I disagree, because: attributional divides in online redistributive discourse*

Abstract

This paper examines how redistributive issues are framed and attributed in politically engaged online discourse. We collect data from Reddit, focusing on debates within the UKPolitics subreddit from 2020 to 2025, and focus on the discussions of council tax, universal basic income, minimum wage, income tax, national insurance, and poverty. We use BERTopic to recover the main thematic structure of the corpus, and apply large language models (LLMs) to annotate comment stance and attribution in relation to the argument advanced in the original post. The analysis shows that online redistributive discourse is organised less by a simple pro- versus anti-redistribution divide than by issue-specific attributional patterns. Government and political parties dominate users' attribution in tax-related debates, while minimum wage discussions more often centre on employers and labour-market dynamics. These attributional patterns also vary around major policy events, with several high-salience announcements followed by shifts from partisan attribution toward institutional government blame. We further present two conversation-tree case studies that help illustrating how these broader patterns appear within nested exchanges. The paper contributes to the policy studies with a computationally informed account of how redistributive discourse is organised by attribution and issue-specific framing in a politically engaged online arena, it also shows how these patterns speak to the changing legitimacy and moral language of redistribution in the contemporary UK.

*Current version: April 2026

1. Introduction

The post-COVID period witnesses deeper disagreement across social groups on major social and political problems. The waves of strikes across sectors in the UK demonstrate the pressure many people feel around the economy, distribution, and the redistribution of resources. Despite the widespread support visible at the picket line, debates and disagreements on related issues are frequently found online. Under the pressures of surging living costs, the long-tail Brexit influence, and the difficult recovery from an unprecedented pandemic, people are managing both the realization of their own interests and the strain of judging what others should receive. We are witnessing a more divided public sphere on social and political issues in the UK and across the world (Jungkunz, 2021; Rivas-de Roca and García-Gordillo, 2022). Disagreement over redistributive issues is especially visible in cyberspace, where people debate the necessity, legitimacy, and feasibility of social security, tax cuts, strikes, and other related topics.

Scholars of ideologies and belief systems have long addressed viewpoint divisions when explaining how people form groups of connected worldviews, but they often begin from the assumption that relatively structured sets of beliefs already exist and can be observed through surveys on political and social issues (Converse, 2006; Barker and Tinnick, 2006; Jost et al., 2009; Martin and Desmond, 2010; Baldassarri and Goldberg, 2014; Boutyline and Vaisey, 2017; Rawlings, 2022). Yet, as Feinberg and Willer Feinberg and Willer (2013) suggest, many issues have multiple aspects, and interpretations can be drawn from several psychological and moral principles. Most of this literature relies on conventional surveys and experiments, though recent work has expanded the analysis using digital traces such as text, networks, and web-tracking data (Yarchi et al., 2021; Xing et al., 2022; Urman and Makhortykh, 2023). We can also find related clues in controversy analysis, where scholars map the distribution and dynamics of disagreement around controversial issues online, particularly in scientific communication (Marres, 2015; Bruns et al., 2017; Willaert et al., 2020).

These strands of work have done an impressive job of measuring and mapping different aspects of public opinion division, but more work is required to understand how redistributive issues are framed and argued over in public cyberspace. This is particularly important because redistributive debates rarely reduce to a single ideological axis. They are often issue-specific, morally loaded, and tied to concrete policy designs. Rather than treating online disagreement as evidence of a stable attitudinal cleavage, this paper focuses on attributional divides: who or what people hold responsible when they argue over redistribution, and how those attributional patterns vary across issues and policy

moments.

In public expressions around redistributive issues, people may rely on varied cognitive frameworks, and the resulting disagreements are often multidimensional. Studies on social policy and public welfare suggest that people may interpret welfare programmes in relation to the benefits they themselves might get, linking preferences to self-interest, risk, or even personality (Hasenfeld and Rafferty, 1989; Shapiro and Young, 1989; Taylor-Gooby, 1982; Lü, 2014; Johnston et al., 2015; Mengel and Weidenholzer, 2023). There are also ideological models through which people understand redistribution, for instance conventional conservative versus liberal positions, leading to more familiar pro-/anti-welfare divisions and tight connections to electoral politics (Janowitz and Segal, 1967; Hunzaker and Valentino, 2019). Political notions such as socialism, liberalism, and capitalism may also affect where people stand on related issues such as migration and entitlement. People may also resort to moral values such as fairness, justice, or deservingness when agreeing or disagreeing over allocations of benefits and calls for redistribution (Hasenfeld and Rafferty, 1989; Van Oorschot, 2006). In other words, the frameworks people use to interpret redistributive issues are complex and overlapping. Even on the same issue, people may react differently when it is presented under different conditions or from different angles (Phelan et al., 1997; Harell et al., 2016). These variations may be magnified in online arenas, where people encounter more diverse viewpoints, discourses, and platform effects (Bakshy et al., 2015; Levendusky and Malhotra, 2016; Sîrbu et al., 2019). It is therefore reasonable to expect that online expressions around redistributive issues will appear in an intricate and issue-dependent form.

This paper investigates the questions of how do redistributive issues get framed and attributed in politically engaged online discourse, and how do those attributional patterns vary across issue domains and policy events? To answer these questions, we map redistributive discourse in the UK using digital data from cyberspace and unpack how different issues are organised by distinct patterns of responsibility, evaluation, and disagreement. The paper follows work on issue mapping, online controversy, and digital public debate (Yasseri et al., 2012; Marres, 2015; Willaert et al., 2020; Gaisbauer et al., 2021), but shifts attention from broad claims about attitudinal cleavage to the more specific structure of attributional divides. Empirically, it combines topic modelling, LLM-based stance and attribution annotation, and event-linked analysis of temporal variation. It contributes a computationally informed account of how redistributive discourse is organised in a politically engaged online arena, and offers an interpretive implication for the changing legitimacy and moral language of redistribution in the UK.

2. Redistributive issues in the UK: the tensions in the Beveridge dream

Redistribution is never not sustained by tax instruments and benefit rules alone: it also depends on whether citizens continue to regard social protection as reciprocal, fair, and collectively useful. A substantial literature shows that support for welfare and redistribution is shaped by institutional context, deservingness judgments, perceptions of government quality, and exposure to economic risk, rather than by income alone (Larsen, 2008; van Oorschot, 2006; Svallfors, 2013; Rehm et al., 2012,?). What makes the recent UK case especially important is that several of the foundations that make redistribution politically workable have come under pressure at the same time. Labour-market insecurity has widened, the fiscal demands on the state have intensified, and public debates increasingly frame welfare through the language of burden, contribution, and unfairness rather than shared protection. Recent developments in the UK are therefore better understood not primarily as the fading legacy of an earlier welfare philosophy, but as evidence of a broader dis-alignment between fiscal capacity, labour-market structure, and a widely shared commitment to solidaristic redistribution.

Research on welfare attitudes has long shown that solidaristic support is strongest where social risks are widely recognised as common and where recipients are not sharply separated from contributors in moral terms (Larsen, 2008). Once those boundaries harden, support becomes more conditional. This is why deservingness remains so central in comparative welfare research: citizens do not simply ask whether redistribution is affordable, but whether those who receive support are sufficiently similar, needy, and deserving to warrant it (van Oorschot, 2006). In the UK, this problem has become more acute under conditions of rising living costs and prolonged economic strain. That pressure does not produce a single attitudinal response but a more complex picture. Blekesaune (2013) shows that widespread economic strain can increase support for redistribution, but more recent work suggests that this effect is easily offset where fiscal scarcity, political distrust, or claimant differentiation remain strong. Rehm et al. (2012) likewise show that exposure to risk can generate support for welfare, while leaving room for divisions based on how risks and burdens are socially interpreted. The point, then, is not that hardship automatically destroys solidarity, but that solidarity becomes more fragile when hardship is filtered through narratives of competition, blame, and selective deservingness.

Recent UK evidence points in precisely that direction. For instance, de Vries et al. (2025) find that the pandemic did not produce a broad and lasting increase in support for working-age unemployment benefits in the UK; rather, claimants who lost work specifically because of Covid were treated as a special case, leaving broader welfare attitudes

largely unchanged. This suggests that even a major collective shock did not restore a generalised sense of shared risk. Alongside this is the persistence of claimant stigma. Baumberg (2016) shows that stigma remains substantial in Britain and can deter claiming, while Patrick (2017) and Patrick and Simpson (2020) show that welfare institutions are often experienced in ways that diminish dignity and weaken social citizenship. Fletcher and Flint (2018) and Fletcher and Wright (2018) similarly document the punitive logic of conditionality and sanctions, showing how welfare governance can recast claimants as objects of surveillance and deterrence rather than members of a shared community of risk. In other words, the erosion of redistribution's social foundations is not simply about declining generosity, but about the growing difficulty of sustaining a moral economy in which taxpayers and claimants still recognise themselves as part of the same reciprocal system.

Alongside the erosion of the social foundations of redistribution, the changing legitimacy of a tax-funded welfare system is also in tension. The issue is not only whether the state can raise enough revenue, but whether it can do so in ways that remain politically intelligible and normatively acceptable. Work on attitudes to taxes and social spending suggests that willingness to support public spending rises where government is perceived as efficient and fair, and where citizens believe public institutions can transform taxation into legitimate collective outcomes (Edlund, 2006; Svallfors, 2013). Conversely, when institutional trust is weak, tax burdens are more likely to be experienced as extraction rather than insurance. This is especially relevant in the contemporary UK, where effective tax burdens have risen while the visible quality, accessibility, and universality of social protection have often appeared more contested. Garritzmann et al. (2023) show that support for welfare reform depends in part on political trust and government satisfaction when reforms involve clear fiscal trade-offs. Public's attitudes toward progressive taxation are shaped by the wider structure of the welfare state itself, and tax legitimacy and welfare design cannot be separated analytically (Berens and Gelepithis, 2019). Citizens are more likely to accept taxation when they see the system as coherent, reciprocal, and broadly shared; they are less likely to do so when taxes remain high but protection appears fragmented or selectively distributed.

This tension is sharpened by changes in labour-market structure. Redistribution has historically been easier to legitimate where employment trajectories are relatively stable and where large parts of the population can imagine themselves moving through the same institutions over the life course. That condition has weakened. Research on post-industrial welfare states has long argued that labour-market change creates new pressures on both the financing and the social legitimacy of welfare systems (Jæger and

Kvist, 2003). More recent work on the knowledge economy similarly shows that class structure and social-policy preferences are being reorganised by new forms of labour-market vulnerability and dualisation (Häusermann et al., 2022). In the UK context, precarious work, long-term sickness, weak wage growth, and uneven employment security all complicate the older expectation that most citizens contribute regularly and benefit predictably. Once labour-market participation becomes more fragmented, redistribution is more easily reframed as a transfer between distinct groups rather than as mutual insurance across a shared life course, which makes the legitimacy of tax-funded welfare more contingent and politically unstable.

These social and fiscal tensions help explain why the empirical focus of this paper is not simply whether online discourse is positive or negative toward redistribution, but how responsibility/attribution is framed within that discourse. If redistribution is increasingly argued over through questions of burden, reciprocity, misdesign, and selective deservingness, then attribution becomes a useful entry point for studying how people interpret redistributive issues in practice. We do not claim to measure the full state of welfare legitimacy in the UK directly, rathertreat online responsibility claims as one observable arena in which these broader tensions are expressed, translated, and contested in public opinion.

3. Research design, data and methods

3.1. Data collection

We collected data from subreddit forum of 'UKPolitics', a moderated political discussion subreddit with approximately two million subscribers and a reputation for sustained engagement with domestic policy. The choice of platform is deliberate: Reddit's nested comment structure allows analysis not only of stated attitudes but of argumentative exchange and discursive elaboration, making it particularly well suited to studying how claims and arguments are constructed and contested rather than merely tabulated (Gao et al., 2024; Shugars and Beauchamp, 2019). Data were collected via the Reddit API (PRAW) [1], using six search keywords selected to cover the principal redistributive policy domains in contemporary UK social policy: universal basic income, minimum wage, income tax, national insurance, council tax, and poverty. Posts were retrieved for the period January 2020 to October 2025 – a window spanning the COVID-19 emergency, its immediate aftermath, successive Budgets under four Prime Ministers, and the return of

[1] Registered and granted access before the platform changed the academic API policy in Oct 2025.

a Labour government in 2024. After filtering by date and removing automated bot posts and deleted content, the dataset comprises 767 posts, 10,713 first-level comments, and 48,133 nested replies. The analysis in this paper mainly focuses on first-level comments as the primary unit of response to the post content.

As an online subsample, Reddit data cannot be treated as representative of the UK population, and the subreddit of UKPolitics is known skews toward politically engaged, younger, and predominantly male users. This constrains the generalisability of findings to the broader public but does not undermine the analytical value of the corpus: the forum provides a high-volume, temporally continuous record of how redistributive issues are framed, attributed, and contested in a politically active online community, and the analytical interest lies in the structure and variation of those frames rather than their prevalence in the general population. The choice of time frame allows us to observe discursive responses to an unusually dense sequence of redistributive policy actions: from COVID-era welfare expansion to post-pandemic fiscal consolidation and the first Labour Budget in fourteen years, generating substantial variation in political context against which attributional patterns can be read.

A hand-curated policy event dataset covering 53 events (policy or government schmes) across the six research domains (2020 to 2025) was assembled as the basis for event analysis. Events were identified from official parliamentary records (Hansard), HM Treasury and DWP press releases, Budget documents, and secondary reporting in mainstream media such as BBC. For each event, the dataset records the policy name, public announcement date, current legislative status, the primary initiating actor and their party affiliation, and two analytical codings: Beveridge alignment and an explanatory note on the event's connection with the Beveridge welfare state tradition to capture the nature of the issued policies that supports further analysis.

Beveridge alignment was coded by the author on a five-point scale: "Strong" (directly implements the Beveridge social insurance model's universalist or collective-risk principles, e.g., the Coronavirus Job Retention Scheme, which provided state-backed income protection against mass unemployment, a direct analogue to Beveridge's anti-'Idleness' goal); "Aligned" (consistent with the Beveridge anti-Want purpose but delivered through means-tested rather than universal channels, e.g., the £20-per-week Universal Credit uplift); "Partial" (addresses Beveridge poverty-prevention goals through mechanisms that depart from the original vision of nationally uniform entitlements, such as discretionary, time-limited, or locally-administered support, a category that characterises most council tax relief and Household Support Fund extensions); "Mixed" (combines Beveridge-compatible and Beveridge-contradictory elements, as with the 2021 Health and Social

Care Levy, which partially revived earmarked contributory financing but blurred the insurance-tax boundary and proved politically unstable); and "Eroding/Tension" (actively moves against Beveridge's universalist or contributory foundations, as with the 2024 Winter Fuel Payment restriction that replaced a near-universal age-based transfer with a means-tested benefit, or the Conservative government's 2024 ambition to abolish employee National Insurance, which would dissolve the contributory link at the heart of the original Beveridge bargain).

The Beveridge alignment variable is used here as an analytical coding scheme rather than a direct measure of public understanding. It classifies the policy events in relation to Beveridge's universalist and contributory benchmark and provides a structured way to interpret whether event-linked shifts in online attribution occur more often around policies that sustain, partially modify, or erode that benchmark. In other words, the coding helps connect the paper's theoretical discussion of welfare legitimacy to the event analysis follows, while remaining an interpretive device rather than a claim that commenters themselves explicitly invoke Beveridgean principles.

3.2. Methods

To identify the principal thematic clusters within the corpus, we applied BERTopic (Grootendorst, 2022), a framework combining sentence-transformer embeddings, UMAP dimensionality reduction, and HDBSCAN density clustering with a KeyBERT-inspired representation model, to the full comment corpus. In the pre-processing we removed URLs and archive-reference strings while preserving the natural linguistic texture of comments, including colloquial terms, political shorthand, and intertextual references to media coverage. Sentence embeddings were generated using all-MiniLM-L6-v2 with bigram representation. The resulting topics were manually reviewed and consolidated into five substantive discourse groups based on leading keywords and representative documents: tax policy and fiscal debate; welfare and minimum income; political parties and leadership; inequality and class; and social and identity issues. The topic modelling stage maps the discursive landscape within which divides emerge; it is therefore used to identify the main issue domains before the following LLM-annotated stance and attribution analysis. It also produces the document-topic assignments that describe how redistributive discourse is distributed across themes and over time.

The main variables of stance classification and causal attribution extraction for the comment text were performed by LLMs applying a structured prompt to each post-comment pair. This prompt-based approach was chosen over supervised classifiers because the analytical targets, particularly the free-text attribution summary and its ty-

pological classification, require contextual reasoning that is not reducible to pattern-matching on surface features of the comment alone. Each annotation task required the model to read both the post title and the comment body and infer the commenter's evaluative stance toward the post's central claim (approve/disapprove/neutral), articulate the reasoning behind that stance, identify who or what the comment assigns causal or moral responsibility to, and classify that responsibility target into a fixed typology. The prompt specified that all outputs were to be returned in a line-by-line key-value format to maximise robustness across model families.

The attribution target typology was developed inductively from a pilot reading of 280 manually coded training cases and refined iteratively to maximise coverage and minimise ambiguity. The nine categories are: government (state institutions, ministerial decisions, government policy at any level); political parties (specific party actions or inactions, partisan politics, electoral behaviour); economy (market forces, structural economic conditions, macro-level constraints); institutions (non-partisan organisations such as the NHS, DWP, judiciary, local authorities, or supranational bodies); individuals (personal choices, individual behaviour, claimant deservingness, moral responsibility of persons); employers (business decisions, employer wage practices, hiring and firing behaviour); migrants (migration flows, immigration policy consequences, scapegoating of foreign nationals); other (specify) (for clearly identifiable targets outside the eight named categories); and none (for comments where no causal agent is identifiable, typically descriptive, rhetorical, or purely evaluative).

Three large language models were selected based on the accessibility and the relative performance: Google Gemma-3-27B-Instruct, Meta Llama-3.3-70B-Instruct, and DeepSeek-R1-Distill-Qwen-32B, and run via high-performance computing. The three models produced valid outputs for 9,920 common comment IDs. A majority-vote ensemble was constructed over stance and attribution target labels; in the event of a three-way tie, Gemma-3-27B took precedence, as this model produced the most balanced label distributions across all categories and more reasonable attribution summaries compare to the manual coded results. Full three-model agreement reached 60.4% on stance and 43.3% on attribution target, reflecting genuine interpretive ambiguity in political text rather than model failure. To assess annotation quality, ensemble outputs were compared against 280 manually coded gold-standard cases; substantive coherence in argument clusters and attribution categories confirmed the fitness for analytical purpose.

We then conduct the event analysis to connect online responsibility claims with real-world policy schemes. Each comment was matched to its most relevant policy event using a blended similarity score (40% TF-IDF cosine similarity, 60% SentenceBERT co-

sine similarity), applied within a 90-day window restricted to the same thematic category, and with a minimum score threshold of 0.25 to retain only comments with meaningful policy-specific lexical or semantic overlap (see Appendix Table A1 for illustrative examples of matched comments at different score bands). The weighted blend was chosen to capture both lexical overlap (policy-specific terminology) and semantic proximity (comments discussing policy effects in different vocabulary). This procedure matched 4,913 of 9,920 comments (49.5%) to 38 distinct policies. We then compute the quarterly attribution target distributions for each theme. Pre/post distributional shifts around the five highest-frequency policy events were tested using Pearson chi-squared tests on attribution target frequencies within 90-day windows.

To further illustrate how the broader corpus-level patterns appear within concrete exchanges, we visualise the conversation trees of two selected cases. Reply chains within the two high-profile threads were additionally annotated: a national insurance thread (post ID pk5i3o; top 8 comments by vote, 29 qualifying replies) and a UBI thread (post ID g7zste; top 8 comments, 22 replies). Annotation followed the same structured prompt but applied locally to each reply-parent pair using the Gemma-3-12B model via Ollama. Results are rendered as three-layer interactive conversation trees to examine how stances and attributions propagate through nested argumentative exchanges, offering a micro-structural complement to the corpus-level analysis.

4. The Structure of Redistributive Discourse

As seen from Figure 1, BERTopic modelling of the first-layer comments recovered 29 substantive discourse topics alongside a residual outlier cluster of 3,966 documents (39.5%). When aggregated into discourse groups, tax policy and fiscal debate accounts for the largest single bloc after 2021. Welfare and minimum income claims form the second substantive cluster after 2021, followed by political parties and leadership, inequality and class, and social and identity issues. These proportions are broadly consistent across the 2022 to 2025 period as seen in Figure 2, though political party discourse intensifies around the Truss administration in autumn 2022 and the 2024 general election cycle. Rather than indicating a single redistributive divide, the distribution of topics suggests that different issue domains organise discussion through different vocabularies and problem frames.

The linguistic texture of discourse varies substantially by theme. Tax policy discussions are characterised by a dense vocabulary of fiscal process: property valuation bands, council revaluation, employer pass-through, threshold effects, and the distinction between

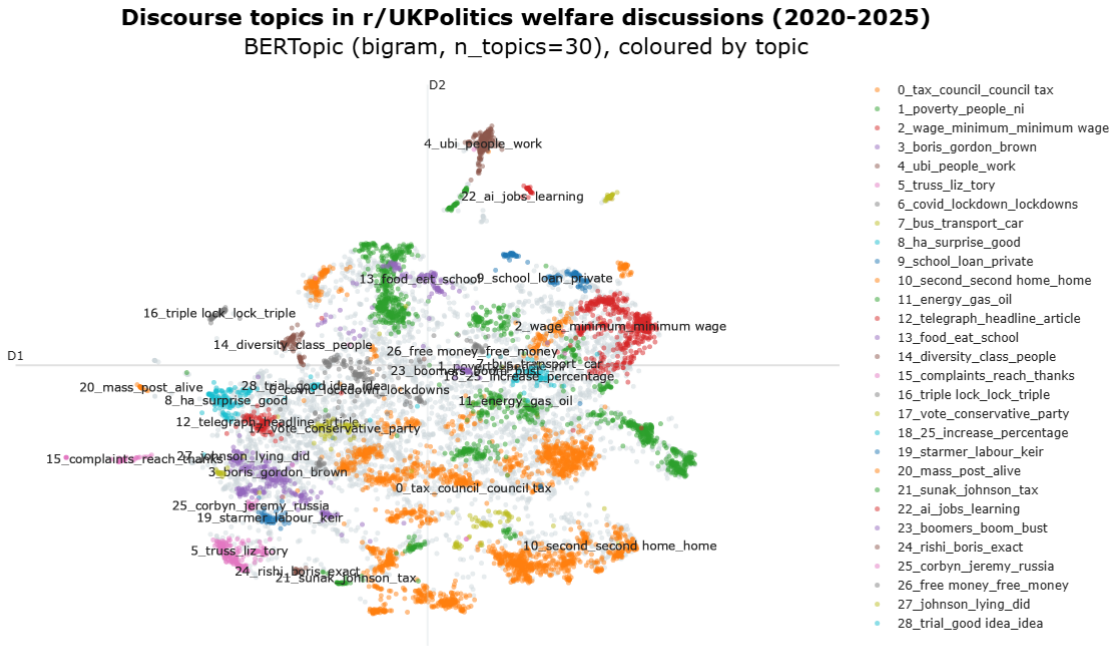

Figure 1: Bertopic by document clusters

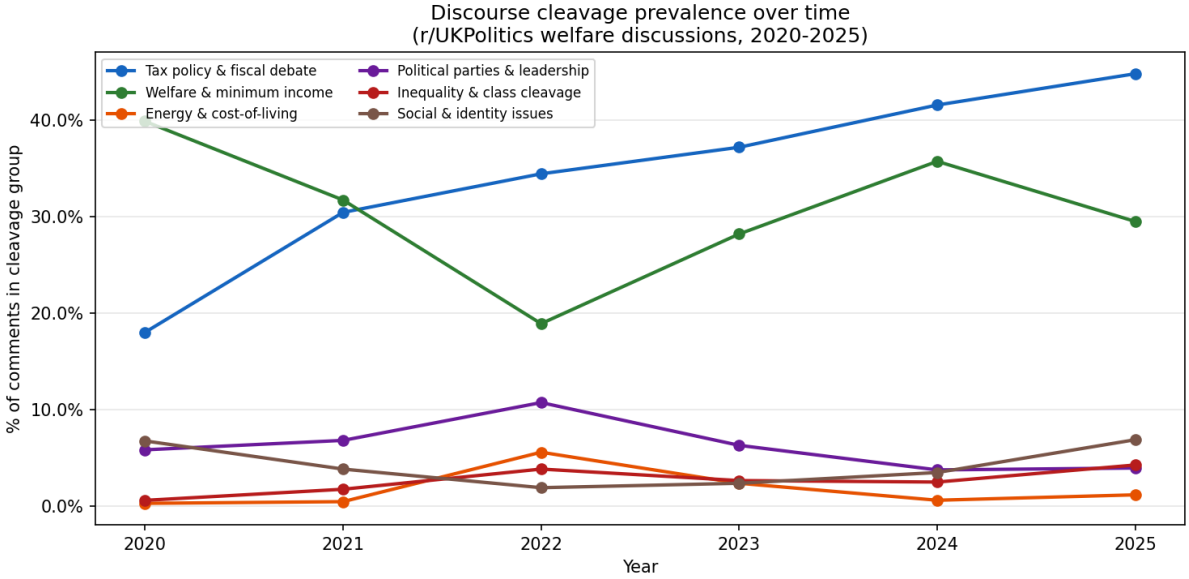

Figure 2: Cleavage group prevalence by year (2020-2025)

direct and indirect taxation. Vocabulary items such as *hypothecation*, *ring-fenced*, *burden*, and *incidence* appear at significantly elevated frequency compared with the overall corpus. Welfare and minimum income discussions, by contrast, foreground relational and evaluative language: *deserve*, *struggling*, *dignity*, *scrounger*, and *in work poverty*, and are more likely to invoke specific demographic categories (single parents, pensioners, young people, disabled people). Political party discourse is dominated by named actor references: Sunak, Truss, Reeves, Tories, Labour, with evaluative adjectives (*incompetent*, *dishonest*, *cowardly*) as secondary features. These inter-topic linguistic differences suggest that different redistributive domains activate qualitatively distinct modes of reasoning and evaluation, a finding consistent with evidence that issue-specific frames, not a single pro/anti redistribution dimension, organise contemporary welfare attitudes (Mengel and Weidenholzer, 2023).

The overall stance/disagreement distribution across the 9,920 ensemble-labelled comments shows a clear majority of disapproving positions (59.6%), with neutral comments second (18.3%) and approval accounting for 22.0%. This aggregate negativity is consistent with the general character of political discussion on Reddit, where critical and sceptical voices predominate. It also suggests that discussion of redistributive policy on r/UKPolitics is frequently organised through disagreement with the post's central claim, rather than straightforward endorsement. Because stance is measured relative to the framing of the post rather than to redistribution in the abstract, this pattern should be read as evidence of a confrontational and evaluative discursive style, not as a direct measure of anti-redistributive preferences.

Table 1: Comment stance distribution by policy theme (%, ensemble majority vote, $n = 9{,}920$)

Policy theme	**Approve**	**Disapprove**	**Neutral**
Council Tax	23.3	54.9	21.8
Income Tax	16.2	67.0	16.8
Minimum Wage	20.8	61.5	17.7
National Insurance	17.5	64.9	17.6
Poverty	26.4	54.5	19.1
Universal Basic Income	24.3	62.5	13.2

Note: Column percentages within each theme. *Approve* = commenter agrees with the post's central claim; *Disapprove* = commenter disagrees; *Neutral* = no clear evaluative position. Stance labels assigned by majority vote across three large language models; full three-model agreement rate = 60.4%.

The cross-theme pattern in Table 1 reveals variations across themes though: National insurance and income tax show the highest disapproval rates, consistent with both themes

being dominated by discussion of contested government interventions (NI levy, threshold freezes, employer NI rises). UBI discussions also reflecting the ideological breadth of the debate which is consistent with UBI occupying a genuinely contested ideational position rather than a taken-for-granted redistributive instrument. Poverty discussions exhibit the highest neutral rate, plausibly because many poverty-related posts are structured as news reports or statistical releases rather than normative claims, inviting commentary that is more diagnostic than evaluative.

5. Patterns of Causal Attribution

The LLM attribution analysis reveals a discourse organised primarily around political agency. Government (27.5%) and political parties (24.2%) together account for over half of all attribution targets. Individuals (11.0%) and economy (9.3%) are the next most common substantive targets, while none – where no clear causal agent is identified – accounts for 16.6%. Tables 2 shows the cross-theme patterns of the attributions. Council tax discussions show the highest rate of government attribution (38.8%), consistent with the strongly institutionalised character of council tax, whose levels are set through legislative process and are not easily construed as market outcomes. The extensive localisation of council tax support since 2013 which replacing a national credit with a patchwork of locally designed relief schemes is reflected in a discourse that treats government as responsible for fiscal design failures rather than attributing hardship to structural economic forces. National insurance shows a similar government-and-parties dominance (32.0% and 29.4% respectively) combined with notably low economy attribution (5.8%), suggesting that commentary constructs NI primarily as a political choice – a tax lever – rather than an economic constraint. This is consistent with Hills (2004) argument that the repeated use of NI as a general revenue instrument, detached from its contributory rationale, has undermined its public legibility as social insurance.

A typical comment that disagrees with the post argument and attribute to the government role is like follows:

> For anyone with a brain this will really stick in the throat – boomers not affected, companies not affected. Hard to see why ordinary folk should pay more when the government has/is wasting and stealing billions.
>
> – r/UKPolitics comment on National Insurance rise, 2021

The grievance is framed in terms of distributive incidence, *who pays*, rather than the principle of contributory financing. The Beveridgean bargain of *benefit in return*

for contributions (Beveridge, 1942) is not invoked positively; rather, the contributory logic is implicitly inverted, with the commenter noting that the cost falls on "ordinary folk" rather than those who benefit from public provision. Meanwhile, institutional waste and corruption are deployed as delegitimating frames, consistent with a broader pattern in UK welfare discourse in which public trust in government spending is low and fiscal disengagement is rationalised through anti-waste arguments (Baumberg, 2016). A similar comment of straightforwardly partisan and directed at state institutions reads as this: "Personally, I neither asked for nor wanted the last two NI cuts Sunak gave me. I'd much rather have public services that actually work." Here the redistributive object (higher NI contributions) is endorsed only as a means to service quality, not as a contribution-based social insurance entitlement.

Table 2: Attribution target distribution by policy theme (%, ensemble majority vote, $n = 9{,}920$)

Policy theme	Gov.	Pol. parties	Economy	Individ.	Employers	Instit.	Migrants	Other	None
Council Tax	38.8	21.7	4.6	12.3	1.9	2.8	0.6	1.2	16.1
Income Tax	36.6	25.3	9.3	10.4	1.4	2.9	0.4	0.6	13.1
Minimum Wage	15.3	14.4	16.3	10.2	22.0	4.4	0.8	0.8	15.7
National Insurance	32.0	29.4	5.8	6.7	6.5	3.8	0.2	0.6	15.0
Poverty	22.6	26.4	9.0	13.1	2.0	4.2	1.1	1.9	19.7
Universal Basic Income	19.5	22.9	16.2	11.5	3.4	6.6	0.7	2.8	16.2

Note: Row percentages; each row sums to 100% within its theme. Attribution target categories: *Gov.* = government (state institutions, ministerial decisions); *Pol. parties* = political parties (partisan actions, electoral behaviour); *Economy* = market forces, structural economic conditions; *Individ.* = individual choices, personal responsibility, deservingness; *Employers* = business decisions, employer wage/hiring practices; *Instit.* = non-partisan institutions (NHS, DWP, local authorities, courts); *Migrants* = migration flows, immigration effects; *None* = no explicit causal agent identifiable. Target labels assigned by majority vote; full three-model agreement rate on all five 9 categories = 43.3%.

Minimum wage discussions diverge sharply from the tax policy pattern. Employers are the single most prominent attribution target (22.0%), followed by economy (16.3%), with government falling to just 15.3%. This profile likely reflects the institutional locus of minimum wage policy: though uprating decisions are formally governmental, the lived consequences are often discussed through employer behaviour: compliance, hours reduction, price pass-through, and hiring decisions, which becomes a central object of debate. The discurse shifts accordingly: vocabulary items such as *CEO bonus*, *sweatshop*, *zero-hours*, and *exploit* appear at elevated frequency in minimum wage threads, indexing an attribution frame centred on employer power and labour-market exploitation. As one user commented: "Good. If 'growth' is basically a bunch of sweatshops and greedy CEOs then what's the point for 99% of us?".

Poverty discussions show the most diffuse attribution structure, with political parties (26.4%) now slightly leading government (22.6%), alongside a higher rate of individual attribution (13.1%) than in most other categories. This is consistent with a large body of evidence on the persistence of deservingness reasoning in UK poverty discourse: the tendency to assess claimant entitlement according to normative criteria of need, control, reciprocity, and identity (van Oorschot, 2006; Baumberg, 2016; Summers et al., 2025). Even amid the COVID-19 pandemic, a period when solidarity rhetoric intensified and first-time benefit claimants brought new groups into the social security system, deservingness frames showed considerable durability: individuals continued to be attributed responsibility for their poverty, and claimants continued to negotiate their own entitlement through appeals to genuine need, lack of control, and prior contribution (Summers et al., 2025). The Reddit corpus mirrors this pattern: even in threads triggered by statistical reports of rising child poverty, individual attribution persists alongside institutional and governmental frames, and sometimes predominates.

> Unfortunately it's a tried and tested strategy to treat the poor badly then blame people even worse off than they are for their misfortune. They fall for it every time.
>
> – r/UKPolitics comment on poverty discussion thread, 2022

The quoted comment here is particularly instructive as it performs a meta-commentary on the attribution structure itself: the user does not simply attribute poverty to government, but criticises the discursive strategy of displacing responsibility downward. It echoes what Larsen (2008) terms the construction of an *undeserving poor* as a political resource, and Svallfors (2013)'s argument that stratified welfare states generate stratified

solidarities, where different social groups are mapped onto different eligibility categories with different moral valuations. That this critique of attribution itself appears organically in the corpus confirms that deservingness is not merely a latent dimension of welfare attitudes but an actively contested discursive object.

UBI discussions are characterised by elevated economy attribution (16.2%) and higher institutional attribution (6.6%) than other themes, reflecting the macroeconomic and structural factor in which UBI arguments are typically conducted, concerns about automation, labour-market restructuring, inflationary effects, and the fiscal architecture of a basic income scheme. Attribution modelling of the llm labelled comment attribution (9,026 attributions, 11 topics, shown in Figure A6 in the online appendix) confirms this: the UBI-specific BERTopic cluster is characterised by terms such as *potential*, *implementation*, *existing*, *financial constraints*, and *DWP redundant* – a vocabulary of systemic hypothetical reasoning rather than political accountability. The pensioner and intergenerational topic further highlights a cross-cutting divide around the triple lock, with attribution directed at age-specific political economy (the voting power of older cohorts, perceived disproportionate protection of pension income) rather than any single institutional actor.

Figure 3 here presents the sankey diagram that visualises the full stance-attribution flow across 9,920 annotated comments. The two widest channels: disapprove → government (n = 1,769) and disapprove → political parties (n = 1,746) confirm that the discourse is principally organised as political accountability claims, with rejection of the post's position and attribution to the state or party system as the dominant co-occurring pattern. A substantial neutral → none flow (n = 1,058) captures descriptive or rhetorical commentary with no clear evaluative or causal structure, concentrated in poverty and council tax threads. Approval flows mirror the same attribution targets but at considerably smaller volume, indicating that political agency is invoked regardless of evaluative direction. Theme-specific divergences visible in the Sankey align closely with the cross-tabulation in Table 2: the minimum wage chain uniquely elevates disapprove → employers as its leading flow; poverty reverses the typical tax-category pattern by leading with disapprove → political parties rather than disapprove → government; and UBI shows the most dispersed disapproval structure, with no single dominant target, reflecting its position as an ideologically open rather than institutionally settled debate.

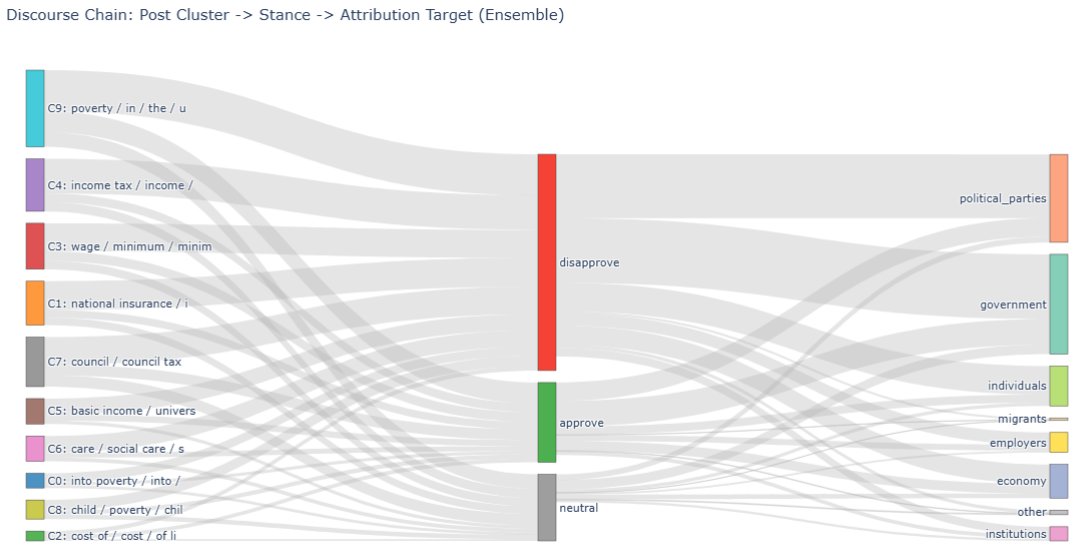

Figure 3: Sankey diagram

6. Event-Driven Variation: Discourse in Response to Policy

6.1. Quarterly Attribution Dynamics

In Figures 4-9, we present the quarterly attribution timelines across all six themes, and they show that the dominance of government and political party attribution is not static but responds, often sharply, to specific policy events. For discussions around national insurance, political party attribution peaks sharply in Q4 2021 following the Health and Social Care Levy announcement, a policy that was politically charged precisely because it deployed the NI mechanism for a purpose (funding social care) outside its traditional contributory rationale, while extending it to pensioners in employment and income, generating accusations of partisan asymmetry. As the levy moved toward implementation and eventual repeal under Truss, party attribution partially recedes and government attribution rises, consistent with a pattern in which attribution becomes more institutionally specific as a controversial policy is legislated, then its reversal is located in executive government rather than partisan agency.

Council tax timelines show a more stable attribution structure with a consistently high government target rate, punctuated by a brief elevation of political party attribution around the 2022-23 council tax rebate and the 2024-25 referendum principles. The relative stability reflects the long-established institutional character of council tax: it has been in continuous operation since 1993, with valuation bands frozen since 1991, and users

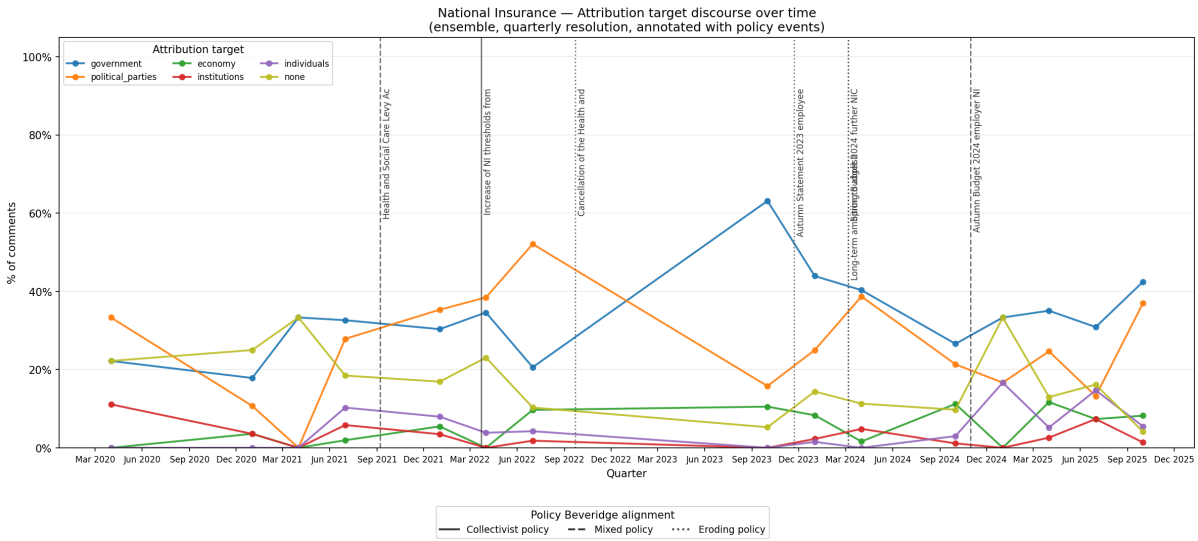

Figure 4: Quarterly attribution target distribution for NI-related discussions (2020-2025)

treat its reform — or failure to reform – as a persistent governance failure requiring political rather than economic explanation. Income tax discussions show a parallel rise in government attribution from 2022 onwards, partly driven by the extended threshold freeze, which functions as a stealth tax but is consistently attributed to government design rather than economic necessity.

For poverty[2], government and political party attribution stays at higher band and coinciding with overlapping crises: post-pandemic benefit retrenchment (the end of the £20 UC uplift in October 2021), energy price shocks, successive benefit freezes, and the announcement of welfare cuts in early 2025. The discourse trajectory is consistent with an institutionalisation of blame: as poverty becomes a sustained news story rather than an acute emergency, attributional responsibility concentrates on government/party as the primary accountable agent rather than fully dispersing across market forces and individual choices.

6.2. Pre/Post Policy Shifts

To further examine the pre/post policy shifts in the debates, we run chi-squared tests for distributional shifts in attribution targets around the a few highest-frequency policy

[2]The initial collection using Reddit's search API with sort='new' retrieved at most approximately 1,000 results per query, exhausting the cap at November 2021 for the high-volume term "poverty". We filled this gap by re-querying across all four available sort orders (new, top, relevance, comments), which access independent search indices and recover posts from 2020-2021 missed by the original collection. Poverty therefore now covers the full 2020-2025 window, including the COVID welfare expansion period.

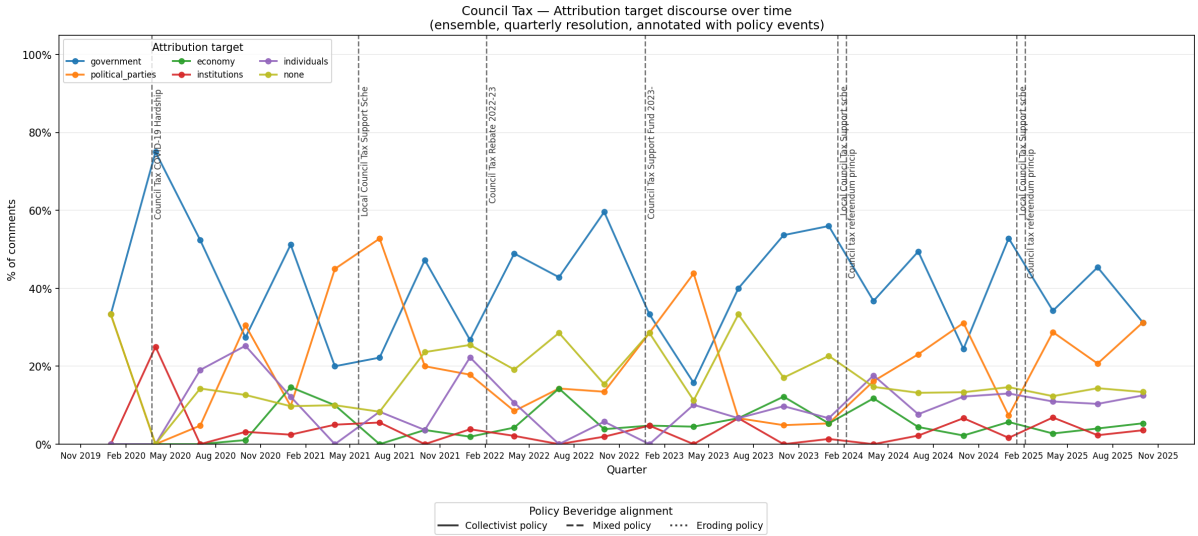

Figure 5: Quarterly attribution target distribution for Council tax-related discussions (2020-2025)

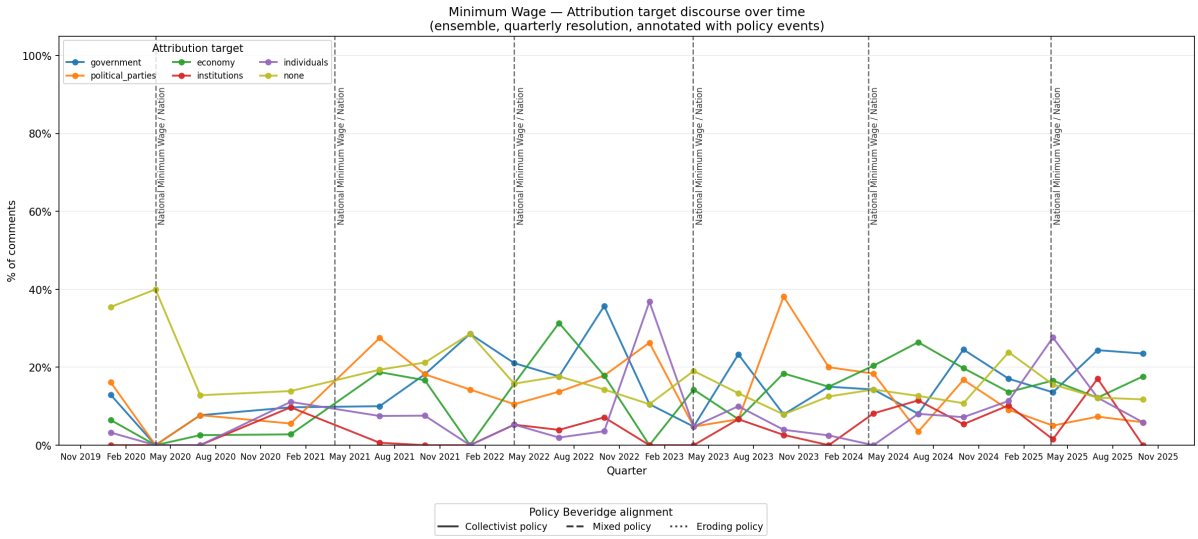

Figure 6: Quarterly attribution target distribution for Minimum wage-related discussions (2020-2025)

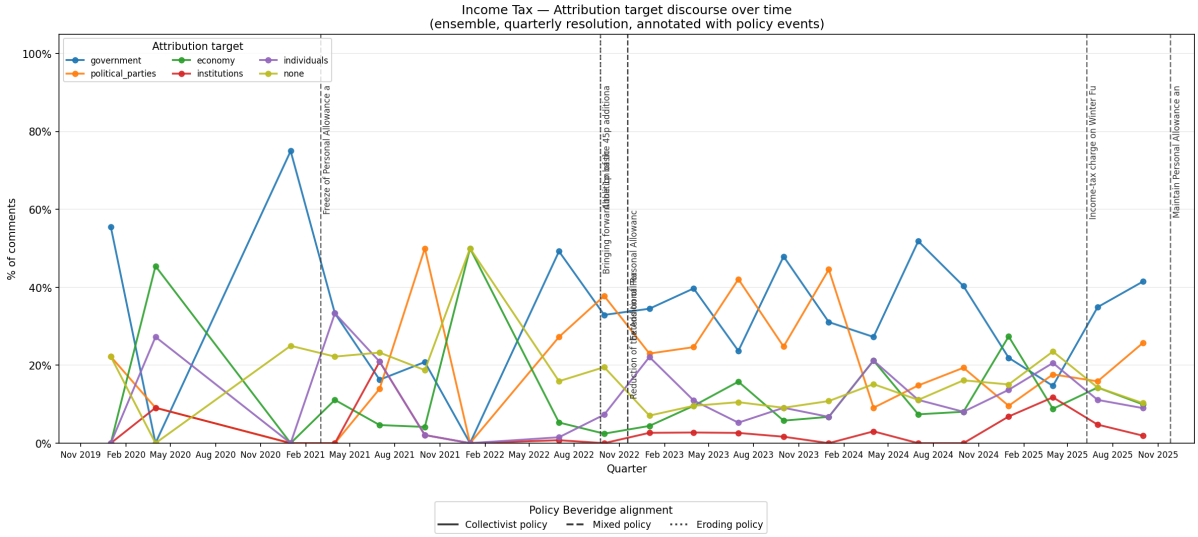

Figure 7: Quarterly attribution target distribution for income tax-related discussions (2020-2025)

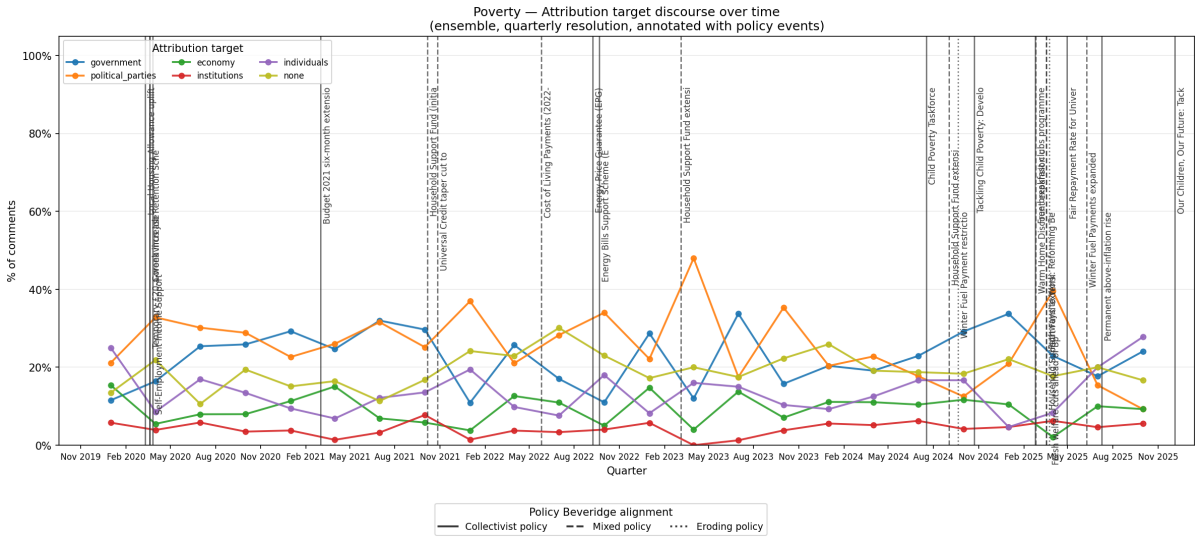

Figure 8: Quarterly attribution target distribution for poverty-related discussions (2020-2025)

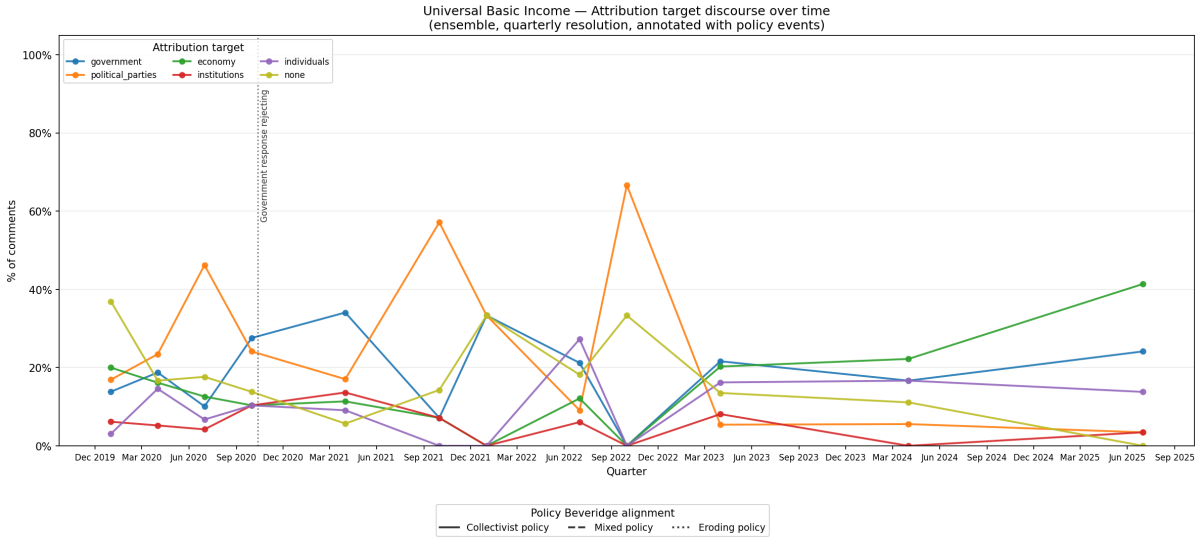

Figure 9: Quarterly attribution target distribution for UBI-related discussions (2020-2025)

events, identifying three statistically significant transitions. The Health and Social Care Levy Act 2021 is associated with a pre-to-post shift in which political parties move from being the dominant attribution target to government. This transition is consistent with the policy trajectory: prior to announcement the levy was embedded in partisan debate (a Conservative proposal, contested by Labour on distributional grounds); once legislated, commentary relocated responsibility from the party proposing it to the state machinery implementing it. The shift also reflects the ambivalence of the policy's Beveridge alignment: the levy nominally revived earmarked contributory financing but blurred National Insurance with general taxation in a way that commenters read as deceptive fiscal design rather than as a legitimate contributory bargain.

The Local Council Tax Support schemes for 2025-26 show a similar pattern: political parties are the dominant attribution target before the announcement, with government taking over as the leading target post-announcement. Council tax support has been administered as a patchwork of locally designed relief schemes since the national credit was abolished in 2013, and its periodic uprating and reform are embedded in a political rather than technical model. The pre-announcement dominance of party attribution reflects the politicised debate around local government funding settlements; the post-announcement shift to government attribution is consistent with the institutional ownership that attaches to formally decided support schemes, particularly as they interact with fiscal pressures on local authorities.

The 2025 National Living Wage uprating shows a distinctive pattern: unlike the

above transitions, where government becomes the post-policy dominant target, the minimum wage uprating shifts attention from structural economic conditions (economy) to employer-focused attribution. This is consistent with the broader finding that minimum wage discourse foregrounds employer behaviour – compliance, hours reduction, and hiring decisions – rather than state accountability. The policy announcement activates concern about how employers will absorb the new wage floor and pass costs through the employment relationship, rather than triggering institutional blame. This aligns with the dominant vocabulary of minimum wage threads, where *CEO bonus*, *zero-hours*, and *exploit* appear alongside uprating announcements rather than government-directed criticism.

Table 3: Statistically significant pre/post attribution target shifts around policy events (Pearson χ^2, ±90-day windows, $p < 0.05$)

Policy	**Theme**	**Policy Type**	N **pre**	N **post**	χ^2	p	**Top target (pre)**	**Top target (post)**
National Minimum Wage / National Living Wage uprating 2025	Minimum wage	*Mixed*	50	109	31.20	<0.001	economy	employers
Local Council Tax Support schemes for 2025-26	Council tax	*Mixed*	53	88	30.64	0.0002	political parties	government
Health and Social Care Levy Act 2021 / 1.25 percentage-point NIC rise	National insurance	*Mixed*	199	256	24.04	0.002	political parties	government

Note: Pre/post windows: ±90 days from public announcement date. Only comments matched at similarity score ≥ 0.25 are included; only policies with $n \geq 5$ in both windows are shown. Two further policies were tested but did not reach significance ($p > 0.05$). *Policy Type* column: *Mixed* = combines Beveridge-compatible elements with structural tensions or contradictions. χ^2 tests use Pearson statistic on attribution target frequency distributions.

Taken together, two of the three significant transitions involve a post-policy shift toward government attribution. This convergence suggests a possible broader pattern: in online redistributive discourse, once a policy crosses the threshold from political debate to legislative or executive decision, users may relocate responsibility from the partisan

domain to the institutional state, concentrating accountability more at the level of government than party. This is consistent with a cognitive-attribution model in which political contestation over a policy proposal activates partisan frames, while policy enactment activates institutional frames. The exception – minimum wage uprating shifting to employer attribution rather than government – is consistent with the structural character of that debate: when the policy consequence is mediated through employer behaviour rather than direct state action, accountability does not automatically flow back to the state. All three significant policies share a Mixed Beveridge alignment, combining elements compatible with the collectivist welfare model alongside structural tensions; none follows a straightforwardly Beveridgean or straightforwardly eroding design, which may partly explain why the attributional response is contested rather than uniform. This does not establish a general rule, but it is consistent with the possibility that policies occupying ambiguous Beveridgean territory are most likely to produce discernible pre-to-post shifts in how online commenters locate responsibility.

6.3. Conversational Dynamics of high-profile cases

This subsection presents two illustrative case studies and use tree analysis for two high-profile cases: a national insurance thread (8 top comments, 29 labelled second-layer replies) and a UBI thread (8 top comments, 22 second-layer replies)[3] to show how the broader corpus-level attributional patterns appear within nested exchanges. In the national insurance thread, first-level comments are relatively consolidated: disapproving stances directed at government and political parties predominate, and reply chains largely sustain this direction. Argumentative divergence is limited; where it occurs, it tends to take the form of qualificatory replies that introduce economic or institutional attributions ("the system was never designed to cope with this demographic"; "it's a problem with the underlying state pension model") rather than full reversals of the dominant evaluative frame.

The UBI thread presents a different pattern. First-level comments are genuinely heterogeneous where reply dynamics involve explicit argumentative engagement that introduces competing frames. Economy attribution appears in the reply layer in contention with the political attributions advanced at first-level:

> Now would have been a good time to test it. Are we just going to wait until everyone is unemployed from automation before we dabble with it?

[3]Theoretically we are able to present the full tree for the thread, however it's not easy to read or possible to interpret directly.

– r/UKPolitics comment on UBI thread, 2020

This comment quoted above illustrates the macro-structural characteristic of UBI discourse: the problem frame is not *government failed to do X* but *structural economic transformation requires policy adaptation*. Reply chains following this kind of economy-attributed first-level comment are more likely to involve explicit disagreement, introducing inflationary concerns, disincentive effects, or fiscal feasibility arguments than the largely reinforcing reply patterns found in national insurance threads. This micro-structural contrast aligns with broader evidence that structural alternatives to existing institutions generate more genuinely pluralised debate than incremental adjustments to established programmes, because the former requires interlocutors to reason from first principles about social provision rather than evaluate a specific governmental act (Berens and Gelepithis, 2019).

Taken as illustrative case studies, these conversational dynamics suggest that the degree of ideational consolidation in online welfare discourse varies with policy type: contested fiscal adjustments to established taxes generate more co-ordinated responses at the first-level comment stage, while discussions of structural alternatives generate more visible pluralism across the reply layer. This distinction supports, rather than independently proves, the broader corpus-level finding that different redistributive issues organise responsibility and disagreement in different ways. For established instruments like national insurance, the discourse is often characterised by a relatively coherent delegitimating narrative directed at government; for less institutionally settled instruments like UBI, the discourse is more openly contested and no single frame clearly dominates.

7. Discussions

This paper examines how redistributive issues are framed and attributed in politically engaged online discourse. Across the corpus, redistributive debates are primarily organised through political and fiscal responsibility claims: government and political parties are jointly identified as the responsible actors in over half of all annotated comments, while the thematic structure recovered by topic modelling does not conform to a simple pro- or anti-redistribution binary. Instead, the discourse appears fragmented across issue domains, with different themes activating different vocabularies, evaluative criteria, and attribution targets. Attribution patterns also vary around specific policy events, and several of the clearest shifts move from partisan attribution toward institutional government blame once a policy is announced or enacted.

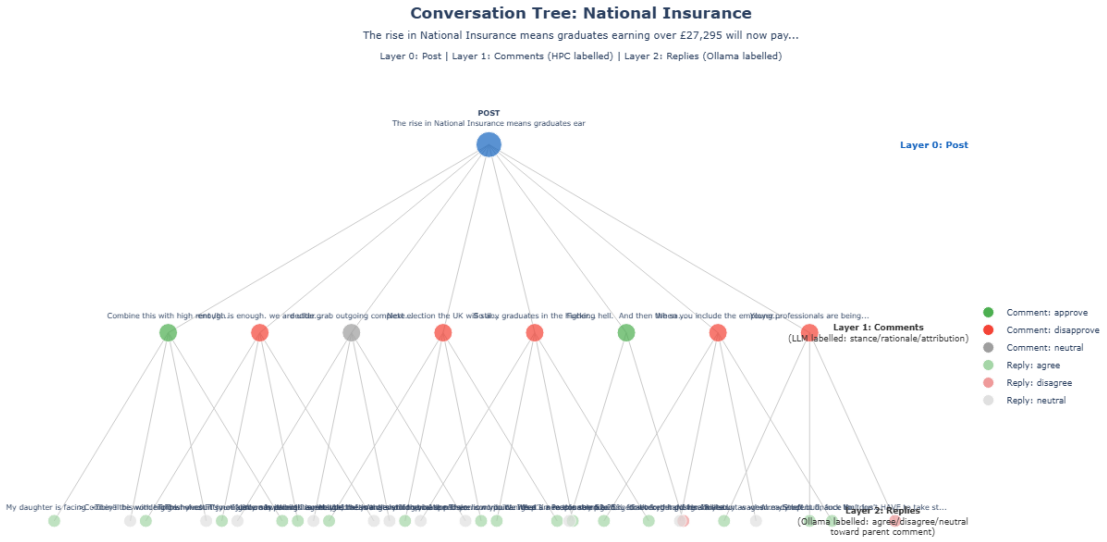

Figure 10: Three-layer conversation tree for national insurance thread (post pk5i3o)

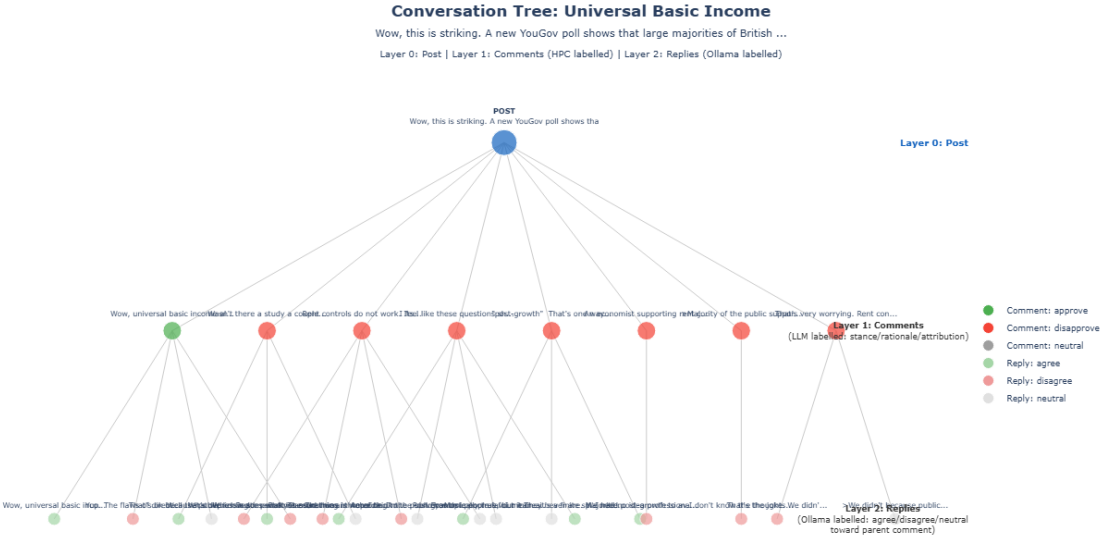

Figure 11: Three-layer conversation tree for universal basic income thread (post g7zste)

The paper's primary contribution is therefore a computationally informed account of how redistributive discourse is organised by attribution and issue-specific framing in a politically engaged online arena. The analysis shows that council tax, national insurance, minimum wage, poverty, income tax, and UBI debates each mobilise somewhat different responsibility structures and cognitive models. Tax-related discussions are especially dominated by government and party attribution, minimum wage debates more often foreground employers and labour-market practices, and UBI debates appear more dispersed and structurally oriented. These differences matter because they suggest that online redistributive discourse is organised not only by support or opposition, but by distinct ways of locating causation, blame, and political obligation.

In addition, the cross-theme variation in attribution, together with the event-linked shifts, is consistent with a broader argument in the social-policy literature: the legitimacy of redistribution is increasingly filtered through deservingness judgements, institutional distrust, and issue-specific framings that vary substantially across policy domains. Within that context, policies that weaken universal or contributory principles are associated with sharper movement toward government attribution. That pattern should be read with caution though: it does not show directly that online commenters consciously defend Beveridgean social insurance, nor does it establish the overall state of welfare legitimacy in Britain. But it does suggest that departures from recognisable universalist or contributory designs may be especially likely to attract concentrated institutional blame in online debate.

The conversation-tree case studies support this broader reading at the micro level. They illustrate that argumentative dynamics within sub-threads do not unfold identically across policy types. Fiscal redistributive debates around established instruments such as national insurance appear more consolidated at the initial comment stage, while debates over more structural alternatives such as UBI remain more openly contested across the reply layer.

There is still plenty of space for the investigation to develop further. The collected data contain metadata such as comment scores and anonymized user IDs, and future work could make better use of this information to examine whether responsibility claims differ across more and less highly rated comments, or whether some patterns are carried disproportionately by particular kinds of users. More broadly, the paper should be read as an initial account of attributional divides in online redistributive discourse, and as a foundation for future work on validation, comparative platform analysis, and the relationship between online discourse and wider public opinion.

References

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132.

Baldassarri, D. and Goldberg, A. (2014). Neither ideologues nor agnostics: Alternative voters' belief system in an age of partisan politics. *American Journal of Sociology*, 120(1):45–95.

Barker, D. C. and Tinnick, J. D. (2006). Competing visions of parental roles and ideological constraint. *American Political Science Review*, 100(2):249–263.

Baumberg, B. (2016). The stigma of claiming benefits: A quantitative study. *Journal of Social Policy*, 45(2):181–199.

Berens, S. and Gelepithis, M. (2019). Welfare state structure, inequality, and public attitudes towards progressive taxation. *Socio-Economic Review*, 17(4):823–850.

Beveridge, W. (1942). Social insurance and allied services.

Blekesaune, M. (2013). Economic strain and public support for redistribution: A comparative analysis of 28 european countries. *Journal of Social Policy*, 42(1):57–72.

Boutyline, A. and Vaisey, S. (2017). Belief network analysis: A relational approach to understanding the structure of attitudes. *American journal of sociology*, 122(5):1371–1447.

Bruns, A., Moon, B., Münch, F., and Sadkowsky, T. (2017). The australian twittersphere in 2016: Mapping the follower/followee network. *Social Media+ Society*, 3(4):2056305117748162.

Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical review*, 18(1-3):1–74.

de Vries, R., Geiger, B. B., Scullion, L., Summers, K., Edmiston, D., Ingold, J., Robertshaw, D., and Young, D. (2025). Welfare attitudes in a crisis: How COVID exceptionalism undermined greater solidarity. *Journal of Social Policy*, 54(3):714–733.

Edlund, J. (2006). Trust in the capability of the welfare state and general welfare state support: Sweden 1997–2002. *Acta Sociologica*, 49(4):395–417.

Feinberg, M. and Willer, R. (2013). The moral roots of environmental attitudes. *Psychological science*, 24(1):56–62.

Fletcher, D. R. and Flint, J. (2018). Welfare conditionality and social marginality: The folly of the tutelary state? *Critical Social Policy*, 38(4):771–791.

Fletcher, D. R. and Wright, S. (2018). A hand up or a slap down? criminalising benefit claimants in Britain via strategies of surveillance, sanctions and deterrence. *Critical Social Policy*, 38(2):323–344.

Gaisbauer, F., Pournaki, A., Banisch, S., and Olbrich, E. (2021). Ideological differences in engagement in public debate on twitter. *Plos one*, 16(3):e0249241.

Gao, T., Ye, M., and Ackland, R. (2024). Multi-layer network analysis of deliberation in an online discussion platform: The case of Reddit. *arXiv preprint*, arXiv:2410.21996.

Garritzmann, J. L., Neimanns, E., and Busemeyer, M. R. (2023). Public opinion towards welfare state reform: The role of political trust and government satisfaction. *European Journal of Political Research*, 62(4):1139–1166.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Harell, A., Soroka, S., and Iyengar, S. (2016). Race, prejudice and attitudes toward redistribution: A comparative experimental approach. *European Journal of Political Research*, 55(4):723–744.

Hasenfeld, Y. and Rafferty, J. A. (1989). The determinants of public attitudes toward the welfare state. *Social Forces*, 67(4):1027–1048.

Häusermann, S., Pinggera, M., Ares, M., and Enggist, M. (2022). Class and social policy in the knowledge economy. *European Journal of Political Research*, 61(2):462–484.

Hills, J. (2004). Heading for retirement? National Insurance, state pensions, and the future of the contributory principle in the UK. *Journal of Social Policy*, 33(3):347–371.

Hunzaker, M. F. and Valentino, L. (2019). Mapping cultural schemas: From theory to method. *American Sociological Review*, 84(5):950–981.

Jæger, M. M. and Kvist, J. (2003). Pressures on welfare states in post-industrial societies: Is more or less better? *Social Policy & Administration*, 37(6):555–572.

Janowitz, M. and Segal, D. R. (1967). Social cleavage and party affiliation: Germany, great britain, and the united states. *American Journal of Sociology*, 72(6):601–618.

Johnston, C. D., Newman, B. J., and Velez, Y. (2015). Ethnic change, personality, and polarization over immigration in the american public. *Public Opinion Quarterly*, 79(3):662–686.

Jost, J. T., Federico, C. M., and Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual review of psychology*, 60:307–337.

Jungkunz, S. (2021). Political polarization during the covid-19 pandemic. *Frontiers in Political Science*, 3:622512.

Larsen, C. A. (2008). The institutional logic of welfare attitudes: How welfare regimes influence public support. *Comparative Political Studies*, 41(2):145–168.

Levendusky, M. and Malhotra, N. (2016). Does media coverage of partisan polarization affect political attitudes? *Political Communication*, 33(2):283–301.

Lü, X. (2014). Social policy and regime legitimacy: the effects of education reform in china. *American Political Science Review*, 108(2):423–437.

Marres, N. (2015). Why map issues? on controversy analysis as a digital method. *Science, Technology, & Human Values*, 40(5):655–686.

Martin, J. L. and Desmond, M. (2010). Political position and social knowledge 1. In *Sociological Forum*, volume 25, pages 1–26. Wiley Online Library.

Mengel, F. and Weidenholzer, E. (2023). Preferences for redistribution. *Journal of Economic Surveys*, 37(5):1660–1677.

Patrick, R. (2017). Wither social citizenship? lived experiences of citizenship in/exclusion for recipients of out-of-work benefits. *Social Policy and Society*, 16(2):293–304.

Patrick, R. and Simpson, M. (2020). Conceptualising dignity in the context of social security: Bottom-up and top-down perspectives. *Social Policy & Administration*, 54(3):475–490.

Phelan, J., Link, B. G., Moore, R. E., and Stueve, A. (1997). The stigma of homelessness: The impact of the label" homeless" on attitudes toward poor persons. *Social psychology quarterly*, pages 323–337.

Rawlings, C. M. (2022). Becoming an ideologue: Social sorting and the microfoundations of polarization. *Sociological Science*, 9:313–345.

Rehm, P., Hacker, J. S., and Schlesinger, M. (2012). Insecure alliances: Risk, inequality, and support for the welfare state. *American Political Science Review*, 106(2):386–406.

Rivas-de Roca, R. and García-Gordillo, M. (2022). Understanding the european public sphere: a review of pending challenges in research. *European politics and society*, 23(3):380–394.

Shapiro, R. Y. and Young, J. T. (1989). Public opinion and the welfare state: The united states in comparative perspective. *Political Science Quarterly*, 104(1):59–89.

Shugars, S. and Beauchamp, N. (2019). Why keep arguing? predicting engagement in political conversations online. *SAGE Open*, 9(1):2158244019828850.

Sîrbu, A., Pedreschi, D., Giannotti, F., and Kertész, J. (2019). Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PloS one*, 14(3):e0213246.

Summers, K., Edmiston, D., Baumberg Geiger, B., Ingold, J., Scullion, L., de Vries, R., and Young, D. (2025). Claiming deservingness: The durability of social security claimant discourses during the Covid-19 pandemic. *The Sociological Review*, 0(0):1–21. Advance online publication.

Svallfors, S. (2013). Government quality, egalitarianism, and attitudes to taxes and social spending: A European comparison. *European Political Science Review*, 5(3):363–380.

Taylor-Gooby, P. (1982). Two cheers for the welfare state: Public opinion and private welfare. *Journal of Public Policy*, 2(4):319–346.

Urman, A. and Makhortykh, M. (2023). You are how (and where) you search? comparative analysis of web search behavior using web tracking data. *Journal of Computational Social Science*, pages 1–16.

Van Oorschot, W. (2006). Making the difference in social europe: deservingness perceptions among citizens of european welfare states. *Journal of European social policy*, 16(1):23–42.

van Oorschot, W. (2006). Making the difference in social Europe: Deservingness perceptions among citizens of European welfare states. *Journal of European Social Policy*, 16(1):23–42.

Willaert, T., Van Eecke, P., Beuls, K., and Steels, L. (2020). Building social media observatories for monitoring online opinion dynamics. *Social Media+ Society*, 6(2):2056305119898778.

Xing, Y., Wang, X., Qiu, C., Li, Y., and He, W. (2022). Research on opinion polarization by big data analytics capabilities in online social networks. *Technology in Society*, 68:101902.

Yarchi, M., Baden, C., and Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, 38(1-2):98–139.

Yasseri, T., Sumi, R., Rung, A., Kornai, A., and Kertész, J. (2012). Dynamics of conflicts in wikipedia. *PloS one*, 7(6):e38869.

Appendix A. Online Appendix

Corpus Retrieval Diagnostics by Keyword Theme

Figure A1 shows the month-by-month post volume for each of the six keyword themes over the 2020–2025 period. The retrieval timeline reveals two important features of corpus coverage. First, temporal density is uneven across themes: poverty and council tax show relatively consistent activity, while UBI and national insurance exhibit concentrated spikes around specific political events (e.g. the 2021 NIC levy announcement and periodic UBI debates). Second, UBI has the sparsest coverage with only 37 posts and a 72% gap rate across the observation window. The poverty gap in 2020–late 2021 was partially addressed through a dedicated recollection run (see main text).

Table A1 summarises retrieval metrics by theme. Poverty has the best temporal continuity (3% gap months) while UBI the worst (72%). Average comments per post though are highest for UBI (21.0) and national insurance (18.1), reflecting the greater engagement intensity of these topics despite lower post volumes. The precision of keyword matching matters: council tax, income tax, and national insurance are conceptually narrow and likely to capture substantively relevant discourse; poverty and UBI are broader terms that may retrieve tangentially related posts. Minimum wage sits in-between. These breadth differences motivate the policy-matching step, which constrains the event analysis to posts with demonstrated semantic proximity to specific policy documents.

Table A1: Corpus retrieval diagnostics by keyword theme (r/UKPolitics, 2020–2025)

Theme	Posts	Cmts	First post	Last post	Gap M	Gap %	Avg. cmts
Poverty	286	3,487	2020-01	2025-10	2	3%	12.2
National Ins	113	2,050	2020-04	2025-10	39	58%	18.1
Council Tax	145	1,607	2020-02	2025-10	17	25%	11.1
Income Tax	105	1,667	2020-02	2025-10	28	41%	15.9
Minimum W.g	81	1,454	2020-01	2025-10	30	43%	18.0
UBI	37	778	2020-03	2025-09	48	72%	21.0
Total	**767**	**10,713**					

Note: Gap M = calendar months within the observation window with zero posts retrieved for that theme. Gap % = gap months as a proportion of total months in range.

Model Agreement and Cross-Model Consistency

Table A2, Figure A2, and Figure A3 report inter-model agreement rates and per-model stance and attribution-target distributions across the three LLM models (Gemma-3-27B, Llama-3.3-70B, DeepSeek-R1-32B) and the majority-vote ensemble (n = 9,920).

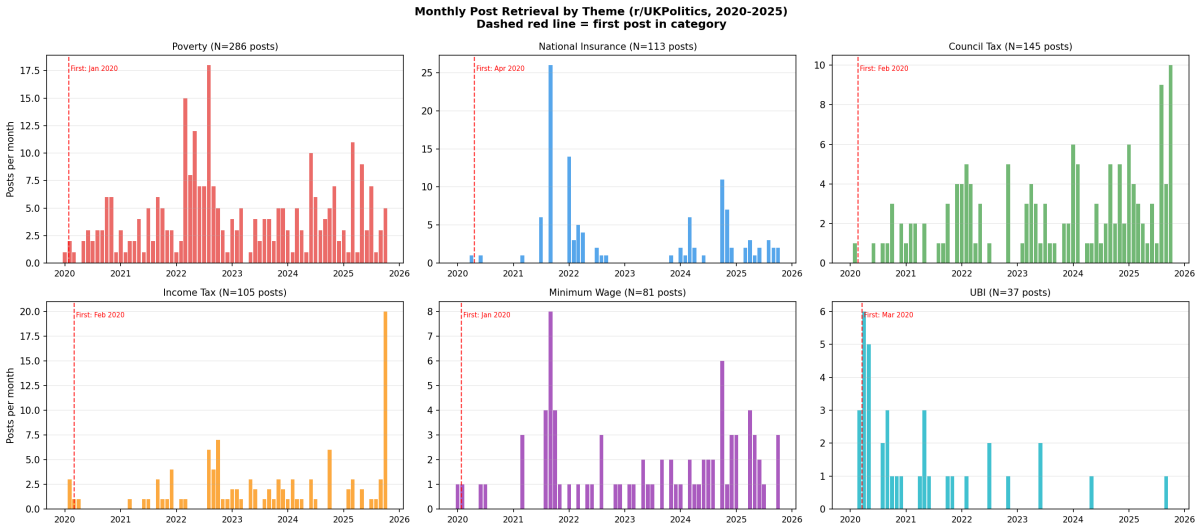

Figure A1: Monthly post volume by keyword theme, r/UKPolitics 2020–2025. Each panel shows the number of posts retrieved per calendar month. Shaded areas reflect periods with no posts retrieved (gap months). Poverty coverage is most complete following the 2020–2021 recollection run; UBI shows the sparsest and most episodic retrieval pattern.

Pairwise stance agreement ranges from 67.4% (Gemma vs. DeepSeek) to 77.8% (Gemma vs. Llama); three-model unanimous agreement stands at 60.4%. For attribution targets, three-model unanimity is lower at 43.3%, consistent with the finer-grained eight-category taxonomy. The overall directional ordering is consistent across all models: disapproval dominates in every topic and every model, and government and political parties are the two most frequently cited attribution targets. The main divergence is in DeepSeek's markedly higher rate of `none` attributions (39.2% vs. 16–17% for the other two models), reflecting a more conservative coding style; the ensemble majority-vote rule effectively dampens this tendency. These results confirm that the substantive patterns reported in the main text are not artefacts of model choice or the Gemma tie-breaking rule.

Berttopic results for Comments and Attribution text

Figures A4 and A5 describe the thematic architecture of the discourse corpus recovered by BERTopic (n = 10,032 comments, nr_topics = 30, sentence embeddings via all-MiniLM-L6-v2 with bigram representation). Figure A4 shows how the 28 substantive topics (plus the residual outlier cluster) distribute across the six keyword categories. Several topics are almost entirely confined to a single category: the UBI cluster (T4, *ubi / people / work*) concentrates 94% of its documents in the UBI category; the minimum wage cluster (T2, *wage / minimum / minimum wage*) concentrates 84% in minimum wage; and the school and private-sector employment cluster (T9, *school / loan / private*)

Table A2: Inter-model agreement rates and stance/target distributions by model (n = 9,920)

	Gemma-27B	Llama-70B	DeepSeek-32B	Ensemble
Pairwise stance agreement				
Gemma-27B vs Llama-70B		77.8%		
Gemma-27B vs DeepSeek-32B		67.4%		
Llama-70B vs DeepSeek-32B		72.0%		
3-model unanimous		60.4%		
Pairwise attribution-target agreement				
Gemma-27B vs Llama-70B		68.6%		
Gemma-27B vs DeepSeek-32B		50.0%		
Llama-70B vs DeepSeek-32B		57.3%		
3-model unanimous		43.3%		
Both unanimous		28.9%		
Stance distribution (%)				
Approve	23.5	18.3	26.9	22.0
Disapprove	65.3	59.7	43.2	59.6
Neutral	11.3	22.0	29.9	18.3
Top three attribution targets (%)				
Government	26.2	30.2	26.4	27.5
Political parties	27.2	26.5	11.7	24.2
None	11.8	17.2	39.2	16.6

Note: Pairwise and unanimous agreement computed on comment-level labels. Ensemble uses majority vote with Gemma-27B as tiebreaker. Top three targets shown are for the ensemble.

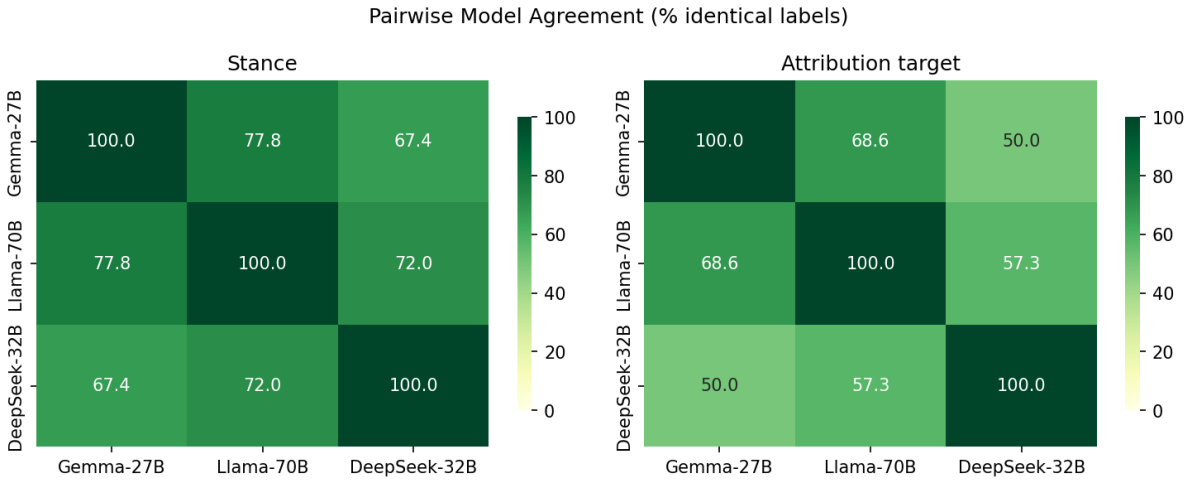

Figure A2: Pairwise inter-model agreement rates (%) for stance and attribution-target labels (n = 9,920 comments, 3-model ensemble: Gemma-3-27B, Llama-3.3-70B, DeepSeek-R1-32B). Left panel: stance (approve / disapprove / neutral) agreement; right panel: attribution target agreement across eight target categories. Darker green indicates higher agreement.

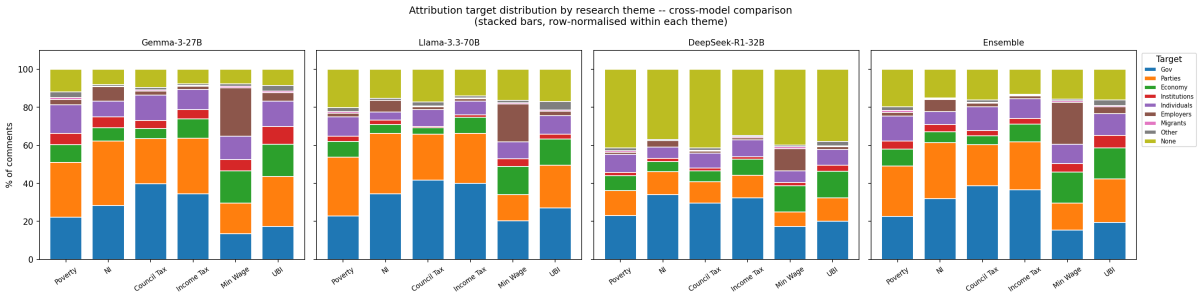

Figure A3: Attribution-target distributions by keyword theme and model, plus ensemble consensus ($n = 9{,}920$).

likewise concentrates 69% in minimum wage, reflecting discussion of graduate wages and public-sector pay. The general tax and fiscal debate cluster (T0, *tax / council / council tax*) spreads more evenly across all six categories. The AI and automation cluster (T22, *ai / jobs / learning*) is predominantly located in UBI (66%) and poverty (20%), consistent with the macroeconomic framing of UBI as a response to technological unemployment described in the main text.

Figure A5 displays the top keyword bars for each of the 28 topics and provides direct evidence for the topic labels used in the analysis. The three main substantive discourse groups visible in the keyword patterns are: (i) *welfare and minimum income* (topics centred on poverty, NI, wage, and UBI vocabulary); (ii) *tax policy and fiscal debate* (topics organised around council tax, income tax rates, percentage thresholds, triple lock, and second homes); and (iii) *political parties and leadership* (topics dominated by named politicians — Truss, Johnson, Starmer, Sunak, Corbyn — and partisan language). Noise topics, which account for 39.5% of the corpus by document count, are characterised by meta-linguistic or artefactual keywords: reaction terms (*ha, surprise, good*), archiving bot commands, and headline fragments, and they are excluded from all attribution and stance analyses in the main text.

Figure A6 shows the BERTopic keyword structure of the 11 topics recovered from the LLM-extracted attribution text (9,026 lines). Running BERTopic on the attribution field rather than on the raw comment text reveals the *causal reasoning vocabulary* embedded in comments rather than their topical surface form. The residual cluster (T0, 4,612 lines) captures the generic attribution descriptions shared across all policy themes: comment-level inference language (*comment implies, suggests, attributes*) combined with policy domain terms (*tax, poverty, council*). The partisan attribution cluster (T1, 742 lines) groups comments explicitly naming political parties and their actions, confirming that Labour and the Conservatives serve as the most frequently invoked named agents of

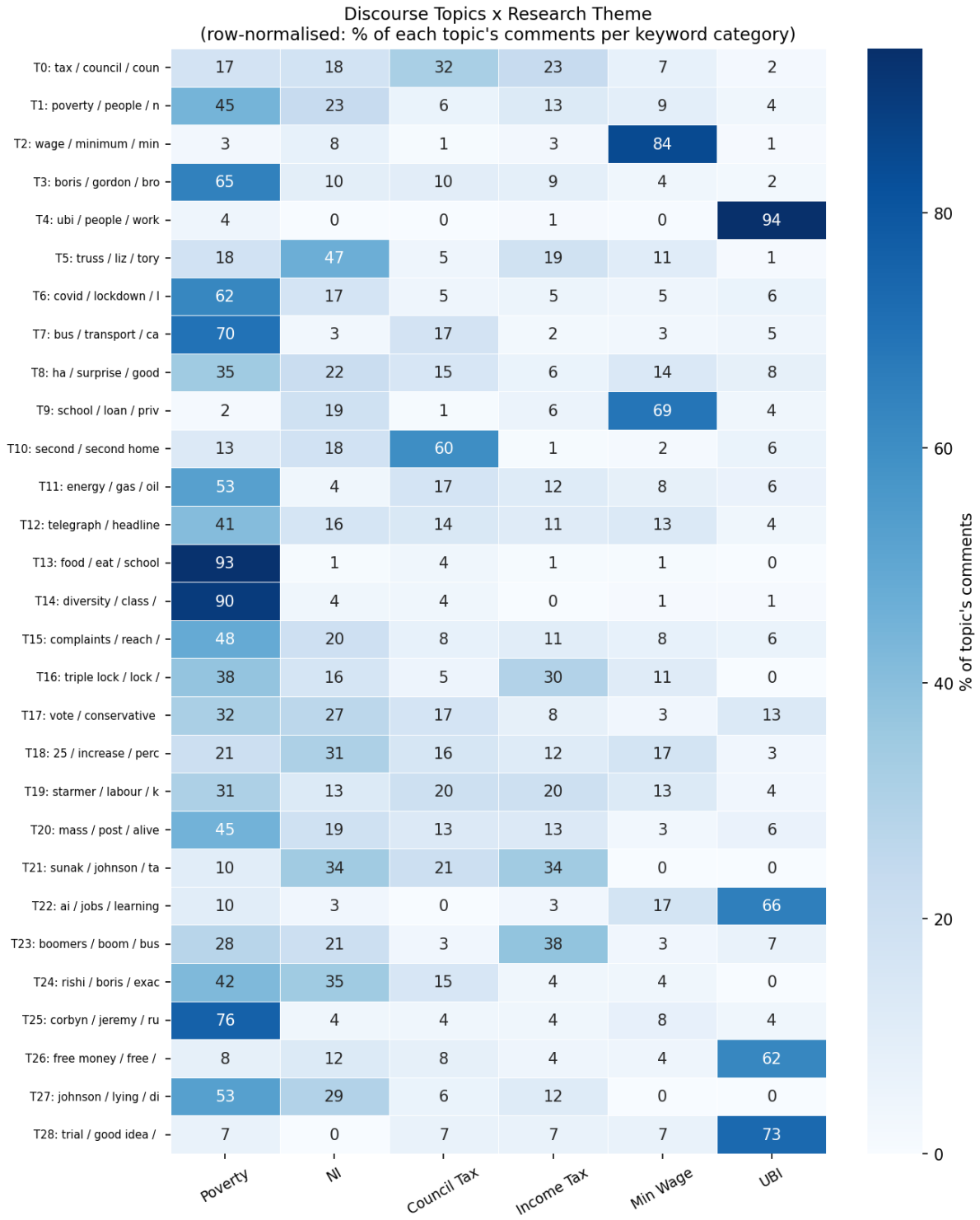

Figure A4: Row-normalised distribution of BERTopic discourse topics across the six keyword categories (r/UKPolitics, 2020–2025, $n = 10{,}032$). Each row represents one topic; colour intensity encodes the percentage of that topic's documents originating from each keyword category (rows sum to 100%).

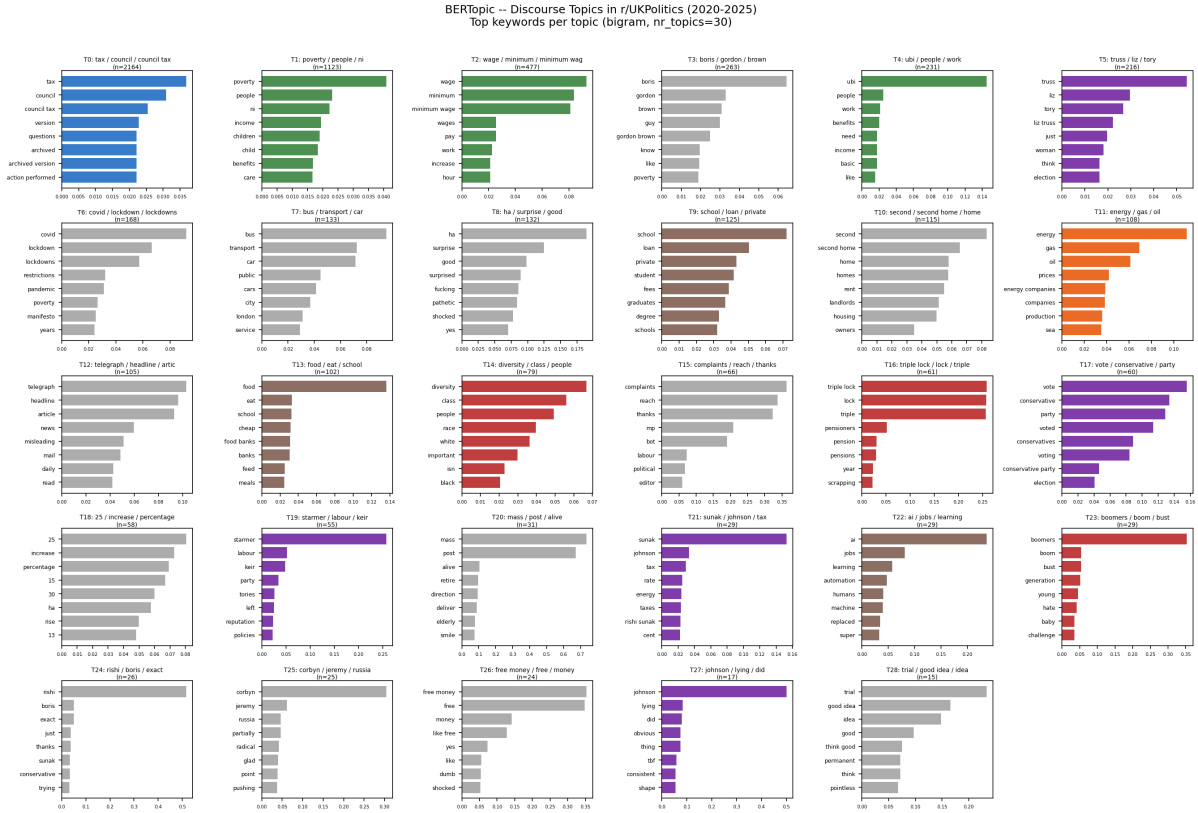

Figure A5: Top keywords per BERTopic discourse topic (bigram, $nr_topics = 30$, all-MiniLM-L6-v2 embeddings). Each panel shows the c-TF-IDF weighted keywords for one topic; panel titles give the topic ID and the auto-generated label. Substantive topics (welfare, tax, political parties, inequality) are distinguishable by policy-domain vocabulary; noise and artefact topics are characterised by meta-linguistic terms, named-politician references without policy content, or archiving-bot commands. Topics are colour-coded by the five discourse groups used in the main text: tax policy and fiscal debate (blue), welfare and minimum income (green), political parties and leadership (orange), inequality and class (purple), and social and identity issues (red).

accountability in the corpus; representative documents in this cluster directly attribute policy outcomes to Conservative decisions or Labour alternatives. The UBI attribution cluster (T2, 371 lines) is characterised by epistemic and conditional vocabulary — *potential, implementation, existing, financial constraints, DWP redundant* — reflecting the hypothetical and systemic elements in which UBI arguments are typically framed: commenters reason about what *would* happen under UBI rather than what a current policy *has* done. The media attribution cluster (T3, 297 lines) groups comments attributing policy narratives or political outcomes to BBC reporting and specific journalists (*Kuenssberg*), pointing to institutional media trust as a recurring dimension of redistributive accountability framing. The Truss-specific cluster (T4, 121 lines) captures the concentrated attributional response to the September 2022 mini-Budget and Liz Truss's tenure, consistent with the intensification of political-party attribution visible in the timeline figures for that period. Smaller clusters (AI/automation, T9; multi-factor inequality, T10) reflect niche but analytically distinct causal frames that are most prominent in poverty discussions. Taken together, the attribution topic structure confirms that causal accountability in online redistributive discourse is not uniformly distributed but organised around recognisable frames: partisan responsibility, institutional critique, conditional systemic reasoning, and media-mediated blame.

Policy–Comment Match Manual Audit

To assess the precision of the policy-matching procedure, 30 matched comment–policy pairs were drawn by stratified random sampling from the borderline score range 0.25–0.40, restricted to comments within ±90 days of the policy announcement date. Three sub-bins of ten rows each were sampled (0.25–0.30, 0.30–0.35, 0.35–0.40) to assess whether match quality degrades near the lower threshold. Each row was manually coded by author as *Good* (comment clearly discusses the matched policy topic), *Borderline* (plausibly related but ambiguous), or *Bad* (clear topical mismatch).

Table A3 reports precision by score bin (precision = Good + Borderline / total). Overall precision is 96.7% across the 30 audited rows. Only one outright mismatch was found, in the lowest sub-bin (0.25–0.30), and four borderline cases total. Precision rises monotonically with score: 90.0% in the 0.25–0.30 bin and 100% in both higher bins. These results confirm that the 0.25 threshold dmits only a small proportion of genuinely off-topic matches, and that the vast majority of matched comments engage substantively with the relevant policy domain.

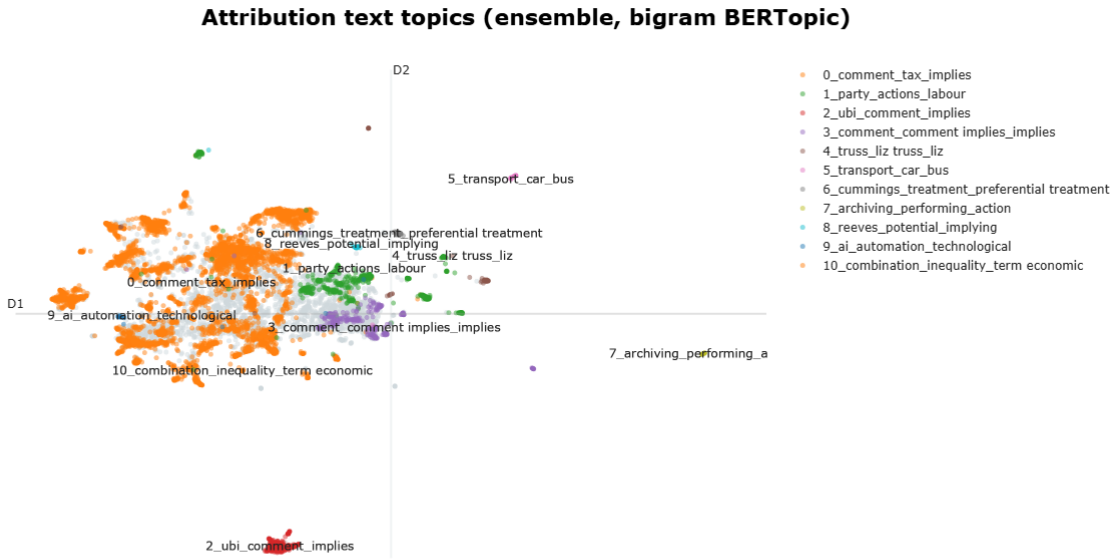

Figure A6: BERTopic keyword structure of attribution text topics (LLM-extracted attribution field, $n = 9{,}026$ documents, 11 topics recovered). Each panel shows the top c-TF-IDF keywords for one attribution topic. Unlike the discourse topic model (Figures A4–A5), which clusters raw comment text, this model clusters the structured attribution rationale extracted by the LLM, thereby surfacing causal-reasoning vocabulary rather than topical surface form.

Table A3: Manual audit precision by match score bin (score 0.25–0.40, ±90-day window)

Score bin	N	Good	Borderline	Bad	Precision
0.25–0.30	10	8	1	1	90.0%
0.30–0.35	10	9	1	0	100.0%
0.35–0.40	10	8	2	0	100.0%
Total	**30**	**25**	**4**	**1**	**96.7%**

Note: Precision = (Good + Borderline) / N. Rows sampled by stratified random draw across six keyword themes within ±90 days of policy date. Manual coding by the author.

Policy–Comment Match Score Examples

As discussed in the main text, to qualify for the event analysis, a comment must be matched to a policy document with a composite similarity score of at least 0.25. The score blends TF-IDF cosine similarity (weight 0.4) and SentenceBERT semantic cosine similarity (weight 0.6), computed between the full text of the Reddit post containing the comment AND the official policy title and description. A score of 0.25 therefore reflects a modest but non-trivial lexical and semantic overlap: the post discusses concepts that are genuinely related to the policy, but may not name it directly. Scores at or above 0.40 indicate very strong, often near-verbatim topical alignment, where the post explicitly references the policy by name or closely paraphrases its core provisions. This similarity score shouldn't be confused/mixed up with normal range of [0, 1] in interpretation as it is not possible to have perfect naming in the online discussions that matches the exact policy name.

Table A4 illustrates this gradient with ten matched pairs drawn from the event-analysis dataset. The first five rows show borderline matches (0.25–0.40): the posts address a thematically related issue without naming the policy precisely, reflecting the semantic reach of the SentenceBERT component. The lower five rows show high-confidence matches ($\geq$0.40): the posts directly reference the policy, and the lexical overlap with the policy description is immediately apparent. Only comments from posts that meet the 0.25 threshold within a $\pm$90-day window around the policy announcement date are included in the event analysis.

Volume Shift vs. Attribution Shift

Figure A7 plots, for each of the eight top discussed/engaged policies, the pre-to-post volume ratio ($n_{\text{post}}/n_{\text{pre}}$) against the Pearson χ^2 statistic and Cramér's V from the attribution-target contingency test. If attribution shifts were driven primarily by volume changes – more comments generating larger χ^2 by chance – we would expect a strong positive correlation between the volume ratio and χ^2. The observed correlations are near-zero and statistically insignificant: $r(\text{volume ratio}, \chi^2) = -0.197$ ($p = 0.672$) and $r(\text{volume ratio}, V) = 0.365$ ($p = 0.420$). Since Cramér's V is sample-size adjusted, the absence of correlation with V is the more decisive test: it confirms that the detected attribution shifts reflect genuine reorientation in causal attribution patterns around policy events, not a statistical artefact of differential comment volumes.

Table A4: Representative policy–post match score examples from the event-analysis dataset

Score	Keyword	Policy	Post title (excerpt)
Borderline matches (0.25–0.40): topically related but not explicitly named			
0.250	Poverty	Pathways to Work: Reforming Benefits and Support to Get Britain Working	NHS nurses recruited from overseas are being pushed into poverty by the Home Office
0.250	Council tax	Council tax referendum principles 2024–25	Council tax to double for 80 per cent of second homes
0.251	National insurance	Long-term ambition to abolish employee National Insurance	Chancellor considering further NI cut as Tories trail Labour by 20 points
0.252	Minimum wage	National Minimum Wage / NLW uprating 2022	Workers in UK are paid less than minimum wage. Comparable to Portugal
0.253	Income tax	Income-tax charge on Winter Fuel Payments above £35,000	Ministers asked to clarify income tax pledge on winter fuel payments
High-confidence matches (≥0.40): direct policy reference or near-verbatim alignment			
0.401	National insurance	Increase of NI thresholds from July 2022	Sunak raises National Insurance threshold to £12,570
0.401	National insurance	Health and Social Care Levy Act 2021 / 1.25pp NIC rise	Boris Johnson defends National Insurance rise as fair
0.403	Poverty	Tackling Child Poverty: Developing Our Strategy	The Tory record on welfare: PIP changes, child poverty, benefit cuts
0.403	Council tax	Council tax referendum principles 2024–25	Council tax: Millions facing 5% increases from April
0.407	Minimum wage	National Minimum Wage / NLW uprating 2025	Autumn budget: Rachel Reeves to raise national minimum wage by 6.7% in 2025

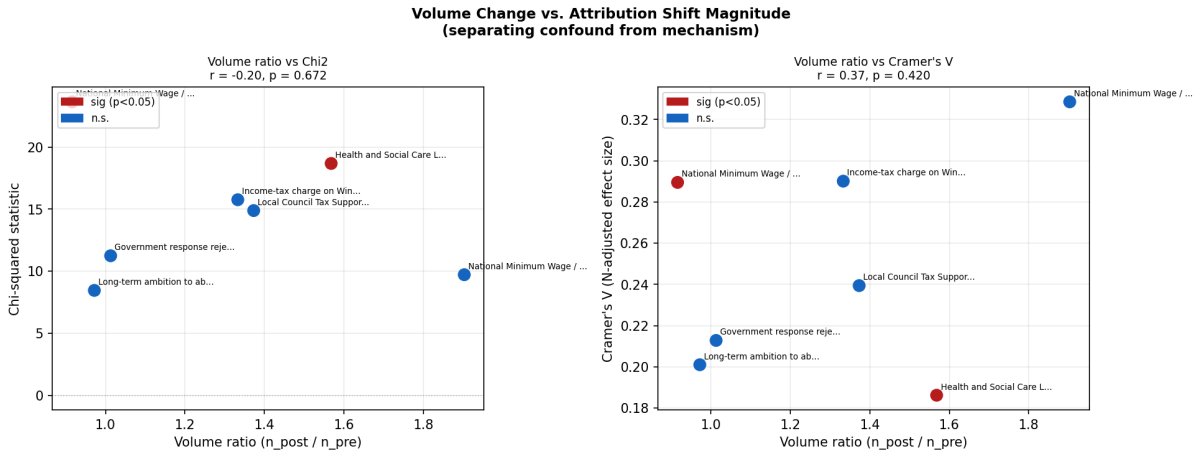

Figure A7: Volume ratio ($n_{\text{post}}/n_{\text{pre}}$) versus χ^2 statistic (left) and Cramér's V (right) for the eight tested policies. Red markers indicate policies reaching $p < 0.05$; blue markers indicate non-significant policies. Near-zero correlations ($r = -0.20$ and $r = 0.37$ respectively, both $p > 0.4$) confirm that attribution shifts are not driven by post-period volume inflation.

Standardised Residuals by Attribution Target

Figure A8 displays standardised Pearson residuals from the pre/post attribution-target contingency tables for the three statistically significant policies. Each cell shows the deviation of the observed cell count from the expected count under independence, expressed in standard deviation units; cells with $|r| > 1.96$ are conventionally regarded as driving the overall χ^2 statistic. The residual patterns reveal target-specific, rather than general, attribution reorientation around each policy event. For the Health and Social Care Levy (NIC rise), government attribution increases post-announcement (residual $+1.42$) while individual attribution declines (-1.72), consistent with a shift from personal responsibility framing to institutional blame as the levy's fiscal mechanism became publicly debated. For the National Minimum Wage uprating 2025, employer attribution rises sharply ($+1.76$) while political-party attribution falls (-2.02), the sole cell exceeding the 1.96 threshold, suggesting that the wage-floor increase refocused blame from partisan actors to business cost-bearers. For the Winter Fuel Payment income-tax charge, the shift is more diffuse across targets, with political-party attribution rising ($+1.30$) and institutional attribution falling (-1.50), but no single cell exceeds the significance threshold. Taken together, the residual analysis confirms that each significant shift is driven by reorientation toward particular attribution targets rather than global distributional volatility.

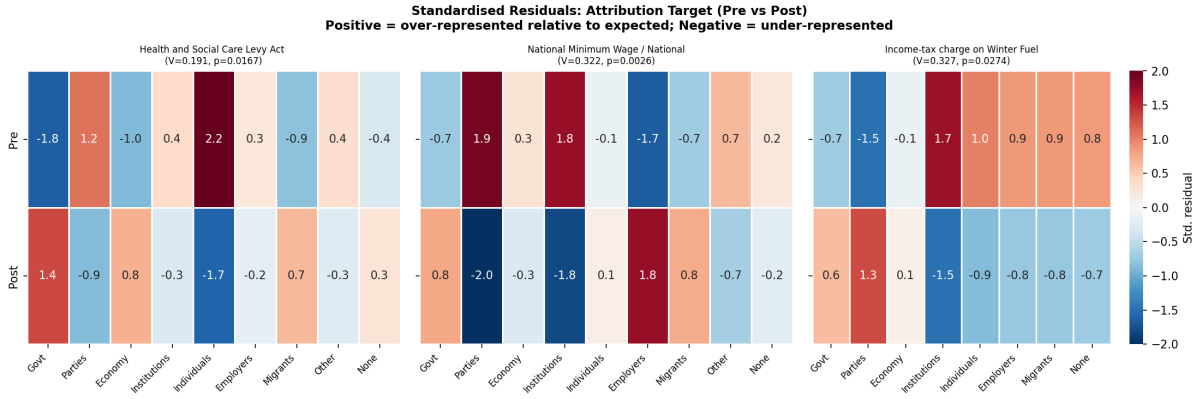

Figure A8: Standardised Pearson residuals for the three statistically significant pre/post policy events. Rows = attribution targets; columns = pre/post period. Colour intensity encodes residual magnitude; cells outlined in bold indicate $|r| > 1.96$. Positive residuals (red) indicate over-representation in that period; negative residuals (blue) indicate under-representation. Only policies with $p < 0.05$ in the ensemble Pearson χ^2 test are shown.