The New Ethos

Volume 1, September 2023

Linguistically Analysing Polarisation on Social Media

Ewelina Gajewska, Katarzyna Budzynska, Barbara Konat, Marcin Koszowy, Editors.

Published online and open access by the Laboratory of The New Ethos – Warsaw University of Technology, Warsaw, Poland

Front page design: Medea Kfoczyńska-Łukasz

Online available at https://newethos.org/publications/

Publication date September, 2023

Digital Object Identifier: 10.17388/WUT.2023.0001.AINS

Aims and Scope

The New Ethos Reports series documents the results of workshop meetings organized by the Laboratory of the New Ethos in collaboration with colleagues who collaborate with the laboratory on the regular basis. Its scope covers topics which are of key importance in the interdisciplinary study of ethos, i.e. people's character, with a special focus on studying ethos-related rhetorical strategies and communicative tendencies in social media that are grasped by new technologies that enable us to look at ethos from a perspective which complements the traditional approaches by delivering resources, theories and technologies of ethos. Specifically, the New Ethos Reports contain:

- a summary of the workshop program along with its key results,
- abstracts of flash talks on crucial research areas related to a workshop theme,
- the summary of panel discussions,
- reports from working groups.

This framework can be extended by suitable contributions relevant to a given workshop's theme, such as longer abstracts, literature overviews, and further research outlines.

Contact

Laboratory of The New Ethos – Warsaw University of Technology, Noakowskiego 18/20 room 501, 00-668 Warsaw, Poland, katarzyna.budzynska@pw.edu.pl

Workshop Organised by the Laboratory of The New Ethos.*

Linguistically Analysing Polarisation on Social Media

Edited by

Ewelina Gajewska[1], Katarzyna Budzynska[2], Barbara Konat[3], and Marcin Koszowy[4]

1 Warsaw University of Technology, PL ewelina.gajewska@pw.edu.pl
2 Warsaw University of Technology, PL katarzyna.budzynska@pw.edu.pl
3 Adam Mickiewicz University, PL bkonat@amu.edu.pl
4 Warsaw University of Technology, PL marcin.koszowy@pw.edu.pl

Abstract

Polarisation of society is the key threat to the transition of our communication to social media. It is often stressed that online technologies, such as recommendation algorithms or clustering friends, encourage and escalate divisions into polarised —extreme & isolated— groups, that is into *in-groups* ('us', 'the good guys', 'pals') and *out-groups* ('them', 'the bad guys', 'evils'). While a lot of attention has been paid in psychology and sociology to investigate causes and effects of polarisation, less work has been done to look at *linguistic manifestations* of polarisation. Even though hate speech understood as offensive and emotional language —extensively studied in computational linguistics— often overlaps with polarisation, they are not the same phenomena: I can be vulgar or emotional towards my friends and I can be perfectly polite and cold with my enemies. In this report, we make the first step towards directly addressing the question: how do we *use language when we are polarised*, that is, what are the features of polarising and polarised language? The document is the result of a *workshop* organised in March 2023 by the *iTRUST project* "Interventions against polarisation in society for TRUSTworthy social media: From diagnosis to therapy". The event was structured around flash talks, panel discussion and working groups, bringing together the iTRUST team and collaborators who contribute their approaches from various areas of philosophy, linguistics, psychology, sociology, media studies and computer science.

Seminar 16–18. March, 2023. Radziejowice, PL – https://newethos.org
Digital Object Identifier 10.17388/WUT.2023.0001.AINS
Keywords Communication Behaviour, Rhetorical Strategies, Computational Argumentation, Polarisation of Society, Discourse Analysis, Natural Language Processing

* We are grateful to Dagstuhl *Leibniz Center for Informatics* for granting us the permission to use their template of reports, licensed under the 'LaTeX Project Public License v1.3c' (see https://github.com/dagstuhl-publishing/styles/blob/master/LICENSE.md).

1 Table of Contents

2 Introduction

Katarzyna Budzynska Warsaw University of Technology, PL
Barbara Konat Adam Mickiewicz University, PL
Marcin Koszowy Warsaw University of Technology, PL

The workshop *Linguistically Analysing Polarisation on Social Media* explores language manifestations of communication behaviours in large groups of social media users in the context of polarisation of society. In particular, we look at discussions on the topic of climate change on Twitter and investigate whether there are linguistic indicators or linguistic traces which reveal that a given online discussion is affected by the problem of polarisation: Can we tell, just by looking at the language used in a discussion, that a group of users is dangerously moving towards a division into two extreme camps? Are there traces on the linguistic surface that guide us to conclude that a discussion is polluted with polarisation? What are the types of rhetorical strategies and styles that users employ to express that they belong to different groups and they do not want to have anything in common with one another? And, on the other hand, what are the types of rhetorical strategies used to express that we belong to the same group and share common view of the world?

Figure 1 depicts the basic dynamics of polarising language. First of all, when we encounter polarisation, we should be able to identify linguistic indicators of the existence of *impenetrable wall* between some participants of the discussion. This is the division between 'us', i.e. 'the good guys' (*in-group*), and 'them', i.e. 'bad guys' (*out-group*) [15].

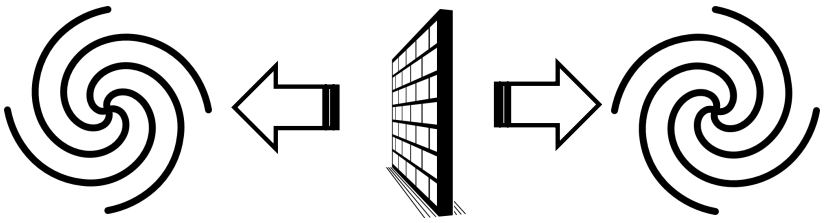

Figure 1 Polarisation in a nutshell. Building a wall between different groups and accelerating the velocity of two opposite forces: (1) gradually moving speakers away from one another between out-groups, (2) while at the same time moving speakers closer to each other within in-groups.

Next, polarisation is far from being a stable state of affairs: once the wall is put up, the distance between out-groups is rapidly growing, as they are pulled apart by the *force of hate* (see arrows in Figure 1). This can be viewed as changing an impenetrable wall into a *bottomless abyss* between people. It means that 'bad guys' become 'evils' or 'demons'. This also means that we are or should be *at war*, as it is presented as our moral duty to fight and defend ourselves against evil forces. As a result, the language of polarisation is filled not only with hate speech, negative emotions and offensive language, but is also full of war metaphors, pro-war rhetoric and descriptions which demonise and dehumanise the enemy.

Finally, the out-groups are pulled apart even more by the force opposite to hate: the *force of love* within in-groups. Imagine two people who stand on the opposite ends of a rope. They wrap this rope around their waist and start to rotate (see vortexes in Figure 1): the rope becomes more and more tight, the atmosphere becomes more and more dense and hostile.

The force of love turns 'good guys' into 'wise ones', 'best pals' or 'one for all, all for one'. Possibly surprisingly, it makes the language of polarisation full of not only negative, but also positive emotions. We should be able to identify many expressions of flattering and agreeing in such discussions, expressions that build intimacy and community amongst members of an in-group and descriptions which idealises and glorifies the friends. Sadly, at this point it does not matter anymore who is right and who is wrong. People, who got drawn into the dynamics of polarisation, becomes whirled into the depth of vortex and eventually everyone loses.

In what follows, we first collect and describe different rhetorical strategies and communication phenomena that play crucial role in linguistic manifestations of polarisation (Section 3: Flash Talks). Then, we sketch the general characteristics of polarising language (Section 4: Panel Discussion). Finally, we discuss in more details the dynamics of communication behaviour typical for polarised online environment (Section 5: Working Groups).

3 Flash Talks

3.1 Ethos

Konrad Kiljan *University of Warsaw, PL*

Ethos plays an increasing role in contemporary political communication, as its participants enjoy previously unprecedented capabilities to shape their own discourse [20] but succeed only if catering to evolving public demands [47]. As the character of the speaker influences the public's reactions of argumentation [52], its perception is heavily affected by affordances of specific platforms [59]. In an overview of ongoing research on that topic, I will discuss how multimodal content analysis enriches our understanding of preferences distribution within specific audiences. Manually annotated corpora and NLP-assisted mining based on psycholinguistic models allow for studies on dynamics of online argumentation highlighting the mobilising and polarising effects of specific rhetorical techniques. Appeals to ethos, in particular, seem to constitute important factors of inter-group polarisation in online spaces centred around the rhetoric of 'good guys versus bad guys'. The gained insights into how various online platforms affect trust in the public sphere [37] lead to evaluation of their impact on democracy.

3.2 Pathos and Emotions

Barbara Konat *Adam Mickiewicz University, PL*

Appeals to emotions have accompanied argumentation since the dawn of the rhetoric. Following the introduction of Aristotelian triad of *logos*, *ethos*, and *pathos* [7], public speakers are aware that the persuasive power lies not only in the strength of the arguments, or the trustworthiness of the rhetor, but also in the ability to elicit, ignite and regulate the emotions of the audience. In *Rhetoric* (Book I, 2), Aristotle ([7]) writes: 'persuasion may come through the hearers, when the speech stirs their emotions. Our judgements when we are pleased and friendly are not the same as when we are pained and hostile'. Following this approach, we are postulating the need of analysis of pathos in polarising language by capturing both emotional appeals and reactions. This task can be realised with the use of computational tools supporting human analyst. Speakers can appeal to emotions in two ways: first, by

using pathotic argument schemes (for example fear appeal, as described by [83]), second, by using emotion-eliciting language (for example words such as 'war' or 'children', see [84]). Human recognition of instances of pathotic argument schemes (i.e. human annotation) can be supported by automatic recognition of emotion-eliciting words used by the speaker to achieve polarising effects. On the side of the audience we can capture the emotional reactions expressed in language using standard sentiment analysis methods [4], which allow us to assess emotions expressed by the recipients of polarising language.

3.3 Rephrase

Marcin Koszowy *Warsaw University of Technology, PL*

The study of *rephrase*, a dynamic and complex communicative phenomenon that consists of modifying an original contribution in order to achieve a variety of persuasive goals [51, 87], has a promising potential for exploring structures, features, and functions of the language of polarisation. There are at least three possible research threads that may initiate a systematic inquiry into the overlap between rephrase and polarisation: (i) the frequency of typically rephrased content as empirical material to identify divisive issues that may mark polarisation processes, (ii) the study of ethotic rephrase that has a potential to boost up speaker's own ethos by at the same time weakening others' ethos – that is itself a polarising tendency, and (iii) employing rephrase as means to intensify an emotional load of a message that may further initiate or reinforce polarisation processes or agendas.

(i) The most general feature of rephrase which is to maintain the similar propositional content to a discourse might already become the subject-matter of the linguistic inquiries into the polarisation phenomena and processes in social media. The study of annotated corpora of rephrased claims and arguments can help linguists and argumentation scholars obtain a novel statistical information about the frequencies of the contents that have been mostly rephrased in a given discourse type or genre. If the frequently rephrased contents can be tested as possibly divisive issues, then the annotation of rephrase might become a way to semi-automatically identify the issues that may lead to polarisation.

(ii) The recent study of rephrase may show how the speakers rephrase ethos for rhetorical gains in order to accomplish two, typically simultaneous tasks, namely boosting up speaker's own ethos by at the same time weakening others' ethos. Given that rephrase might be an effective persuasive means to emphasise differences between speaker's own and other entities' ethos (such as individuals, user groups, communities, institutions, etc.), the knowledge of frequencies of particular rephrase techniques, such as rephrasing ethotic attacks, may become useful for the sake of identifying polarising tendencies.

(iii) The next research thread concerns the overlap between rephrase and expressed pathos. A promising line of inquiry might consist in testing how one of specific rephrase types, namely (de)intensification that changes the degree of previously mentioned qualities overlaps with increasing or decreasing the expressed emotional load of a message. If a given corpus annotated with rephrase types has a relatively high frequency of instances of (de)intensification, that would constitute a point of departure to study to what extent those instances of rephrasing have a potentially high impact of increasing those emotions that might lead towards polarisation.

These three research areas are related to testing hypotheses about the polarising role of rephrase, and, if they happen to be tested positively in terms of proving that some rephrase types or reformulation markers [19] have a polarising impact, such results might

form a contribution to design new strategies to intervene against rephrase-related polarisation tendencies.

3.4 Moral framing

Alina Landowska *University of Social Sciences and Humanities, PL*

Understanding morality expressed in language is essential for detecting *values* behind individual or collective (social) behaviours. In recent years, research directions show an increasing interest in elucidating morals, (e.g., [5, 6, 12, 21, 35, 34, 39, 40, 42, 43]) as well as in quantifying values in natural language processing (e.g., [6, 45, 46, 69, 78, 85]).

Moral values stand for most of human behaviours and are researched in terms of: digital behaviours (e.g., [48]), public life (e.g., [11, 74, 73]), voting (e.g., [82, 61]), political camps (e.g, [22, 30, 76]), justice (e.g., [81]), persuasion (e.g., [28, 55, 86]); leadership (e.g., [27]); message diffusion (e.g., [14, 75]), vaccination (e.g., [16, 48, 77]), social protests (e.g., [9, 60]), hate speech (e.g., [8, 9, 71, 72]), climate change and other environmental issues (e.g., [13, 24, 56]), stem cell (e.g., [18, 53]), abortion (e.g., [54]), terrorism (e.g., [38]), culture war (e.g., [3, 41, 50]).

Moral triggers polarise our debates, whatever they refer to. The important question is why moral framing is so successful. The simplest reason might be that moral communication identifies (potential) transgressions committed by the opponent, and the critique of the opponent's moral characteristics can be easily used against the opponent.

3.5 Entities

Martin Pereira-Farina *University of Santiago de Compostela, ES*

Communication involves discussing things in the world. Speakers, when they talk, commit to the existence of certain things (in a broad sense, both physical things, abstract things, social things, etc.) that are typically shared or disputed by others participating in the communicative interchange [31]. There is a figured world where the speakers produce meaning [29].

This shared world is complex and we propose tackling it through ontological analysis to unpack it. This is achieved by employing *conceptual modelling* [33], which aims to develop a simplified version of the world conveyed in a discourse. It is build upon three key concepts: entities, features and facets.

Entities are non-instantiable things in the world, which can be divided into *atom*, such as Aristotle or a law, or *category* (collections of atoms), such as philosopher or regulation. Each atom is an instance of a category, and a category itself can be an instance of another category. For each entity, a speaker can assert specific features, which predicate a particular value of a property shared by all the instances within a category. For instance, the atom Aristotle has a property named 'Name'. Additionally, features can also encompass associations, denoting specific relationships between entities (atoms or categories); e.g, Aristotle is associated to his students as 'Tutor of'. Lastly, there are facets. When we predicate a value of a property (which can be qualitative or quantitative), such as 'Name = Aristotle' of the atom Aristotle, this is a facet. When we predicate the relationship between the atoms; i.e., 'Tutor of = Alexander the Great', *Alexander the Great* is the reference (facet) for that relationship.

Ontological analysis allows us to reconstruct, in as much detail as necessary, the shared world of the speakers participating in a dialogue. This is particularly relevant in polarisation as it will help us to identify the key points where disagreements are rooted. Understanding the points of disagreements is crucial for seeking how to resolve it. Additionally, it aids in comprehending the speakers' commitments, which also are relevant in the context of polarisation. Furthermore, we can capture how participants change their minds, signifying modifications in their respective worlds, and observe the values and properties the attribute to specific entities. Thus, our main goal is to gain a deeper understanding of polarised debates and dialogues, specifically addressing the entities in the world that are being discussed.

3.6 Concessions

Olena Yaskorska-Shah *Università della Svizzera italiana, CH*

As was discussed in the previous chapter, communication always refers to the shared world of the participants. Since shared beliefs do not need to be persuaded to, they usually remain implicit in the communication, yet take the key role in opinion shaping. Two different communities would not understand each other because they do not have access to the statements, which are not said explicitly in the opposite discourse. This mechanism leads to the phenomena of *deep disagreement* [68] between agents which in the literature is broadly elaborated in the contexts, e.g. of conflict resolution [49] or contemporary politics [1] and political polarisation [23]. This notion is also worth applying in the context of polarisation in social media.

Fogelin [68] says, that 'argumentative exchange can be normal [i.e. the one which would lead to the conflict resolution] when it takes place in the context of broadly shared beliefs and preferences'. As a method for the resolution of deep disagreement, he proposes to find out the deep (i.e., implicit) points of disagreement and discuss them. It is important, that he also notices that deep disagreement is usually not around a single statement, like 'pro-' or 'con-' something, but about the ecosystem of implicit beliefs and preferences. At this point, the question arises how to detect those implicit elements of communication.

Concessions are dialogical phenomena that provide information according to which we can capture the (un)shared beliefs of the communities. Musi et al. [62] analyse concessions as persuasive or non-persuasive types of discourse moves, in which participants introduce their attitude towards some facts. In the research on polarisation in social media, both types provide information about implicit elements of opinion shaping. A data annotation model for concessions has to provide a possibility to show connections between implicit beliefs and explicit statements in order to show the pragmatics of opinion shaping and arguing. This can be expressed e.g. with the Argumentum Model of Topics [67], where the ancient notion of *endoxa* is incorporated as a part of the argumentative practice.

4 Panel Discussion

The material for the discussions on polarising language comprises of tweets on climate change posted by the former US president – Donald Trump. Eighteen such tweets were selected randomly from data set collected with the use of the Trump Twitter Archive[1]:

[1] https://www.thetrumparchive.com/

(1) Donald Trump: *I wonder if the Rutgers coach who had the audacity to yell at the player is a proponent of global warming?*

(2) Donald Trump: *The people that gave you global warming are the same people that gave you ObamaCare!*

(3) Donald Trump: *They changed the name global warming to climate change because the concept of global warming just wasn't working!*

(4) Donald Trump: *The least number of hurricanes in the U.S. in decades. So they change global warming(too cold) to climate change-now what will they call it*

(5) Donald Trump: *We should be focused on magnificently clean and healthy air and not distracted by the expensive hoax that is global warming!*

(6) Donald Trump: *We should be focusing on beautiful, clean air & not on wasteful & very expensive GLOBAL WARMING bullshit! China & others are hurting our air*

(7) Donald Trump: *We are experiencing the coldest weather in more than two decades-most people never remember anything like this. GLOBAL WARMING anyone?*

(8) Donald Trump: *Do you believe this one - Secretary of State John Kerry just stated that the most dangerous weapon of all today is climate change. Laughable.*

(9) Donald Trump: *It's not climate change,it's global warming.Don't let the dollar sucking wiseguys change names midstream because the first name didn't work*

(10) Donald Trump: *The entire country is FREEZING - we desperately need a heavy dose of global warming, and fast! Ice caps size reaches all time high.*

(11) Donald Trump: *There are many Jonathan Gruber types selling the global warming "stuff" - and they really do believe the American public is stupid.*

(12) Donald Trump: *Do you believe @algore is blaming global warming for the hurricane?*

(13) Donald Trump: *We can't destroy the competitiveness of our factories in order to prepare for nonexistent global warming. China is thrilled with us!*

(14) Donald Trump: *The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive.*

(15) Donald Trump: *The Chinese talk of climate change and carbon footprint but don't clean up their factories-but they sell us the equipment to clean up ours!*

(16) Donald Trump: *The badly flawed Paris Climate Agreement protects the polluters, hurts Americans, and cost a fortune. NOT ON MY WATCH!8. I want crystal clean water and the cleanest and the purest air on the planet – we've now got that!*

(17) Donald Trump: *It's really cold outside, they are calling it a major freeze, weeks ahead of normal. Man, we could use a big fat dose of global warming!*

(18) Donald Trump: *Because we have done so well with Energy over the last few years (thank you, Mr. President!), we are a net Energy Exporter, & now the Number One Energy Producer in the World. We don't need Middle Eastern Oil & Gas, & in fact have very few tankers there, but will help our Allies!*

4.1 Polarising Language

John Parkinson Maastricht University, NL
Liesbeth Allein KU Leuven, BE

Katarzyna Budzynska Warsaw University of Technology, PL
Giulia D'Agostino Università della Svizzera italiana, CH
Ewelina Gajewska Warsaw University of Technology, PL
Amalia Haro Marchal University of Granada, ES
Zlata Kikteva University of Passau, DE
Konrad Kiljan University of Warsaw, PL
Barbara Konat Adam Mickiewicz University, PL
Marcin Koszowy Warsaw University of Technology, PL
Alina Landowska University of Social Sciences and Humanities, PL
Maud Oostindie Maastricht University, NL
Martin Pereira-Farina University of Santiago de Compostela, ES
Jennifer Schumann University of Fribourg, CH
Narjes Sheikh Asadi Università della Svizzera italiana, CH
Yana Sviatsilnikava Warsaw University of Technology, PL
Maciej Uberna Warsaw University of Technology, PL
Olena Yaskorska-Shah Università della Svizzera italiana, CH
Ramy Younis University of Fribourg, CH
He Zhang Warsaw University of Technology, PL

When exploring real-world examples of polarising language, we started with the in-group / out-group distinction that emerged in the working groups on anti-vaccination discussion and forms the basis of the semantic approach, discussed in section 5.1. The essential element of the approach is the strategic use that speakers make of further distinctions between demonised out-groups and victim out-groups, coded *OGD* (out-group demonisation) and *OGV* (out-group victimisation) in the Figure 2.

However, another group took this a step further by linking it with narratological approaches that differentiate between the particular and often rather fixed narrative *roles*, which are defined by *ascribed characteristics* and *action*, and the variable, fluid assignment of *actors* to those roles [10], fitting them into story arcs which are themselves meaningful to *audiences* and make the entire construct of in-groups and out-groups legible [36, 58]. Such roles and story arcs describe who does what to whom, when and why. Trump's tweets cast himself in the role of the hero who saves a variety of victims ('Americans', 'manufacturers', 'we') by identifying and punishing a series of bad guys ('the Chinese','Jonathan Gruber types') for their bad actions (yelling and selling, talking about climate change, 'dollar-sucking', and so on).

This goes beyond the OGV/OGD distinction. 'Americans' are labelled as *both* smart enough to 'see' what the bad guys are trying to do – they are being integrated into the in-group in some way – but also as victims, hoodwinked by the demonised bad actors. Part of what is going on here is an agency claim – both bad guys and the hero have agency; everyone else does not, and needs the hero to save them lest they fall victim to the demonised other. In short, the in/out distinction misses the essential role of the *audience*, something that is a construct created by those who try to make authoritative representation claims [44].

The polarised good/bad categories are mutually reinforcing symbolic constructs: what defines an actor as bad is that they perform a bad action; what makes an action bad is that it is performed by a bad actor. However, it is not the actors and their real-world actions per se that are bad, but what the hero *says* they are. The hero exercises power by assigning roles to actors and labelling their actions, shaping the meaning that audiences ascribe to the actions.

This raises an important point in the tweets: that they are an extended effort, over time,

to set rhetorical traps. *Anything* a bad actor says or does becomes associated by the audience with bad action; any mentions of one of the 'bad' acts or attributes in anything other than a disparaging way labels the speaker as themselves being a bad actor. That means that any non-disparaging mention of climate change labels the speaker as themselves a bad guy; which in turn means that *anything* Al Gore says, or China does, for example, can be associated with bad action. It is an attempt to make reasoned debate and consensus impossible: you cannot talk seriously about the subject without labelling yourself as someone who is trying to fool or swindle the American people. This is why such polarisation is such a dangerous strategy: it cannot be resolved simply by appeal to 'the facts' because even raising the issue is what marks you out as unreasonable, a bad guy.

5 Working Groups

Analysis on polarising language was conducted on 5 examples of discussions on the topic of conspiracy theories about COVID-19 vaccines on Reddit[2]:

(19) a. Interesting-Letter67: *It's a division piece, and very common. Make the vaccinated hate the unvaccinated. Divide and conquer. These news agencies aren't going to last through the next war.*

b. Savagfux: *We're already in the midst of the 'next war'.. they're succeeding because everyone is waiting for a bomb to drop.. there's no need for bombs.. they're winning with physiology.. it's crazy that no matter where I go there's 2 things that I always, 100% of the time, end up taking about.. politics and COVID. our heads are spinning.. we're at war friends and it's nothing like we've ever seen before.*

c. LordOFtheNoldor: *Silent weapons for quiet wars*

d. Significant-Knee-150: *This is textbook psychological warfare. We are in a civil war, just not fought with kinetic force, yet.*

e. Savagfux: *It may not end up in any physical / kinetic force locally. I do believe that we will see catastrophic conflict between the US and China in the next 5 years.*

(20) a. empathetical: *I usually can't understand the mentality of vaccinated ppl thinking unvaxxed are screwing them up but then again... they ran to get vaccinated with an untested substance so that is all you need to know to understand why these people believe these things. Braincells = 0.*

b. Homegrown98: *That's not the only explanation for it. I keep talking to my friends and family about all the things that are fishy, all the lies they sell us etc. One of them once said to me „Look, I know that you're right, but I don't want to think about this kind of thing'. They proactively close their eyes to their surroundings and comply because any other way would be uncomfortable or inconvenient to them.*

(21) a. melikestoread: *Don't ever admit the vaccine doesn't work. A virus that initially only kills 1% or less of people is killing 1% or less after vaccinations. Yet tell people it works and they will believe it. If everyone was vaccinated the story would be the vaccine only works for such strain and not for the deadlier strain.*

[2] Examples are available in the *IAT Conspiracy Reddit1* corpus at http://corpora.aifdb.org/IATConspiracyReddit1

Then when more die its because everyone didnt have their 8th vaccine. The cycle continues and pharma laughs and laughs picking up their billions from taxpayer funds. I love seeing all the free COVID test and free COVID vaccine signs just wait until we have 30 trillion in debt and everyone will say whaaat how did that happen? The free wars and free vaccines and free 100s of millions we giveaway in aid to other countries yet we dont have universal healthcare. My biggest peeve is people wanting forced vaccines but wont even guarantee you get proper health coverage in case you have a bad reaction to the COVID vaccine. Why is there no liability to the COVID vaccine companies!

(22) a. supersecretaccount82: *Isn't it pretty well known at this point, even among the dimwitted NPC types who follow CNN, that the vaccinated are still spreading COVID? I mean just a day or two ago 6 guys on the New Orleans Saints tested positive and they're all double vaxxed.*

b. MGTOWayoflife: *The countries with the highest vaccination rates have the highest amount of cases. Reading between the lines that tells me that the vaccines suck for what they are advertised to do. Everyone here knows what they are really supposed to do and in that regard it's working.*

c. cpiq84: *Unfortunately, a lot (not all) fully vaccinated people have completely let their guard down and are going about their lives, pre-pandemic. This is a big factor as to why they are contracting and then spreading the virus. This is a big factor as to why they are contracting and then spreading the virus*

d. Born_crazy-: *Isn't that what they are supposed to do when vaxxed?....if it was one.*

e. sjryan: *No no, you're supposed to get vaxxed and nothing changes, least of all your immunity and transmissibility. The blood clots just mean it's working!*

f. cpiq84: *I guess dude.*

g. ArmedWithBars: *That was the original idea. Then they came to find out that the vaccines weren't nearly as effective at stopping infections like they thought. If anything the only thing they really help is reduced chances of serious side effects. Are you a 65yr old obese diabetic with heart issues? Vaccine might be a good idea cause surviving COVID is a real dice roll. Are you a fit 25yr old with no health issues? Well that's where I'm having an issue seeing the point of vaccinating. There just isn't enough evidence to breakdown at this moment. Coming from someone in NY with family in the healthcare sector I can assure you that COVID isn't fake. I do believe the effectiveness of the vaccines is overblown and the response to this failure is booster shots, which there's no concrete proof this will even help. I'm not sure why masks were such a hard concept for people. They helps prevent the spread of COVID from infected wearers, it doesn't protect the wearer. It's been used in other countries for decades. Your sick? Wear a mask to help protect the people around you. It's a common courtesy, somehow in the US it became a left vs right bullshit fight.*

h. twichy1983: *Mmmmmmmmmmmmmm. I agree with all of that except the wearer not being protected. I've been sporting N95 the whole time. It absolutely protects the wearer. COVID may be smaller than the filter, but its transmission particulates are large enough to get caught.*

i. ArmedWithBars: *N95 is a different ball game. I'm speaking of common medical masks or basic face coverings. I'd be interested in seeing COVID stats if everybody worse a n95 mask nationwide. I'm guessing they would plummet.*

j. Born_crazy-: *I appreciate what you're saying, but you are conflating one thing with another. Respiratory symptoms don't automatically infer infection.*

(23) a. PorcelainPoppy: *I saw this, too. Very convenient narrative. The propaganda in this article is so exploitative. This woman got vaccinated twice. If the vaccines are so effective, how exactly did she get infected with COVID from unvaccinated people? How do we know she didn't die from complications related to the second vaccine she had just received? Also, almost everyone I know who got vaccinated ended up getting some form of COVID from the vaccine itself. Maybe she was infected with COVID when she received the second jab, as many people do fall ill with COVID after receiving the COVID vaccine. The thing that gets me, every time, is the fact that government authorities, the CDC/NIH, and big pharmaceutical companies claim the COVID vaccine is safe and effective, but this story proves that it's actually not effective at preventing COVID-19 at all. This lady had just gotten her second jab. If the vaccine is truly effective, why did she end up catching COVID and dying from COVID within days of the second vaccine? Seems a lot more plausible that complications from the second vaccine actually killed her. We have no way of verifying the validity of her actual cause of death or validating that she indeed became infected with COVID-19 because she had contact with an unvaccinated individual who had COVID-19. If she did indeed become infected with COVID because of contact with an unvaccinated person, then how exactly did the two vaccines protect her? Not very well, apparently.*

b. Homegrown98: *The vaccine she just received? Infected when she received the second jab? The way you word it makes me think you didn't really read the article all too well, it says she got the second jab six months prior. Don't wanna start a fight, just pointing it out.*

5.1 Semantics of Language of Polarisation

Amalia Haro Marchal *University of Granada, ES*
Maud Oostindie *Maastricht University, NL*
Marcin Koszowy *Warsaw University of Technology, PL*
Maciej Uberna *Warsaw University of Technology, PL*
and He Zhang *Warsaw University of Technology, PL*

5.1.1 Background

The phenomenon of political polarisation has been widely studied from different perspectives and for different goals. It can be understood as a process whose purpose consists in changing the public opinion from a normal distribution of views and opinions to a situation in which two or more groups consider people from opposing sides as illegitimate political agents [2]. Recently, much scholarly work has focused on the phenomenon of polarisation in online public spaces (e.g., [57, 64, 79]). In the discussion about this phenomenon, one of the most relevant issues has to do with the different mechanisms through which polarisation arises.

5.1.2 Two mechanisms of polarisation

In analysing some examples from Reddit, our working group identified two different mechanisms of polarisation in online discourse. The first one is what might be called *in-group*

integration (*IGI*), while the second one can be identified as *out-group demonisation* (OGD) (see section 4.1 above). In both mechanisms, the in-group is speaking to its own members. For polarisation to happen, the idea of two distinct and homogenous groups needs to be created. According to this, there is a morally good or enlightened in-group, and a morally bankrupt or 'blind' out-group. These two mechanisms of polarisation take the form of an 'us versus them' dynamic, where nuances and internal differences are downplayed, and a coherent 'other' is created (see also [57]). Furthermore, in this distinction, there is a process of portraying the other in a homogenous way, in which each group acquires a certain image associated with different descriptions. What can be observed is that by portraying the other in a specific way (stupid, evil, sheepish), the in-group is implicitly defined as the opposite (smart, enlightened, good). This is reminiscent of the Orientalist mechanism as described by Said [70], where 'the Orient has helped to define Europe (or the West) as its contrasting image, idea, personality, experience' [pp. 1-2]. In this way, IGI and OGD are two mechanisms that work together to create an image of a black-and-white world, with two clear, coherent, and opposing groups of people.

The two mechanisms of IGI and OGD can be observed in example (20-a) from the Reddit data:

> (20-a) empathetical: *I usually can't understand the mentality of vaccinated ppl thinking unvaxxed are screwing them up but then again... they ran to get vaccinated with an untested substance so that is all you need to know to understand why these people believe these things. Braincells = 0*

In this example, the user empathetical makes use of the 'us vs them' mechanism in different ways. Firstly, the user makes an explicit distinction between the vaccinated and unvaccinated people by pointing out that there is a particular mentality of vaccinated people, who think that the unvaccinated people are 'screwing them up'. The mechanism used here is the one that has been previously called *out-group demonisation* (OGD). By explicitly saying that the vaccinated people think that the unvaccinated are screwing them up, the user is emphasising that there is a group of people who are not smart, i.e., that are portrayed as stupid. And this is reinforced by what follows: '(...) they ran to get vaccinated with an untested substance so that is all you need to know to understand why these people believe these things. Braincells = 0'. Here, the user offers support to the idea that the vaccinated people are not the smart people. Furthermore, in this specific claim, the two mechanisms of polarisation previously mentioned are also at play. The mechanism OGD is at work because the user is again distinguishing the other group by the use of 'they', and 'these people', and continuing this by claiming that they (the vaccinated people) get vaccinated with a substance that is not tested, thus showing how (for instance) stupid they are. This is even more clearly expressed by the utterance of 'Braincells = 0'. By uttering this, the user shows that they, the in-group members, consider 'them', the out-group members, as having zero brain cells, i.e., as those who are stupid in the context of the topic that is being discussed in the Reddit post. Thus, we can observe how there is a definition of the out-group members as stupid people.

The mechanism of the IGI is also working here. The user claims '(...) that is all you need to know to understand why these people believe these things. (...)'. Here, the user is referring to the in-group members by pointing out that all of what they, understood as 'us', need to know to understand the mentality of vaccinated people is the fact that they, 'them', ran to get vaccinated with a substance that is not tested. Again, the distinction between both groups is made by defining the out-group members as people who are stupid, while the in-group members are implicitly portrayed as the smart ones, i.e., those who know that the

vaccine is an untested substance and that people who decided to not get vaccinated with this substance are the smart people.

5.1.3 A third mechanism: out-group victimisation

We established that two separate but intertwined mechanisms are at play in the process of polarisation, namely: *in-group integration* (IGI) and *out-group demonisation* (OGD). After a bit more analysis of the data, however, we identified a third mechanism at play in polarising discussions (see also section 4.1 above). We noticed that, while the in-group continues the mechanism of IGI, they do not only engage in OGD, but also in something which we term *out-group victimisation* (OGV). The in-group creates not one but two clear out-groups: 1) the evil, malicious demons, and 2) the stupid, sheepish victims. While the demons are the true enemy (see also 5.2.3 for a further development of the related notion of 'common enemy' tactic), the victims are merely seduced by the demons. The idea is that if only the victims could see the light, they would join the in-group. The three groups are schematically portrayed in Figure 1, where THEM1 are the demons and THEM2 are the victims. As the image explains, the demons aim to create a false (as perceived by the in-group) division line, between US on the one hand, and THEM1 and THEM2 on the other. The real division line, however, is between THEM1 on the one hand, and US and THEM2 on the other.

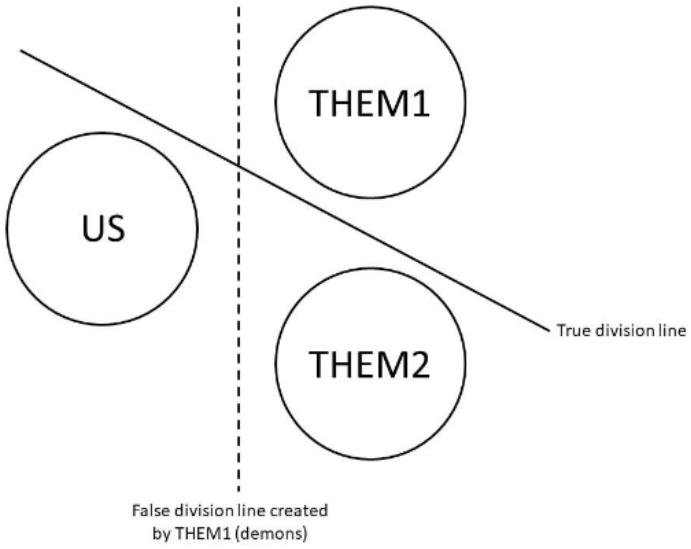

Figure 2 Schematic illustration of polarisation.

These three mechanisms are illustrated in the following example from the Reddit data, which is the response to the comment from user empathetical which is analysed above.

> (20-b) Homegrown98: *That's not the only explanation for it. I keep talking to my friends and family about all the things that are fishy, all the lies they sell us etc. One of them once said to me 'Look, I know that you're right, but I don't want to think about this kind of thing'. They proactively close their eyes to their surroundings and comply because any other way would be uncomfortable or inconvenient to them.*

In this example, the user Homegrown98 makes a clear distinction between two out-groups. The user specifically uses the term them/they to refer to two different groups. The first 'they'

appears in the phrase 'all the lies they sell us etc.', which refers to the THEM1 group, or the demons. The second 'them', in the phrase 'One of them once said to me', refers to THEM2, or the victims/sheep. The mechanism of OGD is visible in the first use of 'they', whereas the mechanism of OGV is present in the second use of 'them'. The implication is that, if only THEM2 (Homegrown98's friends and family) would see the light and not 'close their eyes to their surroundings', they would belong to US, to the in-group.

5.1.4 Conclusions and further remarks

What these three mechanisms (IGI, OGD, OGV) between these three groups (US, THEM1, THEM2) show us, is that polarisation is not as simple and straightforward as we previously suspected. In fact, the aim of polarisation and its precise location are continuously (re)negotiated, and shift between groups. Both in- and out-groups are strategically and discursively created to strengthen the position of the in-group. Linguistic, emotional and ethotic cues that can help us understand these three mechanisms need to be further investigated, but we can make a start with some initial observations.

- Clear linguistic cues that help us understand when polarisation (specifically, IGI, OGD, and OGV) is happening are the uses of terms like: us, we, them, they.
- Emotions that prevail in IGI are trust and joy, whereas the emotions that are visible both in OGD and OGV are disgust, pity, and anger.
- While self-referential ethotic claims prevail in IGI, ethotic attacks are highly visible in OGDand OGV. In OGD the attacks often take the form of proper demonisation, whereas in OGV the attacks mostly imply that the attacked is stupid.
- An important mechanism in polarisation is the notion of 'reported speech' or 'reported intent': '*they* say that *we* are evil' or '*they* want to control *us* by doing this'.

In this section of the report, we have identified the mechanisms that underlie polarisation in the online public sphere. We believe this to be a fruitful starting point for both empirical and conceptual work on polarisation. In order to reduce polarisation and improve deliberation in online spaces, we need to understand the phenomenon better. The list of bullet points above gives clear directions for future research starting points.

5.2 Integration and Disintegration in Language of Polarisation

Zlata Kikteva *University of Passau, DE*
Narjes Sheikh Asadi *Università della Svizzera italiana, CH*
Konrad Kiljan *University of Warsaw, PL*
and Olena Yaskorska-Shah *Università della Svizzera italiana, CH*

5.2.1 Introduction

In the *Integration and Disintegration in Language of Polarisation* section, we further discuss the communication patterns of the polarised groups. Through an analysis of the examples listed at the beginning of section 5, we have identified two major trends of in-group communication upon which we aim to shed some light. The first one, indicating integration, is related to the continuous *reinforcement* of the ideas within the group which is often expressed via 'rephrase to agree', the second one, referring to disintegration, is based on the idea of a *common enemy* for both sides of the polarised discourse introduced by one of the groups. By exploring these trends we aim to gain a deeper understanding of the polarised language.

5.2.2 Reinforcement strategy

In the material selected for discussion of polarisation listed at the beginning of section 5, there are frequent cases when the discussants do not disagree with each other but rather support the same standpoint with different premises. This consolidation of the information could be named as rephrase to agree or reinforcement. According to Van Emeren, Gootendorst, and Henkemans [80], complex argumentation can always be broken down into single arguments for analysing argumentation structure. They refer to this kind of reinforcement as *multiple argumentation*. To clearly illustrate the multiple argumentation, we refer to example (19). The argumentation is reconstructed as all premises support the same standpoint, each argument is assigned the number of the standpoint followed by a number of its own, including 1.1, 1.2, 1.3, and so on (see Figure 3). Each separate argument has an arrow leading to the standpoint. For instance, when the standpoint of 'creating a division between vaccinated and unvaccinated' is presented by Interesting-Letter67, the others agree and bring up some premises for supporting that. To present the results of the analysis concisely, the schematic overview is presented in Figure 3.

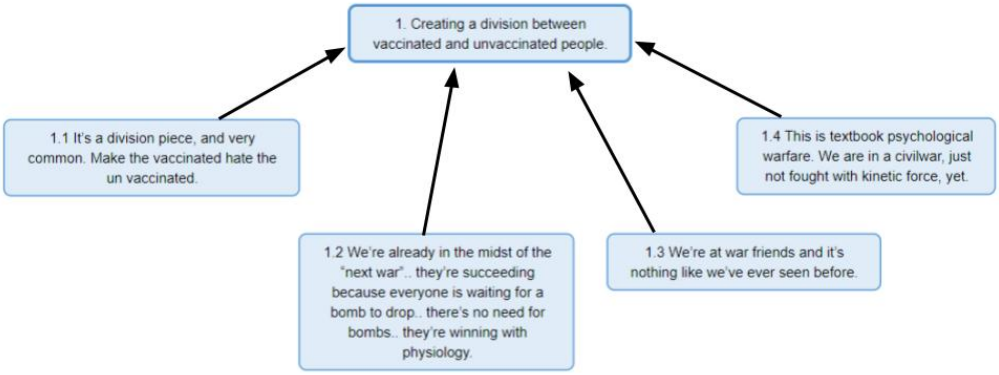

Figure 3 *Schematic overview of a multiple argumentation*

5.2.3 Common enemy tactic

The examples that we are working with are from a forum dedicated to conspiracy theories. According to Douglas et al. [26], conspiracy theories are ways of explaining major political and social events by attributing the reasons behind them to secret plots carried out by powerful actors. In the examples that we studied, we observe several references to some kind of powerful entity that is supposed to be at fault for current problems, such as the news agency CNN and 'COVID vaccine companies'. While the use of a third-party actor is a behaviour typically attributed to believers of conspiracy theories, we argue that it can also be viewed as a tool used by one polarised group in order to bridge (or pretend to bridge) the gap between two groups. We adopt the convention of the in-group and out-group terminology when discussing different sides in the polarised discourse, however, we also discuss a notion of a third party, which we refer to as a *common enemy*. The common enemy is not a necessarily

real entity but rather a rhetorical device used by the in-group. Even when the said entity does exist in real life, it may be perceived differently from the perspective of the out-group.

Not all members of the in-group demonise the out-group and they can go as far as acknowledging the divide between the groups like in example (19-a) when Interesting-Letter67 refers to something, potentially an article as a 'division piece' and when in example (19-b) Savagfux calls the members of the out-group 'friends'.

(19-a) Interesting-Letter67: *It's a division piece, and very common. Make the vaccinated hate the un vaccinated. Divide and conquer. These news agencies aren't going to last through the next war.*
(19-b) Savagfux: *We're already in the midst of the 'next war'.. they're succeeding because everyone is waiting for a bomb to drop..*

This segment illustrates that while the commentators acknowledge polarisation, they do not consider the other group to be their true enemy, instead they hint at a third party that is driving this division forward. While the enemy in this example is left unnamed, commentators in examples (21), (22), (23) mention the third-party actors by names, e.g., government authorities, the CDC/NIH, and big pharmaceutical companies.

Whether the idea of a common enemy can be observed in different polarised groups warrants further exploration as well as the way the use of this strategy can impact communication within polarised communities and beyond them. Finally, it is also worth considering the idea that the out-group can often be viewed as a victim of third-party influence rather than an enemy actor in their own right.

5.3 Brave-hearting Language of Polarisation

Liesbeth Allein *KU Leuven, BE*
Jennifer Schumann *University of Fribourg, CH*
Ewelina Gajewska *Warsaw University of Technology, PL*
and Martin Pereira-Farina *University of Santiago de Compostela, ES*

5.3.1 Background

Braveheart is a 1995 American historical drama based on the epic poem *The Acts and Deeds of Sir William Wallace, Knight of Elderslie.* In this story, the heroic patriot Sir William Wallace unites 13$^{\text{th}}$ century warring Scottish clans to fight against a powerful common enemy, England, in the First War of Scottish Independence. We argue that *brave-hearting* is a rhetorical strategy used to polarise society. Here, brave-hearting is considered the act of framing both agreement and disagreement on a societal topic as a war. Its main characteristics can be summarised as follows:

- Brave-hearting divides society into two *conflicting camps*, where one is presented as the good and the other as the enemy, which follows IGI vs OGD mechanisms (§4.1 and §5.1) and adopts the common enemy tactic (§5.2.3).
- Their conflict on (a) certain topic(s) has arguably brought them to a point of no return and the camps seem to be *at (the verge of) war.*
- Brave-hearting is foremost adopted by the self-declared good.
- The so-called evil enemy is considered more powerful, be it financially, politically and/or socially. Note that power does not necessarily positively correlate with group size; the good could be a woke minority or the man in the street.

- Brave-hearting is marked by specific, war-related vocabulary and rhetorical structures, which is discussed in §5.3.3.

The framing of the in-group/out-group conflict as a war is a prime property of brave-hearting, distinguishing it from related polarisation strategies like robin-hooding (later discussed in §5.4).

5.3.2 Discussed Problems

Why is brave-hearting problematic?

- The war frame adopted in brave-hearting raises a strong call to action. It suggests that *no mediation* is possible and *violence* is the only option left.
- While enforcing a sense of *comradery* within the in-group, brave-hearting simultaneously incites a strong antipathy to, even *hatred* of, the out-group.

5.3.3 The Proposed Approach

As the previous sections have established, brave-hearting creates or further deepens polarising situations. It does so by setting a divisive atmosphere, establishing an in-group that sees itself as less powerful and is driven by a *sense of injustice*, as well as an out-group that, according to the in-group, is considered to be the powerful evil enemy. This type of divisive mechanism can be identified and characterised by looking at a variety of linguistic cues[3] people use (see e.g., [25]). This section discusses a selection of linguistic cues that can point toward brave-hearting in polarising discourse.

In general, people adopting a brave-hearting language, seem to use a vocabulary that does not leave room for many grey areas, meaning that the division created through their discourse is often a matter of black or white. Either people sympathise with the in-group or they are part of the out-group, the enemy. The creation of an in-group can also entail different smaller in-groups merging to fight one common enemy. However, in either case, compromising or finding consensus becomes difficult as mediation does not appear to be possible. Linguistically, this can be reflected in the choice of polarising words such as 'hate', 'division', and other emotionally loaded expressions. By describing the situation around a topic of discussion with these words, it is clear that the opponents fight for two radically different camps and the battle lines are hardened. The verb 'hate', for instance, is connotated with a highly negative attitude toward the opposing party and does not leave room for a middle ground.

But polarisation is not always as extreme from the start. Sometimes it gradually increases in depth throughout the construction of the discourse, culminating in an *escalation* as the following example shows:

> (19-a) Interesting-Letter67: *It's a division piece, and very common. Make the vaccinated hate the unvaccinated. Divide and conquer. These news agencies aren't going to last through the next war.*

Note that the person first talks about a 'division piece', introducing the idea of there being two camps. This is further accentuated by using the verb 'to hate' (see also above). They then talk about 'conquest, leading them to describe the situation as a *war*. A feature that

[3] The linguistic cues mentioned in this section are all highlighted in italics and appear in examples (19-a) to (23-b) listed at the beginning of section 5.

emphasises this sentiment of division, this feeling of underdogs having to fight against the all-powerful, is mirrored in the construction of a lexicon around the idea of *being at war*. To a certain extent, invoking the imagery of war *romanticises* the comradery against a common enemy within the group. People use a whole range of expressions rooted in the word *war* ('quiet wars', 'psychological warfare', etc.) or semantically related to it ('conquer', 'weapons', 'bomb to drop', 'armed'[4], etc.). The use of these words varies in application, meaning that sometimes they are used in quite literal senses[5] and other times they have a more metaphorical nuance[6] to them. But even in their metaphorical use, referring to the concept of war constitutes a frequent (see e.g., [25]) and sometimes strategic move (see e.g., [66] and [65] for a discussion on the use of metaphors in argumentative contexts) in polarising context.

A certain number of linguistic cues can be used to identify the in-group. When referring to like-minded people, expressions like 'friends', 'bonds', and 'friendship' are frequently used to describe themselves. Furthermore, they often refer to personal pronouns in the plural form (e.g., 'we', 'us', 'our') or inclusive expressions (e.g., 'everyone here') to talk about their group. In addition to that, they can sometimes use some type of *jargon* that is understood by people from within the in-group, thus speaking the same language and sharing a common ground, and less likely by people belonging to the opposing field. An example can be seen below:

(22-a) supersecretaccount82: *Isn't it pretty well known at this point, even among the dimwitted NPC types who follow CNN* (...).

The acronym NPC stands for *non-player character* [63] and finds its origins in the gaming community, where it describes a character in a video game that is not controlled by the player. Here, it is used to designate a person that does not think for themselves and is controlled by the enemy. As explained in §5.3.1 and §5.3.2 this is typical of brave-hearting insofar as the in-group feels a sense of superiority, presuming that they know better than the rest, and consequently, leaving them with the impression of being more alert, astute and woke. Yet, brave-hearting not only consists of reinforcing in-group cohesion. It can also make use of the opposite mechanism, namely adopting language that implies a division when talking about the enemy. They frequently refer to the enemy by using personal pronouns in the plural form (e.g., 'they', 'them') or referring directly to the institutions they feel threatened by (e.g., pharma, news agencies).

As this brief synopsis has shown, looking at linguistic cues that are part of brave-hearting strategies can provide a useful way to uncover instances of polarisation within discourse. Polarisation uses divisive language to reinforce the creation of an in- and out-group setting, insisting on the idea that the good in-group needs to pick up arms against a bigger evil, just as Scotland fought against England in *Braveheart*.

5.4 Robin-Hooding Language of Polarisation

Giulia D'Agostino *Università della Svizzera italiana, CH*
Ramy Younis *University of Fribourg, CH*

[4] Note that this type of language is sometimes also used in the usernames as in example (5g).
[5] See e.g., the use of *psychological warfare* in 2d.
[6] See e.g., the use of *the next war* in 2a.

Barbara Konat *Adam Mickiewicz University, PL*
and Yana Sviatsilnikava *Warsaw University of Technology, PL*

5.4.1 Introduction

Robin Hood is a well-known legendary character of the English folklore, risen to popularity since Late Middle Ages. Whatever his origins and his early adventures – different in the numerous versions of the story – he is depicted as a skilled archer and swordsman who acts as a heroic, pure-hearted outlaw robbing from the rich and giving to the poor with the help of his 'Merry Men' band. The *Robin Hood properties* on which we will focus and draw a simile for our analysis are:

(a) The overall 'fighting/rebelling against tyranny' trope;
(b) The 'protection of the underdog' attitude;
(c) The 'construction of a cohesive community' ability.

We will argue that the Robin Hood rhetoric, which might be called robin-hooding, engrafts on the paired tendency of in-group safeguarding and out-group evangelisation, based upon the assumption of a three-pole opposition on two axes.

5.4.2 Fabricating a common enemy and blaming it

With respect to the 'fighting/rebelling against tyranny' trope, first and foremost we claim that in polarised interactions on social media such as those under observation, one is typically able to trace back and recognise two different and independent 'them' referents. On the one hand, we might observe all the social media users that would be reached by some content, all equal on a horizontal level although with different views (thus poles): here we find a common 'us vs. them' dynamic on the grounds of shared beliefs about the outside world. On the other hand, however, we also need to acknowledge some superordinate 'them' – a vertical otherness, nameless or indistinct, that represents the (all-)powerful common enemy, inescapably evil, which is often accused of being the primary cause of misunderstanding, division, and *war* among common people (for a in-depth analysis of conflict dynamics we refer to the discussion in §5.3.1 for brave-hearting). Note that the latter 'them' is not dissimilar from the THEM1 referent introduced in §5.1, whereas the overall strategy is structurally described in §5.2.3.

As argued previously, the language reserved for horizontally levelled 'others' is of pity: the in-group knows best, whereas the out-group is invariably blind/unaware/generally witless. The real enemy, though, is the one who *knows too well* and leverages on pre-existing divides and potential breakpoints to keep the swarming indistinct mass of humanity busy with trivial matters, not to unite against the actual problem - a dynamic already presented as OGD processes. The 'powerful them' are typically blamed for manipulating the public opinion via mass communication to succeed in such an ambitious goal; thus press, journalists, and broadly speaking all famous personalities with an opinion divergent from the in-group belief are accused of being weak-willed minions of the big powers, as exemplified by turn (19-a). It should be noted that this identification of a powerful group that seeks to surreptitiously manipulate the masses suggests an interesting connection between polarising language and the language employed in conspiracy theories. As Jovan Byford ([17]) points out, conspiracy theories are 'marked by a distinct thematic configuration, narrative structure and explanatory logic, as well as by the stubborn presence of a number of common motifs and tropes', and, more specifically, they 'all contain within them the view that a historical or political event

(or a series of events) occurred as a consequence of a carefully worked out plan, plotted in secret by a small group of powerful individuals'.

Since big powers cannot be addressed directly, or are not expected to reply, or in any case engaging in a conversation with them is not the primary goal of the productions under observation, the characteristics we will be describing below are to be intended valid at the horizontal level, i.e., among peers.

5.4.3 Outcasts welcome

Turning to the 'protection of the underdog' attitude, a preliminary observation of the examples should take into account the remark that they all represent what we might call *group enhancement instances*: social platform users establishing or reinforcing a polarised echo chamber they deeply feel connected to, aiming at strengthening the cohesion between its members - similarly to IGI strategies - but also, at the same time, wishing to lure some more 'affine souls' into the group. The latter tendency, however, does not deny the underlying truth of the 'with me or against me' mantra exposed in the previous sections. This leads to permeability between the US and THEM2 groups described in §5.1 via (more or less stable) attraction of the latter into the former.

Under this assumption, it is valuable to recognise the double nature of the interaction. On the one hand the in-group is nurtured, both in positive and negative terms: the *constitutive rules* are laid out with respect to topic and stance, but also the 'others' are framed as those lacking the essential condition for being part of the group and on whom negative traits are projected. On the other hand, it is traceable a (somewhat) welcoming nature of the in-group, an openness towards external members subject to the acceptance of some of the constitutive rules of the group: they can be still saved from ignorance and join the enlightened ones, differently from the 'them from above' (THEM1). To this respect, example (22) constitutes an intriguing instance of concession (see (22-i)) and partial opening up (see (22-j)) towards an out-group member intruding a in-group conversation (see (22-h)):

> (5-h) twichy1983: *Mmmmmmmmmmmmmm. I agree with all of that except the wearer not being protected. I've been sporting N95 the whole time. It absolutely protects the wearer. COVID may be smaller than the filter, but its transmission particulates are large enough to get caught.*
> (5-i) ArmedWithBars: *N95 is a different ball game. I'm speaking of common medical masks or basic face coverings. I'd be interested in seeing COVID stats if everybody worse a n95 mask nationwide. I'm guessing they would plummet.*
> (5-j) Born_crazy: *I appreciate what you're saying, but you are conflating one thing with another. Respiratory symptoms don't automatically infer infection.*

It should not be forgotten that, as a baseline, polarised communication is meant to be in-group aimed ('they' are talked *about*, not *to* or *with*), until an out-group user jumps in the conversation despite not being the intended/primary public [32].

On a distinct note, we remind that the current report focuses on mechanisms that involve people perceiving a 'pole' as a 'group'; dissimilar dynamics can be uncovered instead in the case of 'polarising personalities' or so-called 'influencers', particularly with respect to the element of group construction/enhancement/preservation just discussed in the current section.

5.4.4 Community building

We shall now address the 'construction of a cohesive community' ability of (polarised) users. As stated above in the previous sections, we confirm that the language is meant to be violent and exclusionary, often employs images with a double meaning at the boundary between literal and metaphorical and employ argumentative strategies such as those described in §5.2.2. However, here we prefer to draw attention towards the linguistic features that build in-group closeness and belonging.

We argue that the common language is formulaic towards the in-group: members construct a common vocabulary, not necessarily rich or with a clear-cut meaning, but recognisable as their own as part of their affiliation to the community. They repeatedly refer to it, dribbling the same images from one turn to the other by repeating, rephrasing or reframing them, with a more phatic than content-oriented intention. Reframing occurrences are particularly enticing for further investigation, since the matter discussed in the exchange might slowly drift to a different overall meaning from the one proposed in the beginning, even possibly with the same speaker disagreeing their previous claims, although no one participating in the discussion seems to acknowledge the fact or the issues correlated with it. In example (19), for instance, along the various exchanges user Savagfux first raises a conflict with the previous turn by user Interesting-Letter67 ('we are *already* in the midst of the "next war" [italics ours]'), although never attempting at actively resolving such a disagreement. Later on, however, the same user shifts their framing of such 'next war', both in terms of timing and content: they first state a 'next war' is already fought in physiological terms, referring to COVID-19, and by psychologically exploiting the fear deriving from it for mass control (from which other users take the lead for echoing agreement), but subsequently claim they believe a 'catastrophic conflict' of the traditional type, fought on a battlefield, is going to happen in the near future – not in contrast with what Interesting-Letter67 affirmed in the first place (boldface ours in the quoted turns).

(19-a) Interesting-Letter67: *It's a division piece, and very common. Make the vaccinated hate the un vaccinated. Divide and conquer. These news agencies* ***aren't going to last through the next war****.*
(19-b) Savagfux: ***We're already in the midst of the 'next war'****.. they're succeeding because everyone is waiting for a bomb to drop..* ***there's no need for bombs.. they're winning with physiology****.. it's crazy that no matter where I go there's 2 things that I always, 100% of the time, end up taking about.. politics and COVID. our heads are spinning.. we're at war friends and it's nothing like we've ever seen before.*
(19-c) LordOFtheNoldor: ***Silent weapons for quiet wars****.*
(19-d) Significant-Knee-150: *This is textbook psychological warfare. We are in a civil war, just not fought with kinetic force, yet.*
(19-e) Savagfux: *It may not end up in any physical / kinetic force locally.* ***I do believe that we will see catastrophic conflict between the US and China in the next 5 years****.*

We thus claim that the goal of building in-group cohesiveness and membership seems to take precedence over that of having a critical discussion in such exchanges.

Participants

* Liesbeth Allein
KU Leuven, BE
https://orcid.org/0000-0002-7776-2156

* Katarzyna Budzynska
Warsaw University of Technology, PL
https://orcid.org/0000-0001-9674-9902

* Giulia D'Agostino
Università della Svizzera italiana, CH
https://orcid.org/0009-0007-8918-1440

* Ewelina Gajewska
Warsaw University of Technology, PL
https://orcid.org/0009-0006-6012-4787

* Amalia Haro Marchal
University of Granada, ES
https://orcid.org/0000-0003-3232-976X

* Zlata Kikteva
University of Passau, DE
https://orcid.org/0009-0008-7549-6182

* Konrad Kiljan
University of Warsaw, PL
https://orcid.org/0000-0003-1088-683X

* Barbara Konat
Adam Mickiewicz University, PL
https://orcid.org/0000-0003-2370-4636

* Marcin Koszowy
Warsaw University of Technology, PL
https://orcid.org/0000-0001-5553-7428

* Alina Landowska
University of Social Sciences and Humanities, PL
https://orcid.org/0000-0002-7966-8243

* Maud Oostindie
Maastricht University, NL
https://orcid.org/0009-0007-0051-0480

* John Parkinson
Maastricht University, NL
https://orcid.org/0000-0002-7842-7739

* Martin Pereira-Farina
University of Santiago de Compostela, ES
https://orcid.org/0000-0002-1982-2472

* Jennifer Schumann
University of Fribourg, CH
https://orcid.org/0000-0003-0060-1212

* Narjes Sheikh Asadi
Università della Svizzera italiana, CH
https://orcid.org/0000-0002-6586-610X

* Yana Sviatsilnikava
Warsaw University of Technology, PL

* Maciej Uberna
Warsaw University of Technology, PL
https://orcid.org/0009-0006-8953-8270

* Olena Yaskorska-Shah
Università della Svizzera italiana, CH
https://orcid.org/0000-0003-4669-9462

* Ramy Younis
University of Fribourg, CH
https://orcid.org/0009-0006-7315-9006

* He Zhang
Warsaw University of Technology, PL
https://orcid.org/0009-0003-6610-2675

Acknowledgements The work reported in this report was supported in part by CHIST-ERA under grant 2022/04/Y/ST6/00001; in part by the Research Foundation - Flanders (FWO) under grant G0L0822N through the CHIST-ERA project; in part by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 860621; in part by VW foundation (*VolkswagenStiftung*) under grant 98 541 | 98 542 | 98 544; in part by the Polish National Science Centre (NCN) under grant 2020/39/I/HS1/02861; in part by the Swiss National Science Foundation (SNSF) Division I under grant 200857; in part by the research project PID2019-107478GB-I00 (FPI Predoctoral Fellow PRE2020-095944) of the Spanish Ministry of Science and Innovation; and in part by the Spanish National Research Agency (AEI) through the project PID2020-114758RB-I00 under grant MCIN/AEI/10.13039/501100011033.

Index of Terms

reinforcement Alternatively, 'rephrase to agree', is observed when all discussants support one standpoint with different premises. 16, 17

rephrase A communicative strategy that consists of modifying an original contribution in order to achieve a variety of persuasive goals. 6

robin-hooding A rhetorical strategy used to polarise society. Here, it is considered the act of both unyieldingly distancing poles on a vertical axis and dynamically rearranging poles - trying to subsume them all under a compassionate and truthful one - on a horizontal axis. 21

values The criteria people use to select and justify actions and to evaluate the self, the others and events. 7

References

1. Scott F. Aikin. Deep disagreement, the dark enlightenment, and the rhetoric of the red pill. *Journal of Applied Philosophy*, 36(3):420–435, 2019.
2. Manuel Almagro. Political polarization: Radicalism and immune beliefs. *Philosophy & Social Criticism*, 49(3):309–331, 2023.
3. Mona AlSheddi, Sophie Russell, and Peter Hegarty. How does culture shape our moral identity? moral foundations in saudi arabia and britain. *European Journal of Social Psychology*, 50(1):97–110, 2020.
4. Nourah Alswaidan and Mohamed El Bachir Menai. A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, 62(8):2937–2987, 2020.
5. Kwame Anthony Appiah. *The honor code: How moral revolutions happen*. WW Norton & Company, 2011.
6. Oscar Araque, Lorenzo Gatti, and Kyriaki Kalimeri. Libertymfd: A lexicon to assess the moral foundation of liberty. In *Proceedings of the 2022 ACM Conference on Information Technology for Social Good*, pages 154–160, 2022.
7. Aristotle. Rhetoric. Translated by W. R. Roberts. In D. Ross, W., editor, *The works of Aristotle (vol.11) Rhetorica, de rhetorica ad Alexandrum, poetica*. Oxford: Clarendon Press, 1952.
8. Grant M Armstrong and Julie Wronski. Framing hate: Moral foundations, party cues, and (in) tolerance of offensive speech. *Journal of Social and Political Psychology*, 7(2):695–725, 2019.
9. Mohammad Atari, Aida Mostafazadeh Davani, Drew Kogon, Brendan Kennedy, Nripsuta Ani Saxena, Ian Anderson, and Morteza Dehghani. Morally homogeneous networks and radicalism. *Social Psychological and Personality Science*, 13(6):999–1009, 2022.
10. Rodney Barker. *Making enemies*. Springer, 2006.
11. Michael D Barnett, Haluk CM Öz, and Arthur D Marsden. Economic and social political ideology and homophobia: The mediating role of binding and individualizing moral foundations. *Archives of sexual behavior*, 47:1183–1194, 2018.
12. Paul Bloom. How do morals change? *Nature*, 464(7288):490–490, 2010.
13. Thijs Bouman, Linda Steg, and Henk AL Kiers. Measuring values in environmental research: A test of an environmental portrait value questionnaire. *Frontiers in psychology*, 9:564, 2018.
14. William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318, 2017.
15. Marilynn B Brewer and Roderick M Kramer. The psychology of intergroup attitudes and behavior. *Annual review of psychology*, 36(1):219–243, 1985.
16. Kathryn Bruchmann and Liya LaPierre. Moral foundations predict perceptions of moral permissibility of covid-19 public health guideline violations in united states university students. *Frontiers in Psychology*, 12:795278, 2022.
17. J. Byford. *Conspiracy Theories: A Critical Introduction*. Palgrave Macmilan, 2011.
18. Scott Clifford and Jennifer Jerit. How words do the work of politics: Moral foundations theory and the debate over stem cell research. *The Journal of Politics*, 75(3):659–671, 2013.
19. Maria-Josep Cuenca. Two ways to reformulate: a contrastive analysis of reformulation markers. *Journal of Pragmatics*, 35(7):1069–1093, 2003.
20. Stuart Cunningham. 15. political and media leadership in the age of youtube. *Public Leadership*, page 177, 2008.
21. Oliver Scott Curry. Morality as cooperation: A problem-centred approach. *The evolution of morality*, pages 27–51, 2016.

22 Martin V Day, Susan T Fiske, Emily L Downing, and Thomas E Trail. Shifting liberal and conservative attitudes using moral foundations theory. *Personality and Social Psychology Bulletin*, 40(12):1559–1573, 2014.

23 Jeroen de Ridder. Deep disagreements and political polarisation. *Political epistemology*, 2021.

24 Thomas Dietz, Amy Fitzgerald, and Rachael Shwom. Environmental values. *Annu. Rev. Environ. Resour.*, 30:335–372, 2005.

25 William Donohue and Mark Hamilton. A framework for understanding polarizing language. In *The Routledge Handbook of Language and Persuasion*, pages 207–223. Routledge, 2022.

26 KM Douglas, JE Uscinski, RM Sutton, A Cichocka, T Nefes, CS Ang, and F Deravi. Understanding conspiracy theories. advances in political psychology, 40 (1), 1–33, 2019.

27 Maxim Egorov, Karianne Kalshoven, Armin Pircher Verdorfer, and Claudia Peus. It'sa match: Moralization and the effects of moral foundations congruence on ethical and unethical leadership perception. *Journal of Business Ethics*, 167:707–723, 2020.

28 Matthew Feinberg and Robb Willer. Moral reframing: A technique for effective and persuasive communication across political divides. *Social and Personality Psychology Compass*, 13(12):e12501, 2019.

29 James Paul Gee. *An introduction to discourse analysis: theory and method*. Routledge, New York, 3rd ed edition, 2011.

30 Matthew Gentzkow. Polarization in 2016. *Toulouse Network for Information Technology Whitepaper*, 1, 2016.

31 Bart Geurts. Communication as commitment sharing: speech acts, implicatures, common ground. *Theoretical Linguistics*, 45(1-2):1–30, 2019.

32 Erving Goffman. *Forms of Talk*. Conduct and Communication. University of Pennsylvania Press, Incorporated, 1981.

33 Cesar Gonzalez-Perez. *Information Modelling for Archaeology and Anthropology : Software Engineering Principles for Cultural Heritage*. Springer International Publishing, 2018.

34 Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik, and Peter H Ditto. Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*, volume 47, pages 55–130. Elsevier, 2013.

35 Jesse Graham, Brian A Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H Ditto. Mapping the moral domain. *Journal of personality and social psychology*, 101(2):366, 2011.

36 Alan G Gross. Rhetoric, narrative, and the lifeworld: The construction of collective identity. *Philosophy & rhetoric*, 43(2):118–138, 2010.

37 Jürgen Habermas. Political communication in media society: Does democracy still enjoy an epistemic dimension? the impact of normative theory on empirical research. *Communication theory*, 16(4):411–426, 2006.

38 Lindsay Hahn, Ron Tamborini, Eric Novotny, Clare Grall, and Brian Klebig. Applying moral foundations theory to identify terrorist group motivations. *Political Psychology*, 40(3):507–522, 2019.

39 Jonathan Haidt. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review*, 108(4):814, 2001.

40 Jonathan Haidt. The new synthesis in moral psychology. *science*, 316(5827):998–1002, 2007.

41 Jonathan Haidt. *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.

42 Jonathan Haidt and Jesse Graham. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social justice research*, 20(1):98–116, 2007.

43 Jonathan Haidt and Craig Joseph. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4):55–66, 2004.

44 Maarten A Hajer. *Authoritative governance: Policy making in the age of mediatization*. Oxford University Press, 2009.

45 Joe Hoover, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, et al. Moral foundations twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071, 2020.

46 Frederic R Hopp, Jacob T Fisher, Devin Cornell, Richard Huskey, and René Weber. The extended moral foundations dictionary (emfd): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior research methods*, 53:232–246, 2021.

47 Christian Bøtcher Jacobsen and Lotte Bøgh Andersen. Is leadership in the eye of the beholder? a study of intended and perceived leadership practices and organizational performance. *Public administration review*, 75(6):829–841, 2015.

48 Kyriaki Kalimeri, Mariano G. Beiró, Alessandra Urbinati, Andrea Bonanomi, Alessandro Rosina, and Ciro Cattuto. Human values and attitudes towards vaccination in social media. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 248–254, 2019.

49 Klemens Kappel. Higher order evidence and deep disagreement. *Topoi*, 40(5):1039–1050, 2021.

50 Spassena P Koleva, Jesse Graham, Ravi Iyer, Peter H Ditto, and Jonathan Haidt. Tracing the threads: How five moral concerns (especially purity) help explain culture war attitudes. *Journal of research in personality*, 46(2):184–194, 2012.

51 Barbara Konat, Katarzyna Budzynska, and Patrick Saint-Dizier. Rephrase in argument structure. In *Foundations of the Language of Argumentation: COMMA 2016 Workshop*, pages 32–39, 2016.

52 Marcin Koszowy. *Autorytet w argumentacji i w dialogu. Teorie - modele - aplikacje (Authority in argumentation and in dialogue. Theories - models - applications)*. Białystok: Wydawnictwo UwB, 2019.

53 Bernard Lo and Lindsay Parham. Ethical issues in stem cell research. *Endocrine reviews*, 30(3):204–213, 2009.

54 Christopher Lockhart, Carol HJ Lee, Chris G Sibley, and Danny Osborne. The sanctity of life: The role of purity in attitudes towards abortion and euthanasia. *International Journal of Psychology*, 58(1):16–29, 2023.

55 Andrew Luttrell, Aviva Philipp-Muller, and Richard E Petty. Challenging moral attitudes with moral messages. *Psychological Science*, 30(8):1136–1150, 2019.

56 Nadine A Marshall, Lauric Thiault, Ally Beeden, Roger Beeden, Claudia Benham, Matt I Curnock, Amy Diedrich, Georgina G Gurney, Lindsey Jones, Paul A Marshall, et al. Our environmental value orientations influence how we respond to climate change. *Frontiers in psychology*, 10:938, 2019.

57 Farzana Masroor, Qintarah N Khan, Iman Aib, and Zulfiqar Ali. Polarization and ideological weaving in twitter discourse of politicians. *Social media+ society*, 5(4):2056305119891220, 2019.

58 Frederick W Mayer. *Narrative politics: Stories and collective action*. Oxford University Press, USA, 2014.

59 Logan Molyneux and Shannon C McGregor. Legitimating a platform: Evidence of journalists' role in transferring authority to twitter. *Information, Communication & Society*, 25(11):1577–1595, 2022.

60 Marlon Mooijman, Joe Hoover, Ying Lin, Heng Ji, and Morteza Dehghani. Moralization in social networks and the emergence of violence during protests. *Nature human behaviour*, 2(6):389–396, 2018.

61 G Scott Morgan, Linda J Skitka, and Daniel C Wisneski. Moral and religious convictions and intentions to vote in the 2008 presidential election. *Analyses of Social Issues and Public Policy*, 10(1):307–320, 2010.

62 Elena Musi, Debanjan Ghosh, and Smaranda Muresan. Changemyview through concessions: Do concessions increase persuasion? *Dialogue & Discourse*, 9(1):107–127, 2018.

63 NPC. In *Cambridge Advanced Learner's Dictionary and Thesaurus*. Cambridge University Press, (n.d.).

64 Ozge Ozduzen and Umut Korkut. Enmeshing the mundane and the political: Twitter, lgbti+ outing and macro-political polarisation in turkey. *Contemporary Politics*, 26(5):493–511, 2020.

65 Roosmaryn Pilgram and Lotte van Poppel. The strategic use of metaphor in argumentation. *The Language of Argumentation*, pages 191–212, 2021.

66 Lotte van Poppel. The study of metaphor in argumentation theory. In *Argumentation Through Languages and Cultures*, pages 177–208. Springer, 2020.

67 Eddo Rigotti and Sara Greco. *Inference in Argumentation: A Topics-Based Approach to Argument Schemes*. Springer, 2019.

68 Fogelin Robert. The logic of deep disagreements. *Informal Logic*, 25(1):3–11, 2005.

69 Eyal Sagi and Morteza Dehghani. Measuring moral rhetoric in text. *Social science computer review*, 32(2):132–144, 2014.

70 E Said. Orientalism pantheon books. *New York*, 1978.

71 Kirill Solovev and Nicolas Pröllochs. Moralized language predicts hate speech on social media. *PNAS nexus*, 2(1):pgac281, 2023.

72 Marco Stranisci, Michele De Leonardis, Cristina Bosco, and Viviana Patti. The expression of moral values in the twitter debate: a corpus of conversations. *IJCoL. Italian Journal of Computational Linguistics*, 7(7-1, 2):113–132, 2021.

73 Pontus Strimling, Irina Vartanova, and Kimmo Eriksson. Predicting how us public opinion on moral issues will change from 2018 to 2020 and beyond. *Royal Society Open Science*, 9(4):211068, 2022.

74 Pontus Strimling, Irina Vartanova, Fredrik Jansson, and Kimmo Eriksson. The connection between moral positions and moral arguments drives opinion change. *Nature Human Behaviour*, 3(9):922–930, 2019.

75 Joseph M Stubbersfield, Lewis G Dean, Sana Sheikh, Kevin N Laland, and Catharine P Cross. Social transmission favours the 'morally good'over the 'merely arousing'. *Palgrave Communications*, 5(1), 2019.

76 Sanaz Talaifar and William B Swann Jr. Deep alignment with country shrinks the moral gap between conservatives and liberals. *Political Psychology*, 40(3):657–675, 2019.

77 Hammond Tarry, Valérie Vézina, Jacob Bailey, and Leah Lopes. Political orientation, moral foundations, and covid-19 social distancing. *PloS one*, 17(6):e0267136, 2022.

78 Yla R Tausczik and James W Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.

79 Kathie Treen, Hywel Williams, Saffron O'Neill, and Travis G Coan. Discussion of climate change on reddit: Polarized discourse or deliberative debate? *Environmental Communication*, 16(5):680–698, 2022.

80 Frans H Van Eemeren, Rob Grootendorst, and A Francisca Sn Henkemans. *Argumentation: Analysis, evaluation, presentation*. Routledge, 2002.

81 Tyler J Vaughan, Lisa Bell Holleran, and Jason R Silver. Applying moral foundations theory to the explanation of capital jurors' sentencing decisions. *Justice Quarterly*, 36(7):1176–1205, 2019.

82 Annemarie S Walter and David P Redlawsk. The effects of politician's moral violations on voters' moral emotions. *Political Behavior*, pages 1–27, 2021.

83 Douglas Walton. *Scare tactics: Arguments that appeal to fear and threats*, volume 3. Springer Science & Business Media, 2013.

84 Małgorzata Wierzba, Monika Riegel, Jan Kocoń, Piotr Miłkowski, Arkadiusz Janz, Katarzyna Klessa, Konrad Juszczyk, Barbara Konat, Damian Grimling, Maciej Piasecki, et al. Emotion norms for 6000 polish word meanings with a direct mapping to the polish wordnet. *Behavior Research Methods*, pages 1–16, 2021.

85 Jing Yi Xie, Renato Ferreira Pinto Junior, Graeme Hirst, and Yang Xu. Text-based inference of moral sentiment change. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4654–4663, Hong Kong, China, November 2019. Association for Computational Linguistics.

86 Sijia Yang. *Morality in Tobacco Control Messaging: Effects of Moral Appeals on Persuasion and Retransmission*. PhD thesis, University of Pennsylvania, 2019.

87 R. Younis, D. de Oliveira Fernandes, P. Gygax, M. Koszowy, and S. Oswald. Rephrasing is not arguing, but it is still persuasive: An experimental approach to perlocutionary effects of rephrase. *Journal of Pragmatics*, 210:12–23, 2023.