Research Article

Woke/Anti-Woke dynamics in Norway, 2014-2023

Gabe M Maayan[1,2], Kristen Bestavros[1], Wesley J Wildman[1,2]

Abstract

We present a system dynamics model (SDM) of woke/anti-woke dynamics in Norway, calibrated against Norwegian Citizen Panel data from 2014 to 2023. The SDM synthesizes theories related to woke/anti-woke dynamics, including Moral Foundations Theory and two contrasting types of sensitivity to injustice (focusing on injustice to self and injustice to others). The simulation's two dependent variables measure the total amount of radicalism (left-wing and right-wing combined) relative to the population, and the amount of left-wing radicalism versus right-wing radicalism. Minimizing ideological polarization calls for minimizing the first output measure, while winning ideological culture wars draws attention to the second output measure. Results exhibit the vital role that homophily – the extent to which people socialize and communicate with those ideologically similar to themselves – plays in exacerbating or ameliorating ideological polarization in a society. High homophily increases radicalism in a relatively balanced way. Unbalanced radicalism can occur, but only at low levels of homophily. This suggests practical interventions: intensify culture wars by increasing homophily through media bubbles and decreasing encounters with unfamiliar others; or reduce polarization by decreasing homophily, encouraging positive

[1] Faculty of Computing and Data Sciences, Boston University, United States of America
[2] Just Horizons Alliance, United States of America
Corresponding Author:
Gabe M Maayan, Just Horizons Alliance, United States of America. Email: gmaayan@mindandculture.org

encounters with unfamiliar others, and resisting one-sided radicalism when it arises by emphasizing civil virtues.

Keywords

System-Dynamics model, ideology, polarization, political dynamics, moral foundations theory, Norway

Introduction

Political polarization has become a growing concern in democracies around the world (Gu & Wang, 2022; Schedler, 2023). Alongside this trend, the use of the terms "woke" and "anti-woke" has risen, especially in the U.S. and Europe, and mirrors political radicalism with some distinct differences (Smith et al., 2023). Despite its prominence in political discourse, "wokeism" remains loosely defined, rarely claimed as self-identity, and is often used as part of a polarization strategy by pundits to create 'pernicious polarization,' a mutually distrustful "Us vs. Them" society (McCoy & Somer, 2019; VanDreew et al., 2025).

One distinctive feature of Woke radicalism, compared to other forms of political radicalism, is the strong disgust response it elicits from the Left. Yet, as a recent cultural phenomenon, wokeism has received limited scholarly attention. Existing studies fall short of explaining why Woke/Anti-Woke polarization arises, and why do we see the Left's strong reaction to radical wokeism, moreso than other forms of radicalism. In this study, we propose a theory of Woke/Anti-Woke polarization based on Moral Foundations Theory to answer these questions. To validate the cohesion of our theory, we built a computational model of the theory and demonstrate that it can replicate empirical population ideological

distributions in Norway from three points in time. These empirical ideology distributions were created from Norwegian Citizen Panel data.

There is a long history of studying opinion dynamics and polarization using computational modeling (Axelrod, 1997; Bramson et al., 2017; Centola et al., 2007; G. Deffuant, 2006; G. A. Deffuant, 2002; Flache & Macy, 2006; Hegselmann & Krause, n.d.; Kitts, 2006; Klemm et al., 2005). These studies model a variety of theories and mechanisms, including homophily, network structure, etc. Outside of computational modeling, there is a vast literature studying polarization (Cole et al., 2025; Kołczyńska, 2025; Kubin & von Sikorski, 2021; Marino & Iannelli, 2023; Prior, 2013; Schedler, 2023). Scholars define two distinct forms of polarization. Ideological polarization refers to increasing distance in policy positions between individuals or groups (Dalton, 1987). In contrast, affective polarization refers to increasing emotional or social distance (Leininger et al., 2023; Mason, 2022). Affective polarization leads to stronger identity fusion with political parties, distrust and dislike of the 'other side,' and is closer to the polarization observed with Woke/Anti-Woke radicalism than ideological polarization.

Many forms of polarization are inherently asymmetric, and we observe this with Woke/Anti-Woke polarization too. On the Left, wokeism and woke terminology are rarely used in politics. Instead, the beliefs labeled as 'woke' are generally aligned with traditional liberal dogma and so are not separated out (Smith et al., 2023; VanDreew et al., 2025). Radical wokeism exhibits many of the behaviors of left-wing authoritarianism (LWA) such as desire for social uniformity, concern for hierarchy and moral absolutism, but is still much more of a social movement than a political one (Conway III et al., 2023; Conway et al., 2018;

Costello et al., 2022). In contrast, (anti)woke terminology is used much more frequently in politics on the Right as a tool to increase outgroup hostility. Right-wing authoritarianism (RWA) has been increasing around the world over the past few decades and has contributed to the parallel increase in polarization (Farkhari et al., 2025; Osborne et al., 2023; Salajan & jules, 2024; Vilanova et al., 2023). In this time, the contemporary far-right have gained large amounts of political control and become central actors in mainstream debates (Pirro, 2024; Völker & Saldivia Gonzatti, 2024). As with LWA and radical wokeism, radical anti-wokeism is one facet of RWA. Paradoxically, anti-woke pundits characterize themselves as "fair-minded occupants of a political middle ground," and use arguments based on fairness and colorblindness to oppose wokeism; attributes that are typically associated with liberal arguments (Asen, 2024).

Thus, the paradox of Woke/Anti-Woke radicalism finds Woke radicals moving towards traditionally anti-liberal, authoritarianist values, which creates a strong disgust reaction on the Left, and Anti-Woke radicals using traditionally liberal-value arguments in opposition. The theory outlined below attempts to reconcile this paradox using the findings from Moral Foundations Theory. Our computational model, built using this theory, can replicate real-world ideological distributions, which suggests that the theory is coherent and useful to guide future research. The model presented in this study is not complex enough to investigate detailed policy solutions but could be extended in future research to look at interventions. Regardless, we found the drivers of model behavior to be homophily and different types of justice frustration, which is in line with previous literature on polarization.

In the next sections, we will discuss the history of wokeism in the U.S. and Europe, the base components of our theory, our computational model and empirical data, before finishing with our results and discussion.

Context

In an increasingly polarized society, the word 'woke' has become a politically charged term. It is used as a catch-all descriptor of leftist political ideology, frequently deployed disparagingly by the right to undermine social justice efforts that are deemed too extreme (Murray, 2023; Romano, 2020). Despite how frequently the term has featured in the media over the past ten years, most academic scholarship on wokeness is quite recent. Indeed, the vast majority of academic articles on the topic of wokeness were published following the 2020 killing of George Floyd, a black man, by white police officer Derek Chauvin, in the United States. Floyd's murder prompted the resurgence of the Black Lives Matter movement and led to worldwide protests against racism and police brutality (Cineas, 2021; McWhorter, 2022). This shift marks a pivotal point in the discourse, since much of the existing research on wokeness in its current form has been shaped by the social and political movements that followed. Understanding the evolution of woke ideology, as well as its critiques, is vital for grasping the way wokeness and anti-wokeness replicates in other national settings, including Norway, the focus of this paper.

The first recorded usage of 'woke' as a political term is often attributed to Marcus Garvey, a prolific Jamaican political activist. Garvey believed in the liberation of black people worldwide and declared "Wake up Ethiopia! Wake up Africa!" in a 1923 compilation of his philosophical ideas. The language of Garvey's call to action was echoed in the spoken

outro of Lead Belly's song, "Scottsboro Boys," which urges black people to "stay woke, keep

their eyes open" to the realities of racism and police brutality (Murray, 2023; Romano,

2020). The term "stay woke" became a part of black vernacular, and was discussed in a

1962 *New York Times* article by William Melvin Kelley titled "If You're Woke You Dig It." In the

article, Kelley wrote about the evolution of black slang and its appropriation by white

people. Kelley defined "woke" as an adjective meaning "well-informed, up-to-date."

Contrary to how Marcus Garvey and Lead Belly used the term, Kelley's definition of "woke"

makes no explicit connection to racism or social discrimination; however, he implies that

being woke requires having a strong connection with black culture and knowledge of issues

that specifically affect the black community (Mills, 2020; Murray, 2023). Kelley's article is

often credited as the first modern citation of the term "woke," which is currently defined by

the Oxford English Dictionary as "alert to racial or social discrimination and injustice."

While the term "woke" has long been a part of black vernacular, it did not become truly

mainstream until the late 2000s and early 2010s. Its first mainstream appearance was in

Erykah Badu's 2008 song, "Master Teacher," which utilized the phrase "I stay woke" several

times in its chorus (Murray, 2023; Romano, 2020). The phrase "stay woke" gained traction

on social media, especially among black users. Six years later, after the murder of Michael

Brown by police, the phrase "stay woke" became a call to action, used by racial justice

activists to encourage protests against police brutality and racism. It was at this point that

the term "woke" started to be used outside the context of racial justice, losing some of its

original meaning. It was used by feminist protesters at the 2017 Women's March to

challenge sexism and misogyny, and it would later be used by the LGBTQ+ community to

champion gay rights (Murray, 2023; Romano, 2020). Wokeism became an ideology built around social justice in all its forms, a popular and appealing concept to progressives, including many liberals in the United States. As wokeism became more popular on the Left, so too did anti-wokeism on the Right.

The broad scope of wokeness in social justice contexts left the definition of the term fuzzy, making it the perfect term for conservatives to use disparagingly as a catch-all for leftist political ideology that is deemed unreasonable or extreme. Florida governor Ron DeSantis has referred to wokeism as "a form of cultural Marxism" which is "basically a war on the truth" (Massie & Sheets, 2023; Murray, 2023). As a result of its increasingly negative connotation, few people self-identify as woke or use the term unironically, even if their beliefs align with those often attributed to wokeism.

The murder of George Floyd prompted the resurgence of wokeness in the context of antiracism. While the immediate aftermath saw worldwide protests and calls for police reform, there was also a significant demand for antiracist literature and research. During this time, sales of civil rights titles skyrocketed by 330% (Cineas, 2021). The anti-racist book club became a common phenomenon across the country, with *White Fragility* by Robin D'Angelo, *Between the World and Me* by Ta-Nehisi Coates, and *How to be an Antiracist* by Ibram X. Kendi being some of the most popular readings (Cineas, 2021; Sanneh, 2019). The increase in calls for social change also precipitated a wave of Conservative pundits decrying the dangers of racial activism and wokeness, framing both as existential threats to American culture. Evidence of these claims often takes the form of trending internet slogans; Defund the Police and ACAB (All Cops Are Bastards), both

associated with the Black Lives Matter (BLM) movement, were particularly incendiary to conservatives (Goldberg, 2024; Lind, 2017). Indeed, BLM has frequently been targeted by anti-woke conservatives, with critics of the social justice movement using the slogan All Lives Matter as a retort to Black Lives Matter. Anti-woke conservatives have also resisted demands for police reform, with slogans such as 'Back the Blue' gaining popularity, and a blue version of the American flag appearing on lawns and windows as a sign of solidarity with police – or, to critics, a symbol of white supremacy (Williams, 2020). Responding to the perceived threat posed by wokeism, conservatives have called for the elimination of affirmative action, diversity, equity, and inclusion (DEI) initiatives, and education on topics in social justice, most notably critical race theory (Goldberg, 2024; Lind, 2017). Some of these initiatives have already succeeded – the Supreme Court has ruled that affirmative action is unconstitutional, and many Republican-controlled states have enacted laws restricting what subjects can be taught in K-12 schools (Sherman 2023). More recently, the Trump administration has demanded that higher education institutions terminate their DEI programs or risk losing federal funding, marking a significant escalation in conservative efforts against social justice initiatives (Binkley, 2025).

The majority of discourse surrounding wokeism and anti-wokeism is America-centric; however, in recent years, there have been concerns that both wokeism and anti-wokeism are "spreading" to other countries. In Europe, politicians in France and the U.K. have lamented the rise of wokeness within their borders, claiming it is dangerous and sows division by inventing problems where none exist. Wokeism is particularly incendiary in France, where both progressives and conservatives agree that it is anathema to French

culture and creates more problems than it solves (Caulcutt, 2024). French leaders, including President Emmanuel Macron, have expressed concerns that wokeism is "fragmenting society." Macron has warned that an overemphasis on identity politics undermines France's universalist values by categorizing citizens based on their differences rather than their shared citizenship (Van Zandt, 2021). These sentiments have been echoed by French intellectuals and academics on both sides of the political spectrum, who argue that wokeism encourages a victimhood mentality that they see as foreign to French republican ideals (Caulcutt, 2024; Van Zandt, 2021; Vinocur, 2024). This anti-woke mentality is shared by most French people – according to a poll by specialist Chloe Morín, only 4% of French people support wokeness. Despite this, French people exhibit comparatively high agreement with certain aspects of wokeness – for example, 41% of French people believe racism affects every level of society (Caulcutt, 2024).

Norwegian context. Not all European countries see wokeism as a dangerous American phenomenon, however. For example, the Norwegian culture is generally supportive of addressing systemic racism, promoting gender equality, and increasing LGBTQ+ rights, goals that align with wokeism. These ideas are often seen as a continuation of Norway's commitment to egalitarianism and social justice. In contrast, anti-woke sentiment is viewed with suspicion, particularly because of its association with far-right extremism in the wake of Anders Breivik's 2011 terrorist attacks in Oslo and Utøya, which killed 77 people. Breivik's massacre, driven by extreme anti-immigration and anti-Muslim ideology, was a direct reaction to what he perceived as the erosion of Norwegian and European identity due to multiculturalism and immigration (Rahman et al., 2016; Richards, 2014). Many Norwegians viewed his ideology as a direct result of anti-immigration rhetoric, with some blaming American conservative media for influencing his beliefs (Carle, 2013; Osler & Lybaek, 2014). While the term "woke" was not in use at the time, it is natural to identify Breivik's reactionary right-wing political beliefs as anti-woke.

Norway's response to Breivik's ideology was a wholesale rejection of its extremist narrative, but it also sparked a broader conversation about the influence of anti-woke rhetoric. Norwegian intellectuals and politicians warned that adopting anti-woke positions, especially those focused on fears of multiculturalism and the "decline of Western civilization," risked legitimizing the same kind of toxic discourse that motivated Breivik (Carle, 2013; Osler & Lybaek, 2014). Nonetheless, anti-woke sentiment in Norway has not disappeared. While mainstream political parties condemn Breivik's extremism, concerns about immigration, cultural preservation, and the influence of identity politics persist. The

Progress Party (*Fremskrittspartiet or FrP*), for example, continues to promote stricter immigration controls and a more nationalist agenda, while avoiding the overtly violent rhetoric associated with Breivik (Miller & Listhaug, 1990; Osler & Lybaek, 2014).

In this study, we use data from the Norwegian Citizen Panel of 2014, 2017 and 2023 to calibrate the Woke/Anti-Woke (WAW) model, so it is helpful to understand the political landscape at those times. From 2014-2017, Norway had a conservative government led by Erna Solberg, with the centrist FrP in league with the Conservative Party of Norway (Høyre). FrP is known for criticizing political correctness, especially on issues related to immigration, integration, and freedom of speech. During this period, there was clear opposition to what FrP described as an "immigration-liberal elite" and a kind of political correctness that they believed threatened Norwegian culture. Issues such as the refugee crisis in 2015 and terror threats fueled increased debates about national security, immigration policies, and integration.

From 2017 to 2021, Erna Solberg's government continued, but with growing internal tensions, particularly related to FrP's desire for a tougher line on immigration and political correctness. FrP withdrew from the government in 2020, partly due to dissatisfaction with Høyre's moderate approach. Around 2019-2020, debates about identity politics and "woke" culture gained more attention in Norway, particularly in light of international movements such as Black Lives Matter (BLM) and #MeToo. Discussions about racism, gender identity, and sexual orientation became more prominent, and social media hosted countless debates about what was "acceptable" to say publicly.

After the 2021 election, the previous coalition ended, and the Labour Party (Ap) and the

Centre Party (Sp) formed a new government, led by Jonas Gahr Støre. This government

emphasized social justice, labor issues, and climate action, but has not actively marked

itself as for or against so-called "woke" policies to the same extent as some other parties.

Debates about political correctness, freedom of speech, and identity politics continue to

shape public discourse. The Progress Party, now in opposition, has continued to criticize

what it sees as an excessive focus on gender, identity, and "woke" culture, particularly

regarding issues such as gender-neutral language and measures to promote diversity. As of

2024, the Labour Party signaled a tougher line against immigration, including not accepting

all Ukrainian refugees. The economic situation, a weak currency, and high cost of living

seems to have impacted the Labour Party's approach.

Theory

That different countries and cultures have different responses to wokeism and anti-

wokeism is unsurprising; however, understanding why this is the case is an important

challenge for our model's conceptual framework. One useful lens through which to analyze

these divergent responses is moral foundations theory (MFT), which suggests that

individuals base their moral judgments on a set of universal moral foundations. MFT

identifies five core moral foundations—care/harm, fairness/cheating, ingroup/outgroup,

authority/subversion, and sanctity/degradation—that people prioritize to different degrees

depending on their cultural background and political orientation (Day et al., 2014; Graham

et al., 2009; Kivikangas et al., 2021). MFT is especially useful as a foundational theory

because it has been studied widely across cultures and domains, both within and outside

of political science (Dickinson et al., 2016; Rempala et al., 2016; Vaughan et al., 2019;

Zangari et al., 2025). The first two foundations defined by MFT, care/harm and

fairness/cheating, are often associated with concerns about protecting individuals and

ensuring justice, and we will refer to them as "wellbeing" foundations. The remaining three

foundations—ingroup/outgroup, authority/subversion, and sanctity/degradation—tend to

focus more on maintaining order, group cohesion, and tradition. We call these "binding"

foundations.

Analyzing wokeism and anti-wokeism through the lens of MFT reveals patterns in how

different moral foundations drive support for or resistance to these movements. Liberals,

according to research by Graham, Haidt, and Nosek, tend to emphasize the wellbeing

foundations while de-emphasizing binding foundations (Graham et al., 2009). For liberals,

seeking justice and ensuring fairness for marginalized groups are of utmost importance.

Issues such as racism, gender inequality, and LGBTQ+ rights align closely with the liberal

prioritization of reducing harm and promoting fairness. It is important to note that the

liberal de-emphasis of binding foundations is complete and deliberate; liberalism and

wokeism actively argue against ingroup/outgroup dynamics in favor of inclusivity, and

similarly lament the conservative emphasis on purity, which has its roots in religion and is

seen as harmful through amplifying ingroup/outgroup animosity and hierarchical

authoritarianism (Day et al., 2014; Graham et al., 2009).

Conservatives, on the other hand, emphasize all five moral foundations relatively evenly.

According to Graham, Haidt, and Nosek, conservatives place greater importance on

maintaining social order, loyalty to one's group, and respect for authority and tradition – in other words, binding foundations (Day et al., 2014; Graham et al., 2009). While conservatives also place importance on care/harm and fairness/cheating, the emphasis on wellbeing is expressed differently than liberals. For example, affirmative action is seen by liberals as fair, since it compensates for systemic inequities; while conservatives object to affirmative action because they see it as *un*fair, giving preference to individuals from certain groups over individuals with more demonstrated merit. The broader moral framework exhibited by conservatives makes them more resistant to woke movements, which are often seen as undermining societal stability and challenging traditional hierarchies. Anti-woke rhetoric frequently highlights concerns that woke ideologies threaten national identity or degrade cultural values, aligning with the ingroup/outgroup and sanctity/degradation foundations.

Importantly, the extreme ends of the political spectrum do not necessarily follow the same moral foundations as their moderate neighbors. Radicalism, whether on the far-left or far-right, tends to emphasize ingroup/outgroup dynamics more intensely than moderates, often leading to a heightened sense of threat from those perceived as "other" (Baldassarri & Bearman, 2007; Steel, 2023). On the far-right (radical anti-woke), the ingroup/outgroup dynamic often focuses on nationalism, race, and cultural purity, with a clear delineation between the in-group (nation, race, or tradition) and the out-group (immigrants, minorities, or political opponents). At the same time, their public rhetoric often eschews potent ingroup language and employs universally accessible rational language and categories that people across the political spectrum can easily understand. On the far-left (radical woke),

while race and nationality might be de-emphasized, ingroup-outgroup dynamics appear in the form of loyalty to ideological purity and opposition to those perceived as oppressive or complicit in systems of inequality. Radical woke rhetoric often employs intricate ingroup terms and categories that make sense to the initiated but confound the wider public's understanding. This extreme emphasis on following the 'correct' behavior is associated with the sanctity/degradation moral foundation. Since the very foundation of liberalism rejects such ingroups and outgroups, the radical woke embrace of binding foundations with associated language and power strategies often induces a disgust response, even among more moderate progressives and liberals. Thus, in the Woke/Anti-Woke system, there are three forms of disgust or repulsion. The first is ideological disgust felt between groups at the opposite ends of the spectrum, which drives ideological polarization (Dalton, 1987). The second is in-group/out-group repulsion, again between groups on opposite sides but driven by affective polarization instead (Leininger et al., 2023; Mason, 2022). Finally, the relatively unique form of disgust between those on the Left and the radical Woke due to a different emphasis on moral foundations than expected by an insider group.

Another important factor that influences wokeism and anti-wokeism is justice sensitivity. Tobias Rothmund, Laurits Bromme, and Flavio Azevedo's2020 study, *Justice for the People? How Justice Sensitivity Can Foster and Impair Support for Populist Radical-Right Parties and Politicians in the United States and in Germany*, explores how individuals' justice sensitivity can either promote or deter support for populist radical-right parties. The researchers categorize justice sensitivity into two main types: self-oriented justice sensitivity, which focuses on perceived injustices toward oneself, and other-oriented

justice sensitivity, which centers on empathic awareness of injustices faced by others. Self-oriented justice sensitivity often drives support for anti-woke ideologies, particularly among those who feel that their social or economic standing is under threat. Individuals with high self-oriented sensitivity may perceive themselves as victims of progressive social changes promoted by woke movements, such as shifts in power dynamics, affirmative action, or policies aimed at addressing systemic inequities. In the study, Rothmund, Bromme, and Azevedo found that people who feel victimized by these changes often gravitate toward populist radical-right parties that promise to protect traditional values and resist what they see as unfair redistributions of power. On the other hand, other-oriented justice sensitivity is more likely to foster support for wokeism. Individuals with high levels of other-oriented sensitivity are attuned to the injustices faced by marginalized or disadvantaged groups and are therefore more likely to support policies and movements that seek to address historical and systemic wrongs. According to the study, those with high other-oriented justice sensitivity are drawn to progressive platforms that promote racial, gender, and economic equality (Rothmund et al., 2020).

The Woke/Anti-Woke (WAW) model we present here focuses on Norway, where data for longitudinal calibration and evaluation of the model exist. We believe the theory sketched in the foregoing applies to many western nations where Wokeism is finding a cultural foothold.

Methods

System Dynamics modeling

System dynamics modeling (SDM) is a form of computational model used for understanding the behavior of complex, dynamic systems (Forrester, 1961; Sterman, 2002). In an SDM, a system is defined using a set of differential equations that mathematically formalize high-level theory. The system can then be simulated forward in time to expose the results of the model formalization. These models can be used for a wide variety of applications, from policy analysis and climate change predictions to assembly-line optimization and electric grid stress-testing (Egerer et al., 2021; Shults et al., 2018; Xu et al., 2018; Yang et al., 2013). In the case of this study, we use SDM to demonstrate the coherence of the theory of Woke/Anti-Woke polarization outlined above by showing that a simulation of this theory can recreate the empirical ideological population distribution in Norway at three different points in time. Though relatively rare, SDMs have been used to study various forms of polarization (Bertotti & Menale, 2024; Cano Macias & Ruiz Vera, 2024; Leonard et al., 2021; Shin et al., 2025). Other studies use agent-based modeling (ABM) to study similar phenomena (Baldassarri & Bearman, 2007; Banisch & Olbrich, 2021; Carpentras et al., 2023; Haque et al., 2023; Tokita et al., 2021). ABM is another form of computational modeling that uses a 'bottom up' approach where the modeler defines the system through simple agent interaction rules. In comparison, SDMs use a 'top down' approach where the system is defined by high level theory and are often simpler and more explainable than agent-based models. While this study uses an SDM is used to

demonstrate the coherence of our theory presented above, a future study could use an

agent-based model based on the same theory to investigate direct policy interventions. In

the model, a population is represented as groups of people belonging to different

ideologies. The core elements of an SDM are stocks, which represent quantities of

elements within the system, and flows, that define rates of change in quantities moving

between the stocks. In the WAW model, stocks represent quantities of people in different

ideological states, and the flows represent people moving into a different state.
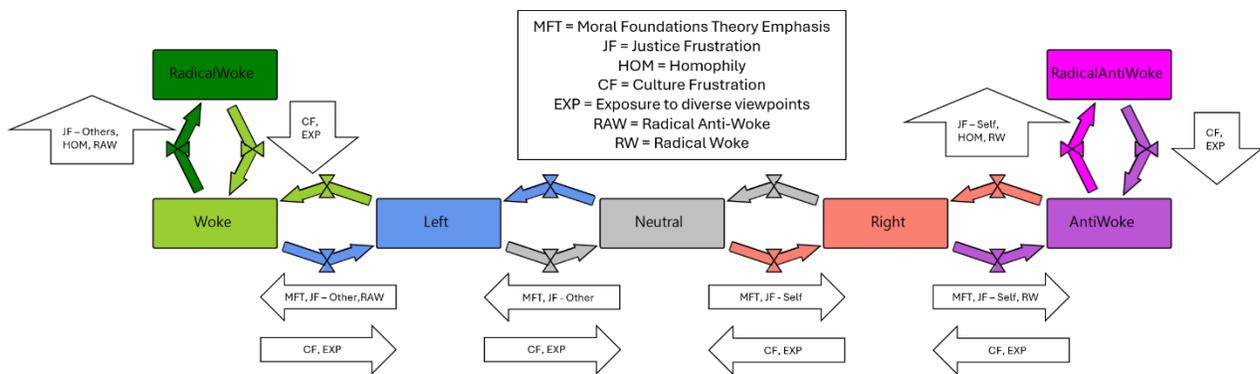
Woke Anti-Woke model description

Figure 1. *Stock-and-Flow diagram for the WAW Model.*

Figure 1 presents our theoretical hypothesis about how people change ideological states in

ways relevant to woke/anti-woke dynamics. This hypothesis is expressed in the WAW

Model. For a detailed description, see the overview, design, and details specification

(Grimm et al., 2020) in the supplemental materials.

Each stock is defined with a differential equation. For example, N, L and R are the values of

the Neutral, Left and Right stocks respectively, which are functions of time. f_c is frustration

with extremist culture, and i_l and i_r are interactions with others of differing beliefs on the

left and right respectively, m is the ratio of cultural emphasis on individualizing vs. binding foundations, and j_l and j_r are the strength of the left and right's justice sensitivities, as defined by Rothmund et al. The Neutral stock is thus defined as:

$$\frac{dN}{dt} = \big(L * avg(f_c, i_l)\big) + \big(R * avg(f_c, i_r)\big) - \big(N * avg(m, j_l)\big) - \big(N * avg(m, j_r)\big) \qquad (1)$$

The first term is the flow from Left to Neutral and represents the assertion that people on the Left move in the Neutral direction because of frustration with extremist culture and interactions with people on the other side of the political spectrum. The second term is similarly defined for the flow from Right to Neutral. The third and fourth terms are the flows from Neutral to Left and Right, respectively, and represent people moving away from a Neutral political opinion because the societal emphasis on moral foundations is similar to their personal beliefs when moving Left or dissimilar when moving Right. In other words, a Left-leaning Neutral individual in a wellbeing-emphasizing society will be more likely to identify as Left, while a Right-leaning Neutral individual in a wellbeing-emphasizing society is more likely to become more entrenched as Right in reaction. The flow also represents people moving away from Neutral because they are feeling frustration at injustices (directed towards themselves when moving Right or directed towards others when moving Left). Full mathematical definitions can be found in the supplemental materials, but the other stocks are similarly defined using the following key elements of the theory of Woke Anti-Woke polarization.

Key elements of the WAW model. First, as discussed above, people with a liberal moral orientation tend to embrace wellbeing foundations and suppress binding foundations, while those with a conservative moral orientation emphasize all foundations roughly equally. When the population emphasizes wellbeing foundations over binding foundations, people tend to migrate outwards in agreement or reactionary disagreement, and when wellbeing and binding foundations are balanced, people tend to move towards Neutral. This ratio of emphasis on wellbeing vs. binding foundations, MFT in Figure 1 or m in equation (1), is dynamic over the course of the simulation. Emphasis on binding foundations is increased by the number of people in the Radical Woke, Right, Anti-Woke and Radical Anti-Woke stocks, and reduced by the number of people in the Woke and Left stocks. This dynamic creates several feedback loops. People moving Left create a societal emphasis on wellbeing foundations which causes more people to move Left in a reinforcing feedback loop. Greater societal emphasis on wellbeing foundations also causes more people to move Right, which reduces the societal emphasis on wellbeing foundations in a balancing feedback loop. This Moral Foundations Theory force drives people between Woke and Anti-Woke but not to the extremes.

Second, justice sensitivity, JF in Figure 1 or j_l/j_r in equation (1), also drives people away from the center and towards the extremes. Consistent with Rothmund et al.'s findings, higher levels of justice sensitivity towards others drives people to the left, while higher levels of justice sensitivity towards oneself drives people to the right. In the model, justice sensitivities are constant values over the course of a single simulation, but in reality, justice

sensitivities vary based on economic and political trends. Justice sensitivities play a part in driving people all the way from Radical Woke to Radical Anti-Woke.

Third, people can be aggravated by the power of people on the opposite extreme (RW and RAW in Figure 1). Thus, people move from Left to Woke as a response to a large number of radical Anti-Woke people. Likewise, people might move from Right to Anti-Woke as a response to a large number of Radical Woke people.

In the model, two forces drive people back toward the center. On the one hand, people can be frustrated with extremist culture of all kinds and gravitate toward a more civil form of ideological disagreement (CF in Figure 1 and f_c in equation (1)). On the other hand, exposure to diverse viewpoints typically induces sufficient empathy to drive people away from extremes (EXP in Figure 1 and i_l/i_r in equation (1)).

Moving into the extremes involves new elements. To move from Woke to Radical Woke requires not only high levels of justice sensitivity towards others and aversive provocation of the opposing extreme, but also the power of homophily (HOM in Figure 1). In relation to homophily, it is attractive to belong to a group that sees things the same way as you do, and that common cause catalyzes a sense of community. In the case of radical wokeism, the sense of solidarity is strong enough that the binding moral foundations are activated to protect the Radical Woke group identity, despite the Left's commitment to resisting binding foundations. Much the same applies to the move from Anti-Woke to Radical Anti-Woke, except that it is justice sensitivity in relation to self that drives the indignation and homophily is reinforced by binding foundations quite naturally.

The general frustration with extremist ideology that drives people away from extremes needs to be articulated carefully because different groups are frustrated by different elements of radicalism. Those in the Left and Woke states are frustrated by anti-wokeism and by radical wokeism's embrace of the binding moral foundations that the left is otherwise committed to eliminating. People in the Neutral and Right states are frustrated by radicalism of all kinds. Finally, people in the Anti-Woke state are frustrated by radical wokeism. Diverse interactions are also key in bringing people to the center. Of course, individual interactions may have polarizing effects but, in this model, diverse interactions have a de-polarizing effect on average.
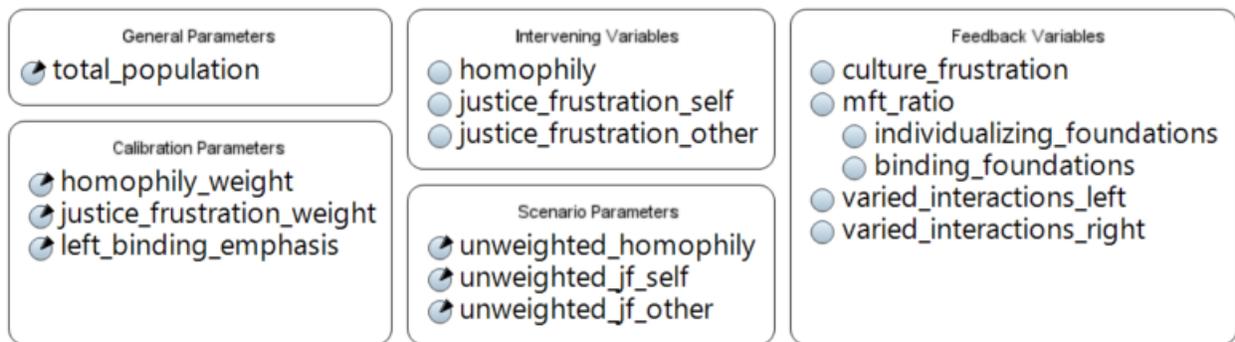

Figure 2. *Model Parameters and Variables.*

Figure 2 lists the parameters and variables used in the WAW SDM. For a complete description, see the supplemental materials. Parameters such as *total_population*, *homophily_weight*, and *unweighted_homophily* are the tuning knobs of the simulation that allow us to experiment with different scenarios and calibrate against real data; they remain constant during a single run of the simulation. Dynamic variables, such as *homophily* and *culture_frustration*, are defined using mathematical formulas and may change during the simulation. For example, *homophily* is calculated using *unweighted_homophily* and

homophily_weight, while *culture_frustration* is calculated using stock values and changes as people flow to different stocks. In all simulation runs, the population is initialized using a truncated normal distribution with no one in the radical stocks.

Wokeness metric through survey data

At the time of this study, there were no longitudinal surveys published in Norway measuring wokeness and anti-wokeness. To calibrate and test the WAW model, we constructed a "wokeness" metric (WM) using the Norwegian Citizen Panel (NCP). The NCP is a internet survey of public opinion in various elements of society and politics. The survey was first run in 2013 and has been carried out three times a year since 2017. Survey respondents are randomly selected from the Norwegian population register of Norwegian citizens 18 years or older, so people may participate in multiple surveys over time.

Survey questions frequently change between rounds, but topics such as immigration, climate change, and policy opinions are relatively constant, permitting contrasts over time. To create a WM, we combined questions related to various "woke" topics in Norway relying on three rounds with the widest variety of relevant subjects: round 3 conducted in 2014, round 8 from 2017, and round 27 from 2023. From round 3, selected questions were related to homosexual rights, immigration, abortion, disability, and climate change. Round 8's selected questions were related to homosexual rights, immigration, and climate change. Finally, round 27's selected questions were related to immigration, abortion, and climate change. Even though there is no measurement of wokeism in the NCP, we believe that an individual's stance on wokeism can be derived from their opinions on the above topics. A

key limitation of this survey and the metric is that they measure beliefs and opinions, rather

than behaviors. Because wokeism is more of a social movement than a political one,

measuring people's behaviors would be much better for determining the polarizing effect of

woke and anti-wokeism. Table 1 presents descriptive statistics of the rounds and their

selected questions.

Table 1. Descriptive statistics of the rounds and questions from the Norwegian Citizen Panel, 2014, 2017, 2023.

Variable	Year	Category	Description	Mean	Med	SD
W03_r3dvh_3	2014	LGBTQ+	Should het. and homo. couples have equal rights? (1-strongly agree to 7-strongly disagree)	2.33	2	1.68
W03_r3dv1	2014	Immigration	How advantageous or disadv. is it for Norway that immigrants settle here? (1-very great advantage to 7-very great disadvantage)	3.53	3	1.37
W03_qo11a	2014	Feminism	Should it be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children? (1-Yes or 2-No)	1.14	1	0.35
W03_qo11b	2014	Feminism	""	1.25	1	0.43
W03_u4_r3cc3aa_3	2014	Disabilities	Should the government use more or less money on disability benefits? (1-much more money to 5-much less money)	2.73	3	0.91
W03_u4_r3cc3ba_3	2014	Disabilities	""	2.76	3	0.92
W03_u4_r3cc3ca_3	2014	Disabilities	""	2.73	3	0.93
W03_u5_r3cc3aa_3	2014	Disabilities	""	2.60	3	0.96
W03_u5_r3ccba_3	2014	Disabilities	""	2.75	3	0.89
W03_u5_r3cc3ca_3	2014	Disabilities	""	2.70	3	0.87
W03_r3km23	2014	Climate	How concerned are you about climate change? (1-very concerned to 5-not concerned at all)	2.73	3	1.02
W03_r3km34	2014	Climate	Should Norway reduce, maintain, or increase oil production? (1-reduce, 2-maintain, or 3-increase)	1.71	2	0.59
W03_r3km35	2014	Climate	""	1.72	2	0.61
W03_p5_1	2014	Demographics	Year of birth (birth decade starting at 1-1939 or earlier to 7-1990 or later)	4.08	4	1.63

W03_r3k8_1	2014	Demographics	Self rated political scale (1-left wing to 11-right wing)	6.09	6	2.27
R8parlikhet	2017	LGBTQ+	Should het. and homo. couples have equal rights? (1-strongly agree to 7-strongly disagree)	2.18	2	1.55
R8dv5	2017	Immigration	How advantageous or disadv. is it for Norway that immigrants settle here? (1-very great advantage to 7-very great disadvantage)	3.62	3	1.37
R8innvandring	2017	Immigration	Assertion: Immigration is a threat to Norwegian culture. (1-strongly agree to 7-strongly disagree)	4.10	4	1.88
R8bekym	2017	Climate	How concerned are you about climate change? (1-very concerned to 5-not concerned at all)	2.55	3	1.01
R8vekst	2017	Climate	Assertion: Economic growth should be safeguarded through the development of industry, even if it conflicts with environmental interests. (1-strongly agree to 7-strongly disagree)	4.27	4	1.65
R8p5_1	2017	Demographics	Year of birth (birth decade starting at 1-1939 or earlier to 7-1990 or later)	4.09	4	1.66
R8k8_1	2017	Demographics	Self rated political scale (1-left wing to 11-right wing)	6.05	6	2.13
r24_pcimm	2023	Immigration	How advantageous or disadv. is it for Norway that immigrants settle here? (1-very great advantage to 7-very great disadvantage)	3.20	3	1.29
r23meme50	2023	Immigration	Generally, how much do you like or dislike immigrants in	4.75	5	1.22

			Norway? (1-intensely dislike to 7-intensely like)			
r27_faabe	2023	Feminism	Should the right to self-determined abortion in Norway be extended? (1-completely disagree to 5-completely agree)	2.81	3	1.41
r23pkkm2	2023	Climate	How concerned are you about climate change? (1-very concerned to 5-not concerned at all)	3.43	4	1.11
r23pk_vekst	2023	Climate	Assertion: Economic growth should be safeguarded through the development of industry, even if it conflicts with environmental interests (1-strongly agree to 7-strongly disagree)	4.31	5	1.67
r27P5_1	2023	Demographics	Year of birth (birth decade starting at 1-1939 or earlier to 7-1990 or later)	4.00	4	1.52
r26k2_bglrs	2023	Demographics	Self rated political scale (1-left wing to 11-right wing)	4.95	5	2.47

Responses to each of the selected questions were on a 5-point or 7-point Likert scale. To create the WM, responses to each question were scaled to have values from -2 to 2, where negative responses are more "woke", positive responses are more "anti-woke", and 0 responses are neutral. Then questions within a single category (e.g. immigration) were averaged to create a respondent's category score. NCP responses are optional, so we needed to handle missing data. We eliminated respondents who answered questions only in a single category. Next, respondent scores were squared and set to their original sign (equation (2)). This has the effect of moving extreme opinions further from 0 and moderate opinions closer to 0. Each individual category score was averaged to create an overall WM

that ranges from -4 to 4. Finally, we rescale the WM to range from 0 to 6 and then partition

and take percentages of the following categories: Radical Woke (0.0-0.5), Woke (0.5-1.5),

Left (1.5-2.5), Neutral (2.5-3.5), Right (3.5-4.5), Anti-Woke (4.5-5.5), Radical Anti-Woke (5.5-

6.0). Table 2 contains the rescaled and binned WM distributions.

Table 2. Woke Metric distributions rescaled and binned to fit the WAW model distribution.

	WM 2014	WM 2017	WM 2023
Radical Woke	3.34%	3.69%	1.39%
Woke	36.07%	19.92%	9.84%
Left	32.55%	37.97%	29.59%
Neutral	15.90%	26.62%	42.73%
Right	8.09%	8.93%	13.24%
Anti-Woke	3.07%	2.45%	2.87%
Radical Anti-Woke	0.97%	0.42%	0.34%

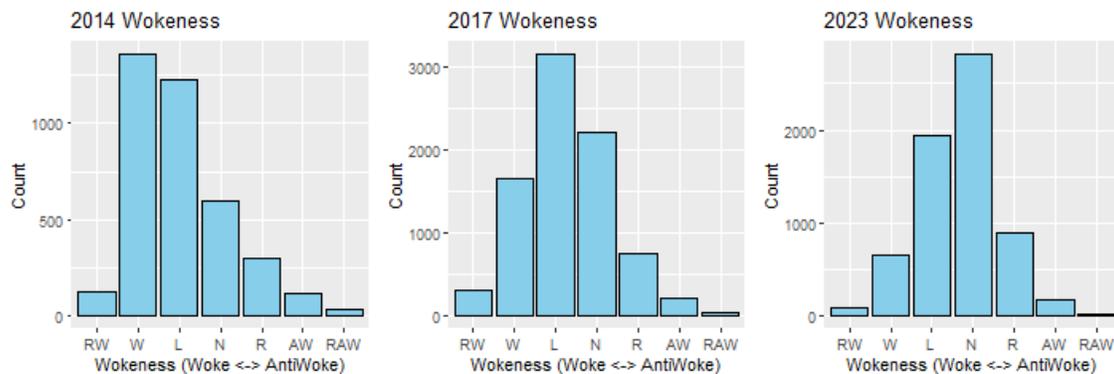$$x^* = sign(x) * x^2 \qquad\qquad 2$$

Figure 3. Woke metric scores from 2014, 2017, and 2023 in Norway, based on Norwegian Citizen Panel data.

Figure 3 shows the overall WM scores from 2014, 2017, and 2023. Between 2014 and 2017,

there was a shift towards the center, which continued between 2017 and 2023. In 2023, we

also observe a scarcity of extreme opinions on either side of the spectrum. The shift to center matches what we hear from our subject-matter experts, but the lack of extremes is contrary to our subject matter experts' consensus that extreme views are becoming more prevalent in Norway over time. We revisit this observation in the discussion section.

Results

Calibration

We first calibrate the model to the data in the Wokeness Metric Survey Data to set the model's calibration parameters (homophily weight, justice frustration weight, and left binding emphasis). We use the Anylogic genetic optimization algorithm to minimize the Kullback-Leibler divergence between the WM distribution and the population distribution produced by the WAW model (Kullback & Leibler, 1951).

We optimize both the calibration and scenario parameters against the three WM distributions and collect optimal parameter values. We then average the values for the calibration parameters to get final values (these are described in the supplemental materials). Finally, we rerun the optimization while varying the scenario parameters and keeping the final calibration parameters static. We observe that there is always more JF-Other than JF-Self, and that homophily is rising over time. Figure 4 shows the population distributions produced by the WAW model in each scenario compared to the WM distributions.
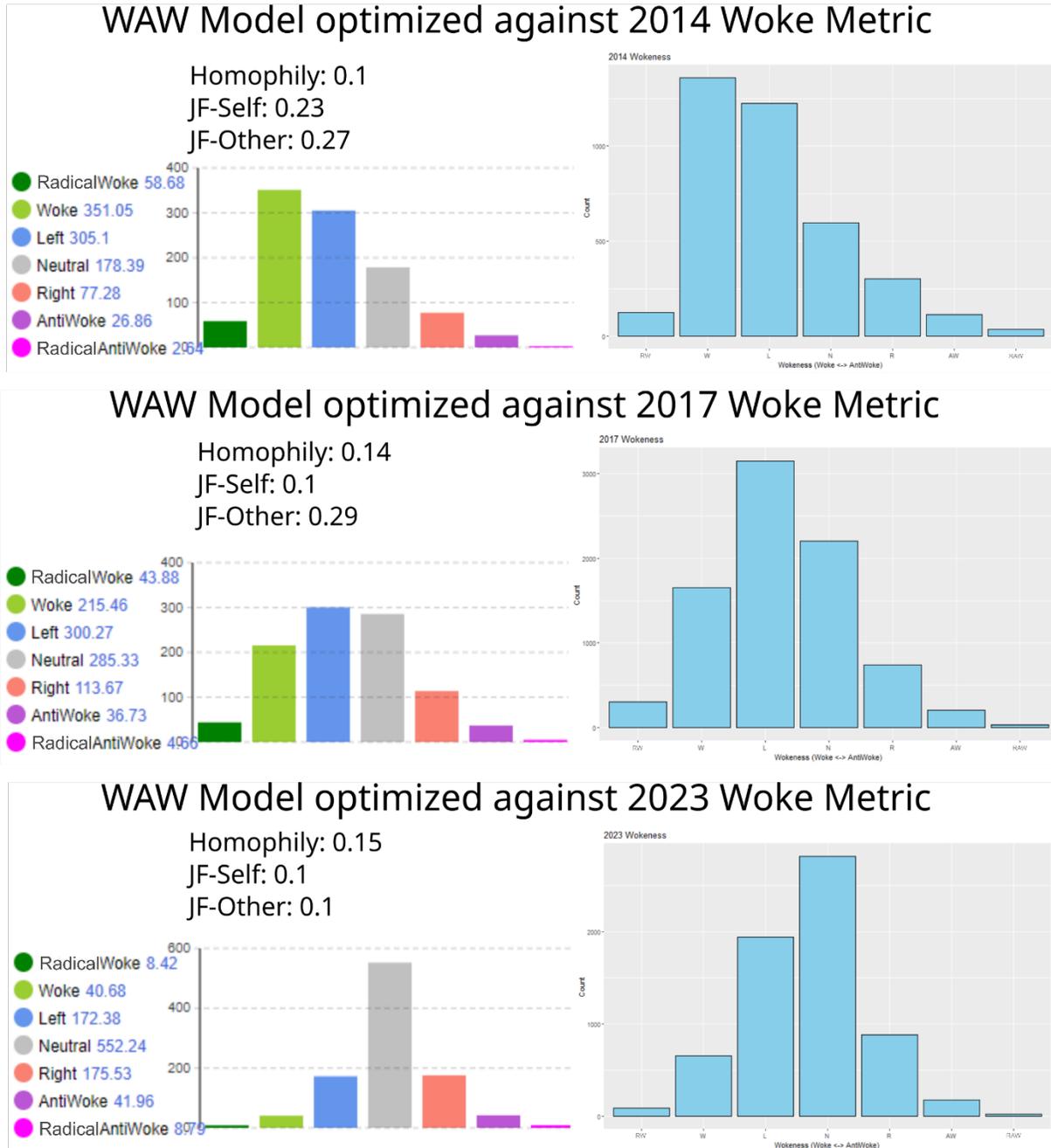
WAW Model optimized against 2014 Woke Metric

Homophily: 0.1
JF-Self: 0.23
JF-Other: 0.27

RadicalWoke 58.68
Woke 351.05
Left 305.1
Neutral 178.39
Right 77.28
AntiWoke 26.86
RadicalAntiWoke 2.64

WAW Model optimized against 2017 Woke Metric

Homophily: 0.14
JF-Self: 0.1
JF-Other: 0.29

RadicalWoke 43.88
Woke 215.46
Left 300.27
Neutral 285.33
Right 113.67
AntiWoke 36.73
RadicalAntiWoke 4.66

WAW Model optimized against 2023 Woke Metric

Homophily: 0.15
JF-Self: 0.1
JF-Other: 0.1

RadicalWoke 8.42
Woke 40.68
Left 172.38
Neutral 552.24
Right 175.53
AntiWoke 41.96
RadicalAntiWoke 8.79

Figure 4. WAW Model calibration results.

Output metrics

One could use the WAW SDM for many different purposes. We construct two output metrics for the WAW model. The first is the ratio of people in the radical stocks to the rest (RtoAll in equation (3)); it ranges from 0 to 1, with higher values indicating higher amounts of radicalism. This ratio indicates the degree of polarized radicalism in the population. Policy professionals wanting to decrease political radicalism might aim to minimize this measure. The second is the ratio of people in the radical woke stock versus those in the radical anti-woke stock (RWtoRAW in equation (4)); it ranges from -1 to 1, with positive values indicating more people in radical woke than in radical anti-woke, negative values indicating more people in radical anti-woke than radical woke, and 0 indicating equal amounts in both extremes. This measure indicates which extreme is predominant in the population. Policy professionals wanting to win a culture war might focus their attention on this dependent variable. The second ratio gives context to the first, allowing us to see if the scales are tipped one way or the other at different levels of radicalism.

$$RtoAll = \frac{RW + RAW}{RW + W + L + N + R + AW + RAW} \qquad 3$$

$$RWtoRAW = \frac{RW - RAW}{RW + RAW} \qquad\qquad 4$$
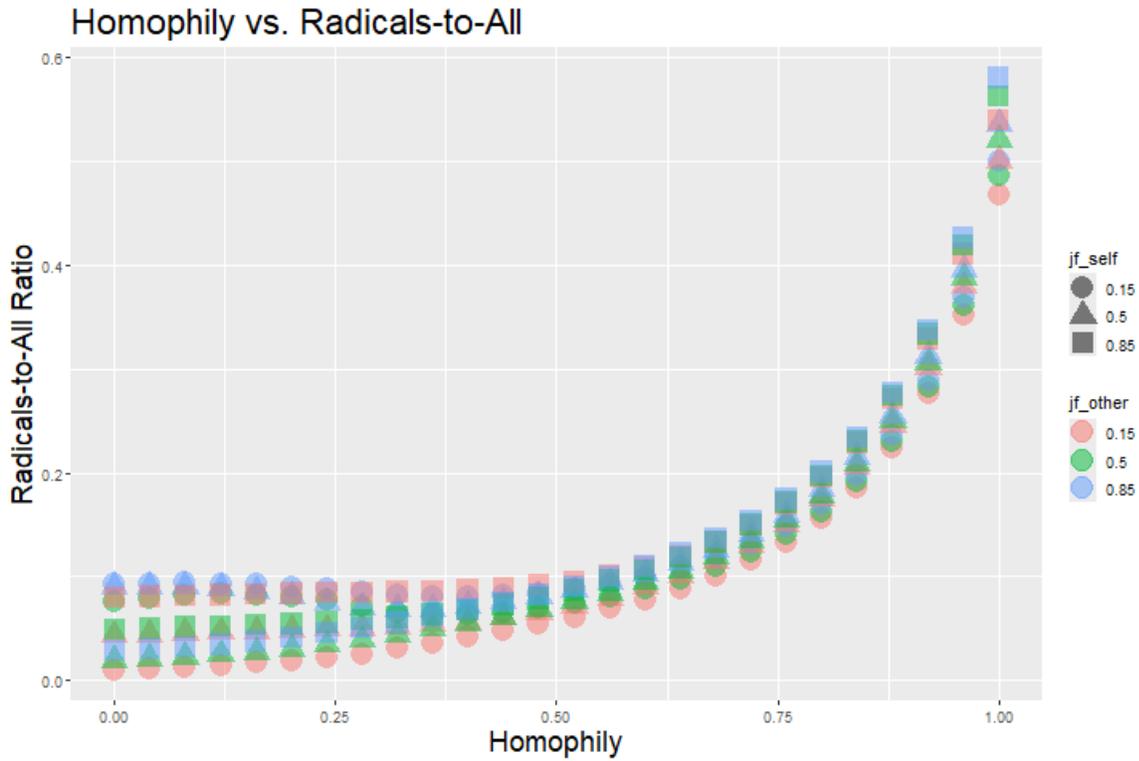
Parameter sweep

Figure 5. Effect of homophily on the ratio of extremes to the total population.

After we calibrated the WAW model, we ran a parameter sweep across the three scenario parameters to visualize the behavior space of the model. Figure 5 shows the effects of varying homophily on the ratio between extremes to the total population at various levels of the two forms of justice sensitivity. In general, higher homophily leads to high radicalism, but at low homophily values, radicalism does not follow the same trends, and we see a bifurcation where lower homophily can instead increase radicalism.

Figure 6 presents the effect on the ratio between the radical woke and radical anti-woke stocks and helps to explain the result of Figure 5. In this figure, lower values indicate more radical anti-wokeism and higher values indicate more radical wokeism. At higher levels of homophily, radical wokeism and radical anti-wokeism are evenly balanced, even with very unbalanced levels of justice frustration. Homophily and the other balancing feedback loops in the model overpower any effect from justice frustration. On the other hand, at low levels of homophily, we see a bifurcation where some justice frustration values lead to radical wokeism dominating radical anti-wokeism, while other values lead to radical anti-wokeism dominating radical wokeism.

This bifurcation occurs with low homophily and any level of justice sensitivity as long JF-Other is greater than or equal to JF-Self. Figure 7 shows this behavior at high and low levels of justice frustration.
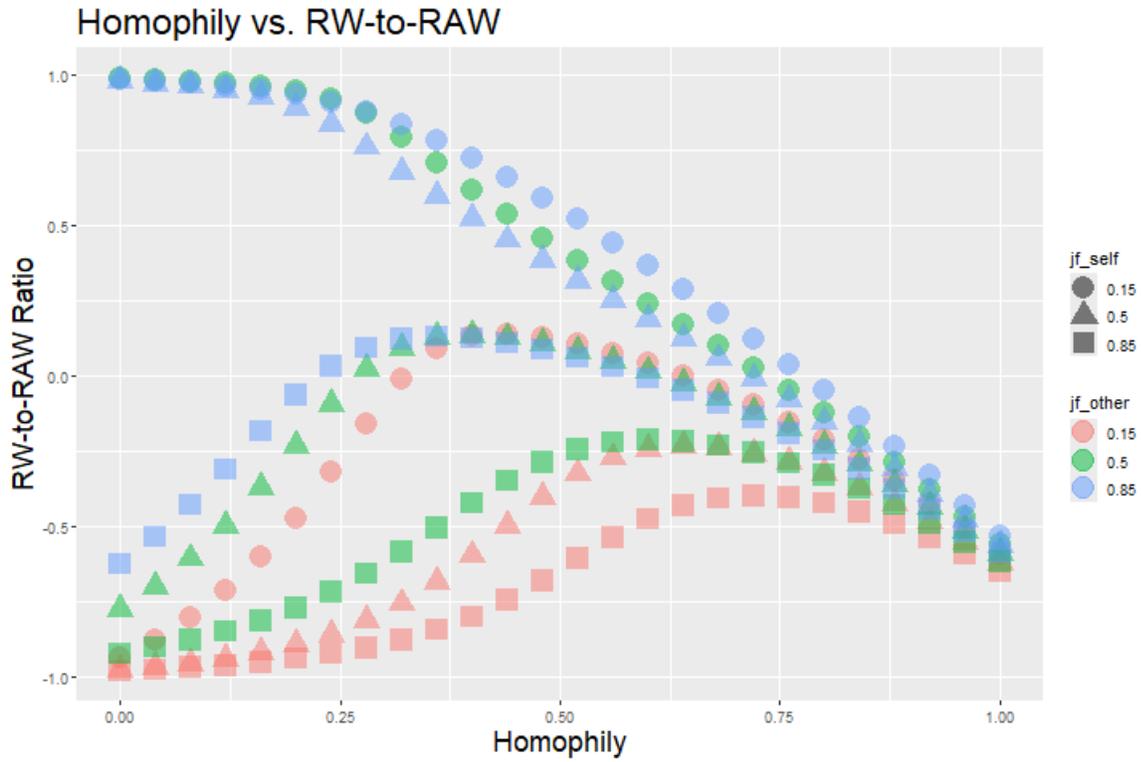

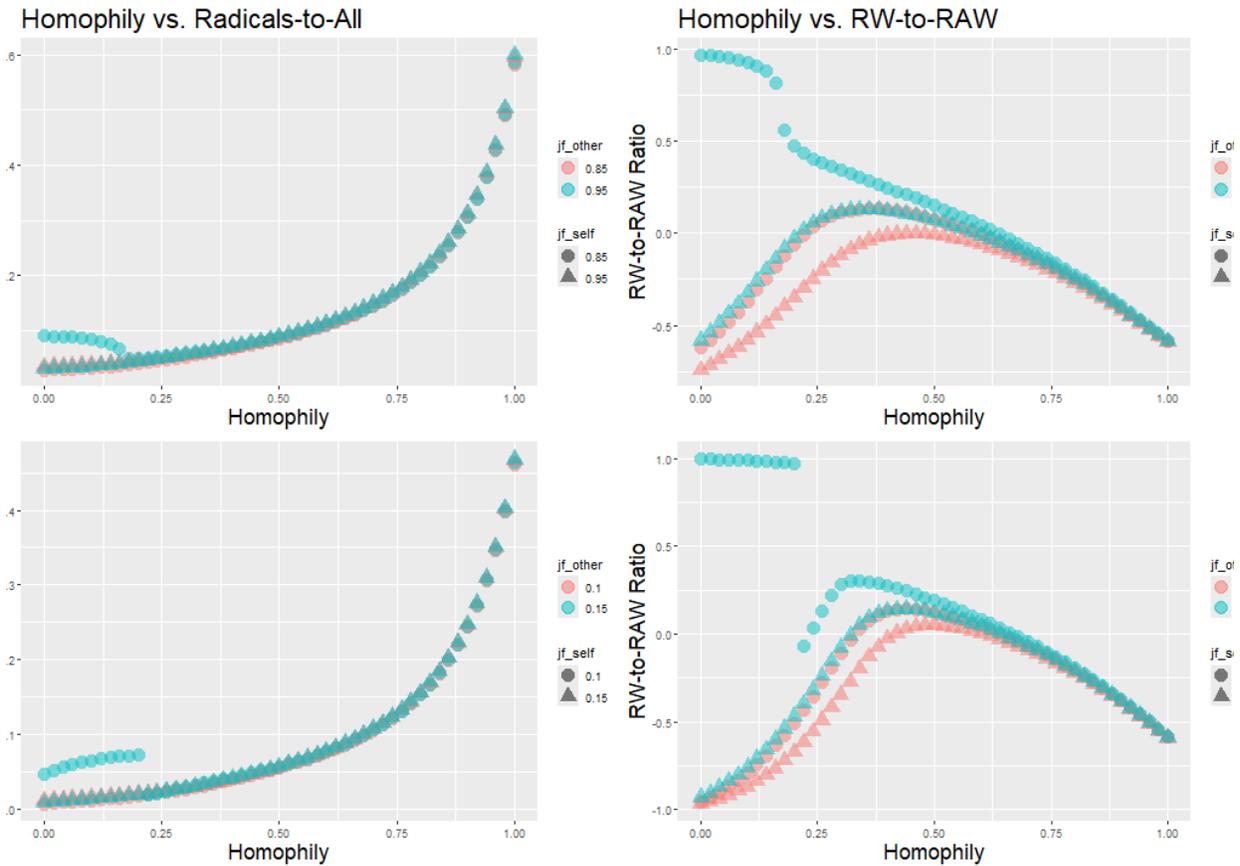**Figure 6.** Effect of homophily on the ratio of *radical woke to radical anti-woke.*

Figure 7. Effect of homophily on the ratio of extremes to the total population (left) and the ratio of *radical woke to radical anti-woke (right). Top figures show the effect at high justice frustration levels and bottom figures show the effect at low justice frustration levels.*

This behavior yields two insights. First, the system produces radical anti-wokeism more easily than radical wokeism. At equal justice frustration values, the system always leans towards radical anti-wokeism (all red shapes, green triangles and squares, and blue squares in Figure 6. Second, at low homophily values, it is possible to observe a sudden shift in radicalism with very small changes in parameters. We revisit this insight in the discussion section.

For the same values at which we see a sudden shift towards radical wokeism, we also see a jump in overall radicalism. In these scenarios, the entire population shifts to the Left side of Neutral, and so Radical Woke becomes larger from the sheer size of the Woke stock. Figure 8 shows the stock distribution with very low homophily, while Figure 9 shows the distribution with a slightly higher homophily.
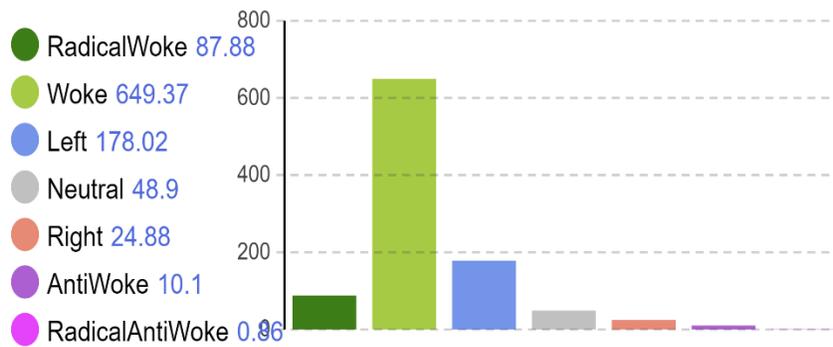

Figure 8. Stock distribution with low homophily (Homophily=0.05, JF-Self=0.5, JF-Other=0.55).
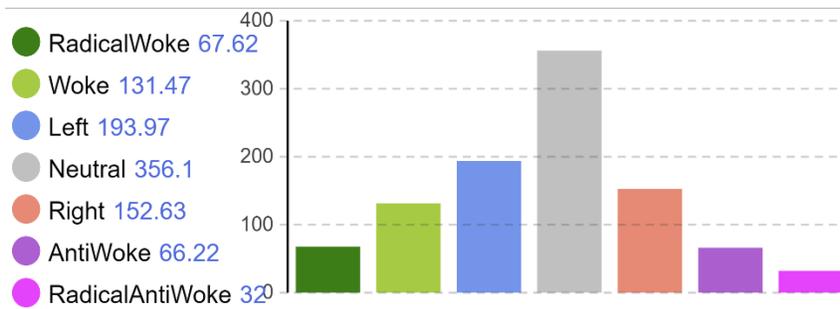

Figure 9. Stock distribution with medium homophily (Homophily=0.4, JF-Self=0.5, JF-Other=0.55).

Discussion

In the behavior of the model, we find that homophily increases radicalism, but at low homophily we instead observe a sudden increase in radicalism as well as a sudden shift to radical anti-wokeism or, more frequently, radical wokeism. Low levels of homophily can produce higher levels of radicalism in the simulation because the population tends to converge toward one side of the ideological spectrum. Although such a complete shift is unlikely to occur in real societies, a population that leans heavily toward one ideological pole will naturally exhibit more radicalism on that side. This occurs because the ideological distance to the extreme is reduced, and individuals can still become trapped in echo chambers, even when their preference homophily is low. In the model, low homophily corresponds to highly diverse social environments, which are most often characteristic of urban settings (McPherson et al., 2001). And indeed, urban environments are frequently the birthplace of social and political movements.

In his 1983 book, *The City and the Grassroots*, Manuel Castells identified three characteristics of political activities in cities that are unique. (Castells, 1983) The characteristics are *collective consumption*, which shapes the grievances and interests of an urban population; *local networks/territorial identities*, which provide meaning and purpose; and the *local state*, which shapes and reinforces the concerns of urban activists (Castells, 1983). Richard Sennett takes this a step further. Building on an observation from Georg Simmel's letters, Sennet writes, "the stranger is a bearer of a new kind of freedom" (Vertovec & Cohen, 2002, Chapter 4, p. 43). The freedom to interact with diverse others

makes it possible to move away from convention and support new social or political movements.

In the Woke Metric distributions, we see a heavily woke leaning population in 2014 with a fair amount of radical woke opinions. In 2017 and 2023, we see an increasingly centrist population with fewer extreme opinions. In 2014, Norway had a conservative government that has become more liberal over time. One explanation for the WM trends is that the Norwegian population tends to have strong oppositional views when the government is more conservative, and more neutral views when the government is more liberal.

On the other hand, the WM distributions paint a different picture than the reports from our collaborators in Norway, who say that woke and anti-woke extremism has only been getting more prevalent over time in political and media discourse. The NCP data show that, instead of being very polarized, the Norwegian population is generally centrist in their beliefs and getting more centrist over time. This apparent paradox highlights the difference between *beliefs*, which are measured by the NCP data, and *behaviors*, which are observed by our Norwegian colleagues. The gap between perceptions of beliefs and actual held beliefs has been well studied and has many causes, such as political elite discourse, polarized online and traditional media, and social media echo chambers (Lee, 2022; Levendusky & Malhotra, 2016; Westfall et al., 2015; Yang et al., 2016; Yudkin et al., 2019). It also may derive from polarization within political elites rather than within the general population, as is common in polarizing societies (Prior, 2013).

Policy interventions

The WAW model in its current state is not complex enough to investigate detailed policy interventions. Nevertheless, the results suggest several basic policy interventions related to the model's parameters, homophily, self-oriented justice frustration and other-oriented justice frustration. Except at very low levels of homophily, reducing homophily reduces overall radicalism. This can be achieved through policies that encourage cross-group contact, such as urban planning measures that create mixed neighbourhoods and schools or workplace diversity programs. Reducing homophily is also possible by decreasing 'echo chambers' through social media algorithm reform or media literacy education. Justice sensitivities play a role in balancing opinions but require more complex and thorough policy interventions. To reduce self-oriented justice frustration, a policymaker may attempt to increase overall economic security, which will reduce a sense of threat to one's own position. To reduce other-oriented justice frustration, the policymaker should instead focus on decreasing societal inequalities.

Conclusions and future work

The WAW model was successfully able to recreate the empirical population ideology distributions from three different years of the Woke Metric. This suggests that the underlying theory is coherent and potentially useful for guiding future research. In the theory, a person's beliefs react to others' behaviors via Moral Foundations Theory. The primary limitation of the validation efforts in this study is the Woke Metric measures individuals' beliefs around "woke topics," rather than potentially polarizing woke or anti-

woke behaviors. Therefore, future work should include collecting data on behaviors related to woke/anti-woke dynamics, as well as on perceptions of others' beliefs so that we can more accurately calibrate the WAW model to Norwegian woke/anti-woke dynamics. There have been a few surveys that ask Norwegian respondents if they know of the term "woke" or self-identify as woke, but none that quantify woke and anti-woke behaviors, or that link those behaviors with underlying beliefs and social environments. In future work, we also plan to model woke/anti-woke decision processes at an individual level using an agent-based model, rather than just at the system level. An individual-level model will allow us to examine the behaviour and decision mechanisms that would cause an individual to step toward or away from radicalism, such as traumatic events or persuasive encounters with others, as well as investigate detailed policy interventions.

The WAW model presents a theory of population-level woke/anti-woke dynamics, driven by Moral Foundations Theory and Rothmund et al.'s theory of two types of justice sensitivity. This social dynamic is rarely studied formally, and displays fascinating characteristics, including ideological asymmetry wherein the left move towards woke ideals which provokes a very strong reaction in the right. We quantify the underlying theory in a systems dynamics simulation and calibrate the simulation against longitudinal data from the Norwegian Citizen Panel. This provides insights into both balanced and unbalanced polarization. It also provides empirical support for policy interventions, mostly confirming commonsense intuitions, but also revealing possibilities such as bifurcation at low levels of homophily. The WAW model can be calibrated to many types of societies and help explain shifts in polarisation and radicalism over time.

References

Asen, R. (2024). Anti-Woke Publics. *Political Communication*, *41*(6), 1029–1034. https://doi.org/10.1080/10584609.2024.2425782

Axelrod, R. (1997). The Dissemination of Culture: A Model with Local Convergence and Global Polarization. *Journal of Conflict Resolution*, *41*(2), 203–226. https://doi.org/10.1177/0022002797041002001

Baldassarri, D., & Bearman, P. (2007). Dynamics of Political Polarization. *American Sociological Review*, *72*(5), 784–811. https://doi.org/10.1177/000312240707200507

Banisch, S., & Olbrich, E. (2021). An Argument Communication Model of Polarization and Ideological Alignment. *Journal of Artificial Societies and Social Simulation*, *24*(1), 1.

Bertotti, M. L., & Menale, M. (2024). Opinion dynamics models describing the emergence of polarization phenomena. *Journal of Computational Social Science*, *7*(3), 2591–2612. https://doi.org/10.1007/s42001-024-00319-x

Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., & Holman, B. (2017). Understanding Polarization: Meanings, Measures, and Model Evaluation. *Philosophy of Science*, *84*(1), 115–159. https://doi.org/10.1086/688938

Cano Macias, R., & Ruiz Vera, J. M. (2024). Dynamics of opinion polarization in a population. *Mathematical Social Sciences*, *128*, 31–40. https://doi.org/10.1016/j.mathsocsci.2024.01.009

Carpentras, D., Lueders, A., Maher, P. J., O'Reilly, C., & Quayle, M. (2023). How Polarization Extends to New Topics: An Agent-Based Model Derived from Experimental Data. *Journal of Artificial Societies and Social Simulation*, *26*(3), 2.

Centola, D., González-Avella, J. C., Eguíluz, V. M., & San Miguel, M. (2007). Homophily, Cultural Drift, and the Co-Evolution of Cultural Groups. *Journal of Conflict Resolution*, *51*(6), 905–929. https://doi.org/10.1177/0022002707307632

Cole, J. C., Gillis, A. J., van der Linden, S., Cohen, M. A., & Vandenbergh, M. P. (2025). Social Psychological Perspectives on Political Polarization: Insights and Implications for Climate Change. *Perspectives on Psychological Science*, *20*(1), 115–141. https://doi.org/10.1177/17456916231186409

Conway III, L. G., Zubrod, A., Chan, L., McFarland, J. D., & Van de Vliert, E. (2023). Is the myth of left-wing authoritarianism itself a myth? *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1041391

Conway, L. G., Houck, S. C., Gornick, L. J., & Repke, M. A. (2018). Finding the Loch Ness Monster: Left-Wing Authoritarianism in the United States. *Political Psychology*, *39*(5), 1049–1067. https://doi.org/10.1111/pops.12470

Costello, T. H., Bowes, S. M., Stevens, S. T., Waldman, I. D., Tasimi, A., & Lilienfeld, S. O. (2022). Clarifying the structure and nature of left-wing authoritarianism. *Journal of Personality and Social Psychology*, *122*(1), 135–170. https://doi.org/10.1037/pspp0000341

Dalton, R. J. (1987). Generational Change in Elite Political Beliefs: The Growth of Ideological Polarization. *The Journal of Politics*, *49*(4), 976–997. https://doi.org/10.2307/2130780

Deffuant, G. (2006, June 30). *Comparing Extremism Propagation Patterns in Continuous Opinion Models* [Text.Article]. JASSS. https://www.jasss.org/9/3/8.html

Deffuant, G. A. (2002, October 31). *How can extremism prevail? A study based on the relative agreement interaction model* [Text.Article]. JASSS. https://web-archive.southampton.ac.uk/cogprints.org/4355/1/1.html

Dickinson, J. L., McLeod, P., Bloomfield, R., & Allred, S. (2016). Which Moral Foundations Predict Willingness to Make Lifestyle Changes to Avert Climate Change in the USA? *PLOS ONE*, *11*(10), e0163852. https://doi.org/10.1371/journal.pone.0163852

Egerer, S., Cotera, R. V., Celliers, L., & Costa, M. M. (2021). A leverage points analysis of a qualitative system dynamics model for climate change adaptation in agriculture. *Agricultural Systems*, *189*, 103052. https://doi.org/10.1016/j.agsy.2021.103052

Farkhari, F., Scharbert, J., Kroencke, L., Schwarzer, C., Koch, J. F., van Zalk, M. H. W., Schlipphak, B., & Back, M. D. (2025). Right-wing authoritarianism and perceptions that minoritized groups pose a threat: The moderating roles of individual- and country-level religiosity and marginalization. *The British Journal of Social Psychology*, *64*(1), e12830. https://doi.org/10.1111/bjso.12830

Flache, A., & Macy, M. W. (2006). *What sustains cultural diversity and what undermines it? Axelrod and beyond* (No. arXiv:physics/0604201). arXiv. https://doi.org/10.48550/arXiv.physics/0604201

Forrester, J. (1961). Industrial Dynamics. *Waltham MA, Pegasus Communications*.

Gu, Y., & Wang, Z. (2022). Income Inequality and Global Political Polarization: The Economic Origin of Political Polarization in the World. *Journal of Chinese Political Science*, *27*(2), 375–398. https://doi.org/10.1007/s11366-021-09772-1

Haque, A., Ajmeri, N., & Singh, M. P. (2023). Understanding dynamics of polarization via multiagent social simulation. *AI & SOCIETY*, *38*(4), 1373–1389. https://doi.org/10.1007/s00146-022-01626-5

Hegselmann, R., & Krause, U. (n.d.). *OPINION DYNAMICS AND BOUNDED CONFIDENCE MODELS, ANALYSIS, AND SIMULATION*.

Kitts, J. A. (2006). Social influence and the emergence of norms amid ties of amity and enmity. *Simulation Modelling Practice and Theory*, *14*(4), 407–422. https://doi.org/10.1016/j.simpat.2005.09.006

Klemm, K., Eguíluz, V. M., Toral, R., & Miguel, M. S. (2005). Globalization, polarization and cultural drift. *Journal of Economic Dynamics and Control*, *29*(1), 321–334. https://doi.org/10.1016/j.jedc.2003.08.005

Kołczyńska, M. (2025). Does polarization increase participation? A systematic literature review and meta-analysis. *European Political Science Review*, *17*(4), 553–568. https://doi.org/10.1017/S1755773925000116

Kubin, E., & von Sikorski, C. (2021). The Role of (Social) Media in Political Polarization: A Systematic Review. *Annals of the International Communication Association*, *45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Leininger, A., Grünewald, F., & Buntfuß, N. (2023). *Ideological and affective polarization in multiparty systems*. SocArXiv. https://doi.org/10.31235/osf.io/mz6rs

Leonard, N. E., Lipsitz, K., Bizyaeva, A., Franci, A., & Lelkes, Y. (2021). The nonlinear feedback dynamics of asymmetric political polarization. *Proceedings of the National Academy of Sciences*, *118*(50), e2102149118. https://doi.org/10.1073/pnas.2102149118

Marino, G., & Iannelli, L. (2023). Seven years of studying the associations between political polarization and problematic information: A literature review. *Frontiers in Sociology*, *8*. https://doi.org/10.3389/fsoc.2023.1174161

Mason, L. (2022). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.

McCoy, J., & Somer, M. (2019). Toward a Theory of Pernicious Polarization and How It Harms Democracies: Comparative Evidence and Possible Remedies. *The ANNALS of the American Academy of Political and Social Science*, *681*, 234–271. https://doi.org/10.1177/0002716218818782

Osborne, D., Costello, T. H., Duckitt, J., & Sibley, C. G. (2023). The psychological causes and societal consequences of authoritarianism. *Nature Reviews Psychology*, *2*(4), 220–232. https://doi.org/10.1038/s44159-023-00161-4

Pirro, A. L. P. (2024). The Contemporary Far Right from Contra to Control. *Political Communication*, *41*(6), 1017–1022. https://doi.org/10.1080/10584609.2024.2414256

Prior, M. (2013). Media and Political Polarization. *Annual Review of Political Science*, *16*(Volume 16, 2013), 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242

Rempala, D. M., Okdie, B. M., & Garvey, K. J. (2016). Articulating ideology: How liberals and conservatives justify political affiliations using morality-based explanations. *Motivation and Emotion*, *40*(5), 703–719. https://doi.org/10.1007/s11031-016-9563-9

Salajan, F. D., & jules, tavis d. (2024). The Global Resurgence of Authoritarianism and Its Existential Threats to Education: Implications for Scholarship in Comparative and International Education. *Comparative Education Review*, *68*(3), 319–344. https://doi.org/10.1086/732119

Schedler, A. (2023). Rethinking Political Polarization. *Political Science Quarterly*, *138*(3), 335–359. https://doi.org/10.1093/psquar/qqad038

Shin, Y. A., Constantino, S. M., Gross, L. J., Kinzig, A., Lacasse, K., & Beckage, B. (2025). Culture Mediates Climate Opinion Change: A System Dynamics Model of Risk Perception, Polarization, and Policy Effectiveness. *Climate*, *13*(9), 194. https://doi.org/10.3390/cli13090194

Shults, F. L., Lane, J. E., Wildman, W. J., Diallo, S., Lynch, C. J., & Gore, R. (2018). Modelling terror management theory: Computer simulations of the impact of mortality salience on religiosity. *Religion, Brain & Behavior*, *8*(1), 77–100. https://doi.org/10.1080/2153599X.2016.1238846

Smith, D. S., Boag, L., Keegan, C., & Butler-Warke, A. (2023). Land of Woke and Glory? The Conceptualisation and Framing of "Wokeness" in UK Media and Public Discourses. *Javnost - The Public*, *30*(4), 513–533. https://doi.org/10.1080/13183222.2023.2273656

Sterman, J. D. (2002). *System Dynamics: Systems Thinking and Modeling for a Complex World* [Dataset]. https://doi.org/10.1287/420a0422-8060-4775-b8d7-a8f01fa36ffe

Tokita, C. K., Guess, A. M., & Tarnita, C. E. (2021). Polarized information ecosystems can reorganize social networks via information cascades. *Proceedings of the National Academy of Sciences*, *118*(50), e2102147118. https://doi.org/10.1073/pnas.2102147118

VanDreew, B. M., Phillips, J. B., Munis, B. K., & Goidel, S. (2025). What's woke? Ordinary Americans' understandings of wokeness. *Research & Politics*, *12*(2), 20531680251335650. https://doi.org/10.1177/20531680251335650

Vaughan, T. J., Bell Holleran, L., & Silver, J. R. (2019). Applying Moral Foundations Theory to the Explanation of Capital Jurors' Sentencing Decisions. *Justice Quarterly*, *36*(7), 1176–1205. https://doi.org/10.1080/07418825.2018.1537400

Vilanova, F., Milfont, T. L., & Costa, A. B. (2023). Short version of the right-wing authoritarianism scale for the Brazilian context. *Psicologia: Reflexão e Crítica*, *36*(1), 17. https://doi.org/10.1186/s41155-023-00260-4

Völker, T., & Saldivia Gonzatti, D. (2024). Discourse Networks of the Far Right: How Far-Right Actors Become Mainstream in Public Debates. *Political Communication*, *41*(3), 353–372. https://doi.org/10.1080/10584609.2024.2308601

Xu, T., Birchfield, A. B., & Overbye, T. J. (2018). Modeling, Tuning, and Validating System Dynamics in Synthetic Electric Grids. *IEEE Transactions on Power Systems*, *33*(6), 6501–6509. https://doi.org/10.1109/TPWRS.2018.2823702

Yang, H., Bukkapatnam, S. T. S., & Barajas, L. G. (2013). Continuous flow modelling of multistage assembly line system dynamics. *International Journal of Computer Integrated Manufacturing*, *26*(5), 401–411. https://doi.org/10.1080/0951192X.2012.719085

Zangari, L., Greco, C. M., Picca, D., & Tagarelli, A. (2025). A survey on moral foundation theory and pre-trained language models: Current advances and challenges. *AI & SOCIETY*, *40*(6), 4973–4998. https://doi.org/10.1007/s00146-025-02225-w