected. It situates bias detection as a site of argumentative injustice because these accounts may not recognize all forms of reasoning equally.

Computational approaches for biased argumentation analyses include frameworks like ABBA for measuring social bias, word embeddings, and attribute-value pairs for identifying underlying algorithmic decisions (Holtermann et al., 2022; Spliethöver & Wachsmuth, 2020; Waller et al., 2024). Mueller et al.

(2026) identify the tension between charitable reconstruction and accurately representing biased reasoning, raising questions about how arguments are represented for subsequent evaluation. Self (2026, in press) identifies reconstruction as a potential site of argumentative injustice because the argumentative force of a narrative argument, for example, is lost when it is reconstructed into a premise-conclusion form.

Bondy (2010) argues that argumentative injustice occurs when an argument is favoured due to the arguer's social position. An argument is rendered structurally inadequate if it has inherent limitations due to non-adherence to rigid, formalized logical structures. Narrative, testimonial, and emotional arguments are rendered structurally inadequate as they fail to conform to linear, explicit and discursive reasoning (Al Tamimi, 2017). I examine how the detection of biased argumentation is a site for argumentative injustice. This paper asks whether automated bias detection can reproduce argumentative injustice when non-conforming arguments, such as testimony, narrative, and emotional reasoning, are discounted and therefore labelled biased.

The empirical analysis investigates contextual sensitivity as one point of entry into this problem. It analyzes 586 contributions from the Canadian House of Commons Hansard recorded between June 16 and June 18, 2026. The purpose is to examine whether the biased arguments detected differ due to different levels of contextual information provided. Each argument is analyzed three times: first as an argument alone, second with the surrounding exchange, and third with the underlying dialogue type. The study examines the role of contextual sensitivity in the detection of biased argumentation. The results help examine whether biased argumentation detection is a site for argumentative injustice by gathering context such as type of dialogues, argumentation form, schemes, evidence, stance and epistemic resources. Epistemic resources refer to the sources, forms, and modes of knowing that speakers invoke when presenting, supporting, or explaining a claim. The empirical analysis therefore investigates contextual sensitivity as one point of entry into the broader philosophical problem.

2. Background

Biased argumentation accounts identify one-sidedness, failures of critical doubt, selective evidence, motivated reasoning, myside bias, credibility, social identity, social position, and evaluative norms.

2.1 Dialogue, One-Sidedness, and Critical Doubt

Walton's work connects bias to failures of critical doubt and inappropriate treatment of alternatives, while emphasizing that argumentative evaluation depends on dialogue type, commitments, critical questions, and argumentation schemes (Walton, 1991, 1999a, 1999b, 2006a, 2006b, 2006c, 2010, 2012). According to Walton (2008), an argument is biased if it pushes only for one side, disregarding the opposite view. He explains how an argument can be considered biased when it unfairly dismisses or ignores opposing viewpoints. It is also important to differentiate between biased arguments and fallacies. A fallacy in informal logic is an error in reasoning that appears persuasive but is logically

flawed due to contextual or rhetorical weaknesses (Walton, 2008; Govier, 2005; Tindale, 2007). Fallacies may arise from ambiguity in arguments (Walton et al., 1996). One-sidedness is not inherently defective, though it becomes problematic when counterarguments or positions are excluded (Walton, 1991, 1999b). Kienpointner and Kindt (1997) distinguish local and global bias, showing how mishandling of relevant information to present a cause can lead to biased arguments. This relevant information can include alternative viewpoints, counterarguments, and counterevidence.

Dziuda (2011) and Zenker (2011) further show that selective disclosure of facts, cherry-picking proofs, and presenting interest-based objections may benefit certain dominant argumentation structures. For example, Pragma-dialectical approaches likewise assess argumentative moves in relation to the obligations of critical discussion, reinforcing norms and rules that can exclude nonconforming engagement (van Eemeren & Grootendorst, 1992, 2004). Van Laar (2006, 2007) has categorized bias in arguments as stance bias, involving the exclusion or underrepresentation of opposing positions, and frame bias, involving the omission of relevant information. Van Laar and Krabbe (2013) note that a dialogue involves responsiveness to criticism, examining how addressing counterarguments helps strengthen one's position.

2.2 Motivated and Cognitive Accounts

Festinger (1957) and Kunda (1990) provide foundational accounts of motivated reasoning, while Thagard (2011) examines how emotional processes influence reasoning. Correia (2011, 2012, 2018) connects motivated irrationality to biased argumentation by examining how arguers interpret and select evidence. He links motivated reasoning with failures of critical thinking by selectively processing information to support preferred positions. Maillat and Oswald (2013) examine cognitive constraints in persuasion dialogue, emphasizing how contextual information helps interpretation. Kobayashi (2016) addresses relational processing of conflicting arguments for bias assimilation, and Konverskyi and Kolotilova (2023) examine confirmation bias in argumentation and how actively questioning one's own reasoning against prior commitments can help mitigate it. Jayez (2026) adds a linguistic account of implicit bias in arguments: discourse relations and argumentative interpretation depend on contextual and non-local features of discourse.

Mercier and Sperber (2011, 2017) and Mercier (2016) argue that motivated reasoning can produce asymmetries between supporting and counterarguments. Reasoning may function differently in self-justification and in the evaluation of opposing positions. Mercier and Heintz (2014) extend this perspective to scientific reasoning for evaluating claims and hypotheses. Banisch and Shamon (2025) further examine how motivated reasoning leads to polarization by reinforcing attitude-consistent evaluations and discounting opposing views.

2.3 Myside Bias

Wolfe and Britt (2008) explain one-sidedness as arguers selectively incorporating information that supports their position while excluding opposing information. According to Walton (2008), an argument is biased if it pushes only for one side, disregarding the opposite view. One-sided reasoning is further

categorized by Van Laar (2014) as stance bias: exclusion of opposing positions, and frame bias: exclusion of a relevant aspect. Biased arguments often use loaded language, emotive terms, or one-sided reasoning to persuade unfairly. Wolfe, Britt, and Butler (2009) relate my-side bias to argumentative structures that are designed to exclude counterpositions. Wolfe (2012) examines how my side bias varies across arguers. Mitigating my-side bias will require considering opposing claims and counterevidence, including strategies for evaluating counterarguments (Christensen-Branum et al., 2019). There is a social dimension to how some arguers receive uptake and credibility, and how minority arguments may be discounted; some positions receive more uptake and credibility. Thompson (2022) argues that persistent social inequalities can prevent marginalized speakers from achieving uptake, connecting this to epistemic injustice (Fricker, 2007).

2.4 Social, Epistemic, and Normative Accounts

Bondy (2010) shows how an arguer can face argumentative injustice when others discredit their arguments or fail to receive uptake. Hundleby (2013, 2023) provides a foundational critique of dominant argumentation structures. For example, dominant politeness norms institutionalize aggression by allowing socially privileged arguers to deploy aggressive tactics under the guise of "polite intellectual sparring" while penalizing marginalized individuals for the same behaviour. This model assumes an adversary with equal social standing. Such norms can obscure differences in social standing and the resources available to arguers.

Hundleby (2023) argues that rendering minority arguments as fallacious can discount their persuasive power and also discredit the arguer. This type of labelling turns argument evaluation into a binary "pin-the-fallacy-on-the-argument" game. Howes and Hundleby (2018) extend this concern to institutionalized bias in science and argumentation. Cohen and Stevens (2016) question whether bias can be separated from argument evaluation practices, while Mizrahi (2018) examines persistent bias in judgments involving expert opinion.

Dominant argumentation structures enforce logocentrism, systematically assign credibility deficits to marginalized speakers and credibility excesses to institutional authorities. Hundleby advocates for implementing meta-level evaluations that explicitly account for social location, power relations, and structural bias. Hinton (2024) further demonstrates the relevance of identity and representational discourse to the normative evaluation of arguments.

2.5 Computational Detection and Reconstruction

Computational approaches translate theoretical accounts of bias into operational criteria for identifying and analyzing biased argumentation. Spliethöver and Wachsmuth (2020) investigate social bias in arguments, while Holtermann et al. (2022) contribute an annotated corpus for bias identification in argumentative text, examining computational representations of social bias. Waller et al. (2024) use argumentation-based analysis to identify reasons underlying bias. Together, these works illustrate different ways of computational detection. However, automated detection also depends on how arguments are represented and reconstructed. Mueller et al. (2026), for example, identify problems

arising when the principle of charity makes a reconstructed argument more plausible by obscuring its underlying reasoning.

3. Methodology

Detecting biased arguments might require operationalizing a conception of bias through its representation of the argument, context, categories, and evaluative criteria. The central question for the present study is: Does the bias classification assigned to a focal contribution change when the contextual information available to the classifier changes? This question shifts attention from detecting bias in arguments to examining whether the procedure used to detect it can itself become a site of argumentative injustice. This empirical question is relevant to the broader problem of argumentative injustice because bias evaluation depends on which features of their reasoning are recognized. The experiment examines contextual sensitivity, which it builds using empirical dimensions of the Hansard debates. The experiment examines contextual sensitivity by comparing classifications across three contextual conditions. The empirical dimensions of the Hansard contributions, including argument form, argument scheme, evidence type, epistemic resource, stance, dialogue type, and argumentative target, provide additional descriptive variables for examining classification patterns.

3.1 Context Sensitivity

Argument is a social activity; it requires ethical norms (Aikin & Vollbrecht, 2020). Doing well at argumentation motivates people to produce stronger reasons, which can help boost adversarial motivation (Zarefsky, 2012). To engage in a debate, one presents the strongest cases, counterarguments, and criticisms that help understand an issue (Aikin, 2011, p. 260). Aikin suggests that minimal dialectical adversariality, such as confrontation of opposing views, is an essential component of epistemic betterment, allowing a rigorous sorting function for reasons. The distinction between virtuous and vicious (ancillary) adversariality plays a critical role in debiasing arguments by moving the focus from simple filtering (detecting loaded language) to variables such as power dynamics, social relations of the arguer and the audience, etc. Adversariality is good for arguers as it supports differences rather than suppressing them (Govier & Hoaglund, 1999). It can help address an arguer's intention as a "fair goal" (Gilbert, 1995). Adversarial argument norms may enhance agency, identity, and the relationship between arguers and audience (Henning, 2018, 2021).

Hundleby (2013) provides a foundational critique of traditional argumentation theory, focusing on the structural biases embedded in adversarial reasoning models. It largely situates bias at the level of argument. The problem of adversariality is aggressive modes of discourse excluding women and socially marginalized groups, such as oppositional framing leading to Polarization (Moulton, 2015; Tannen, 1998). Hundleby warns that assuming all arguers have equal "exit options" or intellectual authority hides systematic argumentative injustices (Hundleby, 2013, p. 255).

Consideration of alternative views and opposing arguments is crucial for reaching reasoned judgements (Finocchiaro, 1994; Perkins, 1989; Perkins et al., 1983). Verification of evidence and arguments helps reach a reasoned judgement by making dialectical inquiry (Kuhn, 1991; Bailin & Battersby, 2021). Innocenti identified the sources of force in arguing: 1) intellectual force exerted by an arguer adhering to logical and epistemic norms, and 2) conventional force, such as an arguer adhering to the rules for critical discussions, like intersubjective agreements and problem-solving efficacy.

Howes and Hundleby (2018) argue that emotions, including anger, can provide significant epistemic value by providing information on injustice. This framework is utilized by Howell (2023) to justify protest as a reasoned argument. Arguers can address aggression to “allow rationality to play its intended role.” Civility can be a weaponized norm that silences marginalized speakers (Hundleby, 2013, p. 259).

Hundleby developed the status quo fallacy to address implicit bias in arguments. The status quo fallacy is identified as a critical reasoning error in which dominant social norms are assumed to be the standard. On this view, the goal of an argument shifts from “defeating an opponent” to collaboratively improving collective reasoning (Govier & Hoaglund, 1999; Moulton, 2015).

Hundleby (2021) advocated for an “outside-in” methodological approach, focusing on perspectives which are outside the dominant domains. Hundleby and Howes (2021) emphasize that contextual vulnerability can lead to new opportunities in argumentation. These opportunities can help develop human-centric models which see the arguer and the audience with both agency and vulnerability, rather than as simple actors and passive objects. Addressing rhetorical disadvantage and epistemic injustice can help develop AI models that can mitigate subtle systemic biases (Haro Marchal, 2025; Howes & Yap, 2026). Govier (2021) and Aberdein (2016) hint at discursive markers of subordinate status in data, which can perpetuate racial and gender bias. Hundleby identifies social bias as the case in which an arguer unreflectively assumes that social norms are natural or correct (Hundleby, 2013; Hundleby et al., 2016). A core component of debiasing is accounting for the social location and recognizing Hundleby’s warning that “who is arguing matters.” I am exploring this contextual vulnerability to find variables to address context sensitivity and answer how bias detection is a site of argumentative injustice.

3.2 Dataset

The dataset contained 586 Canadian House of Commons Hansard contributions from June 16–18, 2026. It contains 105,723 words and 158 unique arguers. Contributions range from 6 to 3,228 words, with a mean of 177 words and a median of 100 words. The unit of analysis is the complete non-empty Hansard contribution represented by one CSV row (LiPaD, 2026). Contributions were not divided into sentences or individual claims. Original Hansard metadata, including publication, speaker, constituency, caucus, province, date, time, page, and row identifier, were retained. Parliamentary debate was selected because contributions occur within recognizable exchanges and institutional dialogue practices, allowing the same argumentative contribution to be evaluated both in isolation and in context.

3.3 Research Design

This study examines whether detection of biased arguments is sensitive to contextual information. The design holds the focal contribution constant while progressively varying the information available to the classifier. Each contribution is evaluated independently under three conditions: (1) argument only, (2) argument plus exchange, and (3) argument plus exchange plus dialogue context.

In the argument-only condition, the classifier receives the complete focal contribution without surrounding material. In the argument-plus-exchange condition, it receives the focal contribution together with up to two preceding and two following contributions. In the argument-plus-dialogue-context condition, the same exchange is supplemented by an explicit characterization of the dialogue type. The experiment therefore tests contextual sensitivity rather than classification accuracy. A classification change indicates that the evaluation of the same contribution changed when contextual information was introduced; it does not establish that either classification was correct. The focal contribution, model, classification categories, and evaluation task remain constant across conditions. No information from one condition is supplied to another. A Python-based classification pipeline using APIs processed each non-empty Hansard CSV row as one complete contribution. A structured annotation classifier assigned categories for argument form, argument scheme, evidence type, epistemic resource, stance, dialogue type, and target. A bias classifier then returned a classification (biased, not_biased, or ambiguous), bias type, and rationale under three contextual conditions: argument-only, argument-plus-exchange, and argument-plus-dialogue-context.

3.4 Empirical Dimensions

There are seven empirical dimensions: 1) argument form, 2) argument scheme, 3) evidence type, 4) epistemic resource, 5) stance, 6) dialogue type, and 7) target. These dimensions capture relevant epistemic, argumentative, and contextual features of the contributions. They serve as descriptive and analytical variables rather than ground-truth indicators of bias.

Argument forms can be premise-conclusion, testimonial, narrative, experiential, or mixed. Argument schemes include argument from authority, expert opinion, negative/positive consequence, causal, analogy, example, practical reasoning, or other. Evidence types can be statistical, expert, testimonial, experiential, institutional, historical, anecdotal, or none. Epistemic resources included data, expertise, testimony, personal experience, constituent experience, narrative, institutional authority, emotion, or common knowledge. Stance was coded as support, oppose, or neutral. Dialogue type comprised Persuasion, Negotiation, Inquiry, Deliberation, Information-seeking, Eristic, Discovery, or Mixed. Target was coded as policy, institution, government, opposing position, individual, or other. Evidence type and epistemic resource were distinguished because they capture different aspects of a contribution: the former identifies what is presented as evidence, while the latter identifies the source or mode of knowing invoked.

Bias was classified separately as biased, not biased, or ambiguous. The classifier was designed with buffers to stop it from inferring bias from political content, emotional language, controversy, partisanship, criticism, persuasion, testimony, experiential reasoning, or speaker identity alone. Bias-type

features include prejudicial framing, selective treatment, stereotyping, unsupported generalization, asymmetrical reasoning, and exclusionary reasoning. These categories were descriptive and could co-occur within a contribution. Dialogue types are defined, but with the exception that there might be a mixed category when no single dialogue type adequately characterizes the exchange.

4. Results

The results compare bias classifications across three contextual conditions: C1 (argument only), C2 (argument plus surrounding exchange), and C3 (argument plus surrounding exchange and dialogue-type label). Classifications are coded as biased (B), not biased (Nb), or ambiguous (A). Table 1 shows a shift in bias classifications as contextual information was introduced. It shows that adding contextual information reduced ambiguous classifications from 43.5% in C1 to 25.8% in C2. Not biased classification increased substantially, while biased classifications rose gradually from 20.8% in C1 to 25.6% in C3. These findings indicate that the classifier's evaluations varied with the contextual information provided,

Table 1
Bias classification across contextual conditions

Bias classification	Argument only, n (%)	Argument + exchange, n (%)	Argument + dialogue context, n (%)
Biased	122 (20.8%)	131 (22.4%)	150 (25.6%)
Not biased	209 (35.7%)	304 (51.9%)	289 (49.3%)
Ambiguous	255 (43.5%)	151 (25.8%)	147 (25.1%)
Valid cases	586 (100%)	586 (100%)	586 (100%)

Table 2 shows changes in bias-type labels across the three conditions. The results suggest that contextual information influenced the specific bias types assigned by the classifier. Table 3 shows that changes in bias classification varied across dialogue types. Table 4 shows that changes in bias classification varied by stance, with biased classifications increasing gradually from C1 to C3.

Table 2
Most common bias types across three conditions

Bias type	C1, n (%)	C2, n (%)	C3, n (%)
None	163 (27.3%)	152 (25.9%)	147 (25.1%)
Insufficient information	70 (11.9%)	15 (2.6%)	13 (2.2%)
Prejudicial framing	46 (7.8%)	32 (5.5%)	43 (7.3%)
Unknown	34 (5.7%)	12 (2.0%)	11 (1.8%)
Selective treatment of relevant considerations	27 (4.6%)	24 (4.1%)	17 (2.9%)

Table 3
Bias classification by dialogue type across three conditions

Dialogue type	<i>n</i>	C1: B	C1: Nb	C1: A	C2: B	C2: Nb	C2: A	C3: B	C3: Nb	C3: A
Eristic	178	91	33	54	94	46	38	113	29	36
Deliberation	121	9	60	52	7	81	33	8	80	33
Persuasion	105	17	45	43	26	53	26	26	54	25
Information-seeking	82	1	31	50	1	55	26	0	57	25
Mixed	42	2	15	25	3	27	12	3	27	12
Discovery	23	1	19	3	0	19	4	0	18	5
Inquiry	18	1	4	13	0	12	6	0	12	6
Negotiation	17	0	2	15	0	11	6	0	12	5
Total	586	122	209	255	131	304	151	150	289	147

Table 4
Bias Classification by Stance across three Conditions

Stance	<i>n</i>	C1: B	C1: Nb	C1: A	C2: B	C2: Nb	C2: A	C3: B	C3: Nb	C3: A
Oppose	239	111	51	77	111	75	53	128	62	49
Neutral	184	1	64	119	2	124	58	1	122	61
Support	163	10	94	59	18	105	40	21	105	37
Total	586	122	209	255	131	304	151	150	289	147

In Table 5, Premise–conclusion was the most frequent argument form (47.8%), followed by testimonial (22.0%) and narrative forms (20.0%). Mixed forms accounted for 8.7% of contributions, while experiential forms were least common (1.5%). Although premise–conclusion remained the dominant category, testimonial and narrative forms together accounted for 42.0% of the dataset, demonstrating the presence of diverse argumentative forms in parliamentary debates. Table 6 illustrates the distribution of argumentation schemes in the contributions. Practical reasoning was the most frequent argument scheme (27.0%), closely followed by Argument from positive/negative consequence (25.9%). Table 7 shows that more than half of the contributions were classified as containing no identifiable evidence type (54.6%). Testimonial evidence was the most frequent category among those identified (14.7%), followed by institutional evidence (9.2%), statistical evidence (8.5%), and anecdotal evidence (7.0%). Experiential, historical, and expert evidence were comparatively rare, each accounting for approximately 2% of contributions.

Table 5
Distribution of Argument form

Argument form	Count	Percentage
Premise-conclusion	280	47.8%
Testimonial	129	22.0%
Narrative	117	20.0%
Mixed	51	8.7%
Experiential	9	1.5%
Total	586	100.0%

Table 6
Distribution by Argumentation schemes

Argument scheme	Count	Percentage
Practical reasoning	158	27.0%
Consequence	152	25.9%
Other	139	23.7%
Example	57	9.7%
Authority	48	8.2%
Causal	22	3.8%
Analogy	10	1.7%
Total	586	100.0%

Table 7
Distribution of Evidence type

Evidence type	Count	Percentage
None	320	54.6%
Testimonial	86	14.7%
Institutional	54	9.2%
Statistical	50	8.5%
Anecdotal	41	7.0%
Experiential	12	2.0%
Historical	12	2.0%
Expert	11	1.9%
Total	586	100.0%

Table 8
Distribution of epistemic resources

Epistemic resource	Count	Percentage
Institutional authority	172	29.4%
Common knowledge	141	24.1%
Testimony	110	18.8%
Narrative	51	8.7%
Data	42	7.2%
Constituent experience	30	5.1%
Emotion	15	2.6%
Personal experience	13	2.2%
Expertise	12	2.0%
Total	586	100.0%

Table 9
Distribution of argumentative targets

Target	Count	Percentage
Government	172	29.4%
Policy	152	25.9%
Opposing position	115	19.6%
Individual	59	10.1%
Other	56	9.6%
Institution	32	5.5%
Total	586	100.0%

Table 8 shows which epistemic resources were identified by the classifier. Institutional authority was the most frequently identified resource (29.4%), followed by common knowledge (24.1%) and testimony (18.8%). Together, these three categories account for approximately 72% of the contributions, suggesting that the classifier most often identified appeals to institutional standing, shared understandings, and accounts offered by others as resources for supporting claims. Narrative (8.7%) and data (7.2%) appeared less frequently, followed by constituent experience (5.1%). Emotion, personal experience, and expertise were each identified in approximately 2–3% of contributions. Table 9 shows the distribution of argumentative targets. Government was the most frequently identified target (29.4%), followed by policy (25.9%) and opposing positions (19.6%).

Table 10
Bias Classification by Epistemic Resources across all conditions

Epistemic resource	<i>n</i>	C1: B (%)	C1: Nb (%)	C1: A (%)	C2: B (%)	C2: Nb (%)	C2: A (%)	C3: B (%)	C3: Nb (%)	C3: A (%)
Institutional authority	172	9.3	31.4	59.3	15.7	51.7	32.6	17.4	54.1	28.5
Common knowledge	141	34.8	26.2	39.0	32.6	44.7	22.7	41.1	39.7	19.1
Testimony	110	11.8	39.1	49.1	11.8	60.0	28.2	10.9	58.2	30.9
Narrative	51	19.6	52.9	27.5	15.7	64.7	19.6	19.6	51.0	29.4
Data	42	45.2	28.6	26.2	50.0	26.2	23.8	52.4	26.2	21.4
Constituent experience	30	13.3	63.3	23.3	10.0	73.3	16.7	16.7	63.3	20.0
Emotion	15	46.7	26.7	26.7	60.0	26.7	13.3	53.3	33.3	13.3
Personal experience	13	15.4	53.8	30.8	15.4	69.2	15.4	15.4	69.2	15.4
Expertise	12	16.7	50.0	33.3	16.7	58.3	25.0	25.0	50.0	25.0

Table 10 shows that changes in bias classification varied across epistemic resources. Institutional authority and testimony shifted from ambiguous classifications toward not biased when contextual information was introduced. By contrast, data and emotion were associated with comparatively high proportions of biased classifications across all three conditions, with both reaching their highest levels in C3 for data and in C2 for emotion. Overall, these findings suggest that the classifier's sensitivity to context differed according to the epistemic resource identified.

5. Conclusion

This study examined whether the detection of biased argumentation changes when more contextual information is made available. The results show that classifications varied across the three contextual conditions. Table 1 shows that ambiguous classifications decreased substantially when the surrounding exchange was introduced, while not biased classifications increased. The biased classification increased more gradually across the three conditions. Adding the explicit dialogue-type characterization in C3 was associated with a further increase in biased classifications, while not biased and ambiguous classifications remained relatively stable compared with C2. Table 4 similarly shows variation across stance categories. Contributions coded as opposing a position accounted for the largest number of biased classifications in every condition, with the count increasing from 111 in C1 and C2 to 128 in C3. Neutral contributions shifted substantially from ambiguous to not biased when more contextual information was introduced.

The distribution of argumentative forms also matters for understanding the scope of the analysis. Table 5 shows that premise–conclusion forms were the most frequently identified category, followed by testimonial and narrative forms. Testimonial and narrative forms together accounted for 42% of the contributions, indicating that the corpus contains substantial variation. This distribution is relevant to the paper's concern with dominant logocentrism, which can be further investigated to find which argument forms receive uptake.

Table 8 shows that institutional authority, common knowledge, and testimony were the most frequently identified epistemic resources. Table 10 indicates that classification patterns differed across these resources. Institutional authority shifted from predominantly ambiguous classifications in C1 toward not-biased classifications in C2 and C3. Testimony also showed a substantial increase in not-biased classifications when the surrounding exchange was introduced. These patterns deserve attention because the classification distributions associated with different epistemic resources were not alike. However, these findings should therefore be interpreted as descriptive patterns. The classifier was instructed not to infer bias from emotional language, controversy, or stance alone. Table 10 and Table 4 show that this instruction did not fully determine the outcomes. Emotion was classified as biased in 46.7% of contributions under C1, 60.0% under C2, and 53.3% under C3. Opposing stance was classified as biased in 46.4%, 46.4%, and 53.6% of contributions, against 6.1% to 12.9% for supporting contributions and 0.5% to 1.1% for neutral contributions. These patterns could indicate that emotion and opposition co-occur with features the classifier treats as one-sidedness.

More than half of the contributions were classified as containing no identifiable evidence type, while government, policy, and opposing positions were the most frequently identified argumentative targets. The concern is that accounts of biased argumentation might inherently privilege dominant argumentation structures, so context sensitivity matters. It is also the evaluative criteria that determine which features of argumentative contributions become visible to assessment and how those features are interpreted. Testimonial contributions were classified as not biased in 39.1% of cases under C1 and 60.0% under C2, while the biased share stayed between 10.9% and 11.8% across all three conditions (Table 10). Under argument-only evaluation, testimony was ambiguous in 49.1% of cases. Adding the exchange reduced ambiguity to 28.2%. Argument-only evaluation is therefore where testimony is most exposed to discounting, but in these data the exposure takes the form of unresolved ambiguity rather than biased labelling. Institutional authority was classified as not biased in 31.4%, 51.7%, and 54.1% of contributions across the three conditions, while its ambiguous share fell from 59.3% under C1 to 28.5% under C3 (Table 10). If institutional standing reduces the evaluator's uncertainty, the pattern is consistent with a credibility excess in Fricker's (2007) sense.

This concern connects the findings to Bondy's (2010) account of argumentative injustice and feminist critiques of dominant argumentation norms, particularly Hundleby's work on adversariality, exclusion, and the social conditions of argumentative evaluation. This raises the possibility that standards presented as neutral or universal may not fully accommodate differences in social position, communicative practice, and the epistemic resources available to the arguers.

Narrative, testimonial, experiential, and emotional reasoning may communicate relevant knowledge without taking the form of an explicit premise–conclusion argument. If the evaluative criteria fail to distinguish between nonconformity and defective reasoning, they may contribute to the unfair assessment or discounting of argumentative contributions. Thus, the possibility of argumentative injustice becomes relevant.

The findings instead reveal a relevant feature of the evaluative procedure: classification outcomes varied with the contextual information supplied, and this variation was not uniform across all categories. The observed variation in classification distributions does not itself establish argumentative injustice. A change may reflect a legitimate reinterpretation of a contribution, while persistent ambiguity may be appropriate when the available information is insufficient. The normative question is whether the standards and informational conditions governing evaluation adequately recognize the argumentative and epistemic features relevant to a contribution.

Contextual sensitivity matters because a contribution's significance may depend on its surrounding exchange and dialogical function. The broader contribution of this study is therefore to identify contextual sensitivity as one point from which the possible relationship between bias detection and argumentative injustice can be examined. Establishing whether a particular classification constitutes injustice requires further normative and empirical analysis.

Declarations

Author Contribution

The author has written the manuscript and prepared figures and also reviewed the manuscript.

Acknowledgement

I would like to acknowledge Prof. Hundleby's work on context in evaluating arguments and her criticism of dominant argumentation structures.

Data Availability

I have used online available data.

References

1. Aberdein, A. (2016). Arguments with losers.
2. Aikin, S. F. (2011). A defense of war and sport metaphors in argument. *Philosophy and Rhetoric*, 44(3), 250–272.

3. Al Tamimi, K. A. W. (2017). *A narrative account of argumentation* (Doctoral dissertation, York University).
4. Aikin, S. F., & Vollbrecht, L. (2021). Argumentative ethics.
5. Bailin, S., & Battersby, M. (2021). Is there a role for adversariality in teaching critical thinking?. *Topoi*, 40(5), 951-961.
6. Banisch, S., & Shamon, H. (2025). Biased processing and opinion polarization: experimental refinement of argument communication theory in the context of the energy debate. *Sociological Methods & Research*, 54(1), 187-236.
7. Bondy, P. (2010). Argumentative injustice. *Informal Logic*, 30(3).
8. Howell, T. (2023). Whataboutisms: The good, the bad and the ugly. *Informal logic*, 43(1), 91-112.
9. Christensen-Branum, L., Strong, A., & Jones, C. D. O. (2019). Mitigating myside bias in argumentation. *Journal of Adolescent & Adult Literacy*, 62(4), 435-445.
10. Cohen, D. H., & Stevens, K. (2016). Virtuous vices: On objectivity and bias in argumentation.
11. Correia, V. (2011). Biases and fallacies: The role of motivated irrationality in fallacious reasoning. *Cogency: Journal of reasoning and argumentation*, 3(1), 107-126.
12. Correia, V. (2012). The ethics of argumentation. *Informal Logic*, 32(2), 222-241.
13. Correia, V. (2018). Contextual debiasing and critical thinking: Reasons for optimism. *Topoi*, 37(1), 103-111.
14. Dziuda, W. (2011). Strategic argumentation. *Journal of Economic Theory*, 146(4), 1362-1397.
15. Festinger, L. (1957). A theory of cognitive dissonance Row. *Peterson and company*.
16. Finocchiaro, M. A. (1994). Two empirical approaches to the study of reasoning. *Informal Logic*, 16(1).
17. Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford university press.
18. Gilbert, M. A. (1995). Coalescent argumentation. *Argumentation*, 9(5), 837-852.
19. Govier, T. (2005). *A practical study of argument*. Belmont, CA: Thomson/Wadsworth.
20. Govier, T. (2021). Reflections on minimal adversariality. *Informal Logic*, 41(4), 523-537.
21. Govier, T., & Hoaglund, J. (1999). *The philosophy of argument* (Vol. 3). Newport News, VA: Vale Press.
22. Henning, T. (2018). Bringing wreck. *Symposion*, 5(2), 197-211.
23. Henning, T. M. (2021). "I Said What I Said"—black women and argumentative politeness norms. *Informal Logic*, 41(1), 17-39.
24. Hinton, M. (2024). Argumentation and Identity: A Normative Evaluation of the Arguments of Delegates to the COP26 UN Climate Change Conference: M. Hinton. *Argumentation*, 38(1), 85-108.
25. Haro Marchal, A. (2025). Mind The Gap: Commitment Attributions in Argumentative Exchanges. *Informal Logic*, 45(4), 560-589.
26. Hundleby, C. (2013). Aggression, politeness, and abstract adversaries. *Informal Logic*, 33(2), 238-262.

27. Hundleby, C. (2023). Social justice, fallacies of argument, and persistent bias. *Argumentation*, 37(2), 281-293.
28. Howes, M., & Hundleby, C. (2018). The epistemology of anger in argumentation. *Symposion*, 5(2), 229-254.
29. Howes, M., & Yap, A. (2026). Introduction to the Special Issue on the work of Catherine E. Hundleby. *Informal Logic*, 46(1), 1-22.
30. Holtermann, C., Lauscher, A., & Ponzetto, S. P. (2022, May). Fair and argumentative language modeling for computational argumentation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 7841-7861).
31. Hundleby, C., Brownstein, M., & Saul, J. (2016). Implicit bias and fallacies of argumentation. *Implicit bias and philosophy*, 1, 238-264.
32. Jayez, J. (2026). (Innocent?) Bias in argumentation. The view from language.
33. Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.
34. Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3), 480.
35. Kienpointner, M., & Kindt, W. (1997). On the problem of bias in political argumentation: An investigation into discussions about political asylum in Germany and Austria. *Journal of pragmatics*, 27(5), 555-585.
36. Kobayashi, K. (2016). Relational processing of conflicting arguments: Effects on biased assimilation. *Comprehensive Psychology*, 5, 2165222816657801.
37. Konverskyi, A., & Kolotilova, N. (2023). Confirmation Bias in Argumentation Processes. *Bulletin of Taras Shevchenko National University of Kyiv. Philosophy*, 2(9), 20-24.
38. LiPaD. (2026). Data: Browse the Canadian House of Commons. <https://lipad.ca/data/>
39. Moulton, J. (2015). A paradigm of philosophy: The adversary method. In *Women, knowledge, and reality* (pp. 11-25). Routledge.
40. Mercier, H. (2016). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, 20(9), 689-700.
41. Mercier, H., & Heintz, C. (2014). Scientists' argumentative reasoning. *Topoi*, 33(2), 513-524.
42. Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences*, 34(2), 57-74.
43. Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press-T.
44. Mizrahi, M. (2018). Arguments from expert opinion and persistent bias. *Argumentation*, 32(2), 175-195.
45. Mueller, N., Löwenstein, D., & Vosgerau, G. (2026). Biased Reasoning, Argument Analysis, and the Principle of Charity. *Argumentation*, 1-29.
46. Oswald, S., & Maillat, D. (2013). Biases and constraints in communication: Argumentation, persuasion and manipulation. *Journal of pragmatics*, 59, 137-140.

47. Perkins, D. N. (1989). Reasoning as it is and could be: An empirical perspective. "Thinking across cultures: The third international conference on thinking.
48. Pimenova, O. (2022). Dominant Discourse in Indigenous Consultations: A Comparative Study of the Crown's Reasoning. *International Journal on Minority and Group Rights*, 30(3), 387-416.
49. Pimenova, O. (2025). "Deliberative Context" Is not the Whole Story of Deliberative Reasoning: the Site C Case of Disagreement Management in Indigenous Consultations. *Argumentation*, 39(3), 451-489.
50. Perkins, D. N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. *Thinking: The expanding frontier*, 177-189.
51. Spliethöver, M., & Wachsmuth, H. (2020, December). Argument from old man's view: Assessing social bias in argumentation. In *Proceedings of the 7th Workshop on Argument Mining* (pp. 76-87).
52. Stevens, K., & Cohen, D. H. (2021). Angelic devil's advocates and the forms of adversariality. *Topoi*, 40(5), 899-912.
53. Tannen, D. (1998). The argument culture: Moving from debate to dialogue. *ETC: A Review of General Semantics*, 55(3), 356-357.
54. Thagard, P. (2011). Critical thinking and informal logic: Neuropsychological perspectives. *Informal logic*, 31(3), 152-170.
55. Tindale, C. W. (2007). *Fallacies and argument appraisal*. Cambridge University Press.
56. Tomasetto, C., Mucchi-Faina, A., Alparone, F. R., & Pagliaro, S. (2009). Differential effects of majority and minority influence on argumentation strategies. *Social Influence*, 4(1), 33-45.
57. Van Laar, J. A. (2006). Don't say that!. *Argumentation*, 20(4), 495-510.
58. Van Laar, J. A. (2007). One-sided arguments. *Synthese*, 154(2), 307-327.
59. Van Eemeren, F. H., & Grootendorst, R. (1992). Relevance reviewed: The case of argumentum ad hominem. *Argumentation*, 6(2), 141-159.
60. van Laar, J. A., & Krabbe, E. C. (2013). The burden of criticism: Consequences of taking a critical stance. *Argumentation*, 27(2), 201-224.
61. Waller, M., Rodrigues, O., & Cocarascu, O. (2024, March). Identifying reasons for bias: An argumentation-based approach. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 19, pp. 21664-21672).
62. Walton, D. N. (1991). Bias, critical doubt, and fallacies. *Argumentation and Advocacy*, 28(1), 1-22.
63. Walton, D. N. (1995). A Pragmatic Theory of Fallacy Studies in.
64. Walton, D. (1999a). Francis Bacon: Human bias and the four idols. *Argumentation*, 13, 385-389.
65. Walton, D. (1999b). *One-sided arguments: A dialectical analysis of bias*. SUNY Press.
66. Walton, D. (1999c). Ethotic arguments and fallacies: The credibility function in multi-agent dialogue systems. *Pragmatics & Cognition*, 7(1), 177-203.
67. Walton, D. (2005). *Fundamentals of critical argumentation*. Cambridge University Press.

68. Walton, D. (2006a). Examination dialogue: An argumentation framework for critically questioning an expert opinion. *Journal of pragmatics*, 38(5), 745-777.
69. Walton, D. N. (2006b). Poisoning the well. *Argumentation*, 20(3), 273-307.
70. Walton, D. (2008). *Informal logic: A pragmatic approach*. Cambridge University Press.
71. Walton, D. (2010). Why fallacies appear to be better arguments than they are. *Informal logic*, 30(2), 159-184.
72. Walton, D. (2012). Using argumentation schemes for argument extraction: A bottom-up method. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 6(3), 33-61.
73. Wolfe, C. R. (2012). Individual differences in the “myside bias” in reasoning and written argumentation. *Written Communication*, 29(4), 477-501.
74. Wolfe, C. R., & Britt, M. A. (2008). The locus of the myside bias in written argumentation. *Thinking & reasoning*, 14(1), 1-27.
75. Walton, D. N., Gabbay, D. M., & Barwise, J. (1996). *Fallacies arising from ambiguity*. Dordrecht: Kluwer Academic Publishers.
76. Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication*, 26(2), 183-209.
77. Zarefsky, D. (2012). A challenge and an opportunity for argumentation studies. *Argumentation and Advocacy*, 48(3), 175-178.
78. Zenker, F. (2011). Experts and bias: When is the interest-based objection to expert argumentation sound?. *Argumentation*, 25(3), 355-370.