

New Leader, New Crowd: Labour’s Facebook Comment Section from Miliband to Corbyn

Tristan Hotham*

2026

Preprint, not peer reviewed

Abstract

Jeremy Corbyn’s election as Labour leader in September 2015 brought a surge of new members and a reputation for an unusually energetic, and unusually combative, online following. I test what changed in the place where that following met the party most directly: the comment section of Labour’s own Facebook page. I compare every comment that could be retrieved from posts on the page between 1 January and 12 April in 2015, under Ed Miliband (174,560 comments on 461 posts), and in the same weeks of 2016, under Corbyn (60,019 comments on 131 posts). I score each comment with an open toxicity classifier, track who and what it mentions, follow the 1,964 accounts that commented in both years, and read a random sample of 300 comments with AI assistance to estimate stance. In the same weeks of the year, the comment section under Corbyn was smaller, more conversational and slightly less toxic: 8.9% of comments were toxic in 2016 against 10.9% in 2015. About half of that fall reflects a change of crowd: the accounts seen only in the 2015 window had been more toxic than those seen only in 2016. The other half is a smaller fall among the accounts that commented in both years, which is shared with the rest of the page and may reflect the period rather than the people. What changed most was the argument: comments about UKIP fell from 9.6% to 3.0%, and comments about the EU, the leader and Labour’s internal factions rose. Comments the classifier flagged as identity attacks and that mentioned Jewish people rose from 12 to 18. On reading, most in 2016 accused Labour of tolerating antisemitism or challenged other commenters, but the numbers are too small, and the classifier too unreliable, to estimate how common antisemitic comments were. The comparison is descriptive: two periods on one page cannot separate the effect of a leader from the effect of an election year, and comments from 2015 were retrieved a year after they were written and fewer of them survived. Within those limits, the evidence suggests that this comment section, in these weeks, did not become more hostile under Corbyn, and that the main change was in what people argued about.

Keywords: Labour Party; Jeremy Corbyn; Ed Miliband; Facebook; comment sections; incivility; toxicity; party factions

*Independent researcher. ORCID 0000-0003-4931-9349. Correspondence: contact@tristanhotham.com. The data were collected during doctoral research funded by the Economic and Social Research Council. The views expressed are the author’s own and do not represent those of any employer.

1 Introduction

Between May and September 2015 Labour changed leader, rules and membership. Ed Miliband resigned after losing the general election. Under a new one-member-one-vote system, with registered supporters able to vote for £3, Jeremy Corbyn won the leadership in the first round. Party membership more than doubled within a year ([Whiteley et al. 2019](#)). Much of the energy behind Corbyn was organised and expressed online, through his own pages, through Momentum and through a network of supportive sites and groups ([Dommett and Temple 2018](#); [Dennis 2020](#); [Gerbaudo 2019](#)). Critics within and outside the party said that the same online energy brought abuse, factional hostility and, later, antisemitism into Labour spaces ([Equality and Human Rights Commission 2020](#)).

Those claims are hard to test because most of the online spaces in question were private, ephemeral or too large to observe. One space was neither: the public comment section of the Labour Party's own Facebook page. Anyone could comment there, the party's posts set the agenda, and in 2015 and 2016 the comments could be collected through Netvizz ([Rieder 2013](#)). During my doctoral research I collected the comments on every post the page made between 1 January and 12 April in both years ([Hotham 2021](#)). That gives a rare comparison: the same page, the same weeks of the calendar, one year under Miliband and one under Corbyn. It is not like for like, since the two years differed in their elections, their news and how the page itself posted, and I return to this throughout.

I ask four questions. Did the people commenting change? Did the comment section become more or less toxic? Did the targets of hostility change, from opponents to factions within the party? And were any changes driven by new people arriving, or by the same people behaving differently?

The answers matter beyond Labour. Party pages are where parties speak to the public most directly on social media, and their comment sections are one of the few places where supporters, opponents and the merely curious meet in front of the party itself ([Stromer-Galley 2014](#); [Kalsnes et al. 2017](#)). If a change of leader goes with a change in who turns up and how they behave there, a party's audience on its own page is not fixed, and claims about a leader's online following can be checked against it.

This paper extends my earlier work on candidate comment sections in the United States ([Hotham 2026e,c](#)) and in the 2016 London mayoral election ([Hotham 2026b](#)), and on how British parties divided work between leader and party pages ([Hotham 2026a,d](#)).

2 Background and expectations

2.1 Comment sections as political spaces

Research on online comments has moved from asking whether comment sections are uncivil, which they often are ([Coe et al. 2014](#); [Rowe 2015](#)), to asking who produces the incivility and why. Two findings are important here. First, a small share of users write a large share of comments, and in some settings the most active users are the most hostile ([Kim et al. 2021](#)). What looks like the tone of a community can be the tone of its most prolific members. Second, incivility tracks the topic and the target. Comments about opponents, contested identities and moral issues are more often uncivil than comments about policy detail ([Rossini 2022](#); [Theocharis et al. 2020](#)). In Britain, studies of abuse directed at MPs find that it follows the news and the target's party and identity ([Ward and McLoughlin 2020](#); [Southern and Harmer 2021](#); [Gorrell et al. 2020](#)).

Both findings point to composition. If a new leader brings a new crowd, the comment section can change

even if nobody’s behaviour does. If the same people stay but the arguments change, the comment section can change even if the crowd does not.

2.2 Labour from Miliband to Corbyn

Three features of the change of leadership suggest specific expectations.

A new and more committed audience. Most of those who joined Labour after May 2015 did so to support Corbyn, and many were returning former members or first-time joiners on the left (Whiteley et al. 2019; Bale et al. 2020). Corbyn’s supporters were active on Facebook, and much of his support was mobilised through it (Dennis 2020; Gerbaudo 2019). I therefore expect a comment section in 2016 with more conversation among supporters, more replies, and a larger share of comments supporting the party.

A different election calendar. In early 2015 Labour was heading into a general election on 7 May. The short campaign began on 30 March, inside the observation window. In early 2016 the next contests were the local, devolved and mayoral elections of 5 May, and the EU referendum campaign was starting. Opposing partisans, especially Conservative and UKIP supporters, had more reason to visit Labour’s page in 2015. I therefore expect fewer comments from outside the party in 2016, and a shift in topic from the general election contest to the EU and to Labour itself.

Factional conflict. Corbyn’s leadership opened a public argument between his supporters and parts of the parliamentary party, often expressed online in terms such as “Blairite”, “red Tory” or “Trot” (Dennis 2020). I expect more comments about internal factions in 2016. Whether this conflict made the comment section more hostile overall is an open question. Hostility from opponents might fall even as hostility within the party rose.

3 Data and methods

3.1 Comments

The data come from Labour’s official Facebook page (“The Labour Party”). I used Netvizz to collect every post the page made between 1 January and 12 April 2015 (461 posts) and between 1 January and 12 April 2016 (131 posts), and all comments and replies that the API returned for each post. Both collections were made in April 2016. Netvizz replaced commenter identities with a one-way hash that was consistent across collections, so I can link comments by the same account across years without knowing who wrote them.

Two features of the collection shape what can be compared. First, the 2015 comments were collected about a year after they were written, and the 2016 comments within weeks. Facebook reported 235,781 comments on the 2015 posts and 72,721 on the 2016 posts. The API returned 74.6% and 86.6% of them, leaving 59,887 and 9,716 missing. Comments deleted in the meantime, by their authors, by the page or by Facebook, are missing, and so may be some replies, since the API’s handling of reply threads changed between the two years. I check how much this could matter in three ways: by testing whether a post’s retrieval rate predicts the toxicity of its comments, by restricting the comparison to posts with high retrieval, and by asking how toxic the missing comments would have to be to erase the difference.

Second, the 2015 window includes the first two weeks of the general election’s short campaign. I repeat every comparison for January to March only, which is outside the short campaign in both years.

I removed 4,320 exact repeats of the same text by the same account on the same post: 1,334 in 2015 and 2,986 in 2016. The difference is probably a collection artefact, and I report results with the repeats kept. The final data contain 174,560 comments in 2015 and 60,019 in 2016.

3.2 Measures

Toxicity. I scored every comment with the original Detoxify model (Hanu and Unitary team 2020), trained on the Jigsaw toxic comment data (Jigsaw and Conversation AI 2018). It returns probabilities for toxicity, severe toxicity, obscenity, threats, insults and identity attacks. I treat a probability of 0.5 or more as a positive. The model is trained mainly on North American comments and is better at comparing rates across large groups than at classifying single comments (Hede et al. 2021). Classifiers of this kind are imperfect and can over-flag comments that mention identity groups (Dixon et al. 2018; Sap et al. 2019). I therefore report identity-attack rates with care and read flagged comments where a specific claim rests on them.

Mentions. I flag comments that mention each party leader, Labour, the Conservatives, UKIP, the SNP, the Liberal Democrats and the Greens. I also flag pejorative internal-faction labels: “Blairite”, “red Tory” and “Tory-lite” on one side, and “Trot”, “hard left” and “loony left” on the other. Broader lists that include neutral terms such as “New Labour”, “Momentum” and “socialist” are reported in the appendix. Mentions of “Jeremy” followed by “Hunt” are not counted as mentions of Corbyn. I also flag mentions of Jewish people, Israel or antisemitism, immigration, the EU, the NHS and the economy. The word lists are in the appendix. They are deliberately simple and over-inclusive. I use them to describe the mix of topics, not as measures of attitude.

Stance. Word lists are poor measures of support. A test list of supportive phrases matched only six of the 86 supportive comments in a validation sample. Instead I drew a random sample of 150 comments from each year and coded each one with the help of an AI model (Claude, Anthropic), whose codes I reviewed, recording whether it supported Labour or its leader (including attacks on Labour’s opponents from a Labour-supporting position), criticised Labour or its leader, or did neither. I also flagged comments about Labour’s internal factions. The coding was done without reference to the year, although many comments reveal it. The same procedure and categories were used to read the 30 comments flagged as identity attacks that mention Jewish people, coding whether each was antisemitic, accused Labour or others of antisemitism, challenged antisemitism, or none of these. Because the coding was AI-assisted and reviewed by one person, I treat it as an estimate that needs independent checking, and I do not publish the comment texts.

3.3 Analysis

The unit of comparison is the comment, but comments on the same post are not independent. I report confidence intervals from 1,000 bootstrap resamples of posts within each year, and regressions with standard errors clustered by post. Controls are comment length (log characters), whether the comment is a reply, and calendar month. I also report standard errors clustered by both post and account. To separate composition from behaviour, I group accounts by the windows in which they were seen: only in 2015, in both, or only in 2016. These groups are defined by the two windows, not by membership of the party or loyalty to a leader: an account seen only in 2015 may have commented on the page later in 2015 or after April 2016. I compare the 1,964 accounts seen in both windows with themselves, using account fixed effects, and decompose the overall change into the part due to the change of crowd and the part

due to the change among those seen in both. Because accounts seen in both windows changed too, and that change may reflect the period rather than the people, I treat it as an estimate of the period effect and apply it to the 2015 crowd to estimate a counterfactual.

This is a comparison of two periods on one page. It can show what changed. It cannot show that the change of leader caused it, because the election calendar, the national mood and Facebook itself also changed between the two periods.

4 Findings

4.1 Who commented

The comment section in 2016 was smaller but busier per post (Table 1). The page posted far less often under Corbyn in these weeks (131 posts against 461), a change in the party's own strategy that is itself part of what differs between the two periods, and each post drew more comments: a median of 427 retrieved comments per post against 324. About 10,000 accounts commented in 2016 against 30,000 in 2015. In both years participation was extremely concentrated. The most active 1% of commenters wrote 48% of all comments in 2015 and 45% in 2016, and well over half of commenters wrote only once.

The comment section became more conversational. Replies to other commenters made up 55% of comments in 2015 and 67% in 2016.

The crowd changed more than its size suggests. Only 1,964 accounts commented in both windows. They were 20% of the 2016 commenters but wrote 38% of the 2016 comments, and they had written 37% of the 2015 comments as well. The rest of the 2016 comment section came from 8,092 accounts not seen in the 2015 window, and 28,191 accounts seen in 2015 were not seen in the 2016 window.

4.2 Toxicity fell slightly

The comment section was a little less toxic in 2016 (Figure 1). In 2015, 10.9% of comments were classified as toxic; in 2016, 8.9%. The difference of 2.0 percentage points (95% interval 1.4 to 2.8) holds with controls for comment length, replies and calendar month (1.8 points, $p < 0.001$), with standard errors clustered by both post and account (SE 0.55, $p < 0.001$), with controls for the type of post (1.8 points), and with controls for the topics comments mentioned (1.7 points; Table 2). Weighting each post equally rather than each comment gives 10.4% against 8.9%. Keeping the repeated comments gives 10.8% against 8.5%. The gap appears in every month: 2016 was lower in January, February, March and the first days of April alike. Insults fell from 3.7% to 3.0% and obscenity from 2.9% to 2.3%. Identity attacks were rare in both years and did not change (0.17% and 0.15%).

The difference in retrieval is the main threat to this comparison, and the checks I can make do not suggest it drives the result. Within each year, the toxicity of a post's comment section was unrelated to how many of its comments were retrieved (correlations of -0.03 in 2015 and 0.06 in 2016). Restricting the comparison to the 298 posts where at least 80% of comments were retrieved gives a larger difference (2.7 points, $p < 0.001$); restricting it to the 87 posts with at least 90% gives a similar one (1.8 points), though with few posts it is imprecise ($p = 0.15$). If deleted 2015 comments were more often toxic than retrieved ones, the true 2015 rate would be higher and the fall larger. The fall would disappear only if missing 2016 comments were much more toxic than retrieved ones: for the 9,716 missing 2016 comments to close the gap on their own, about 24% of them would have to be toxic, against 8.9% of those retrieved.

Table 1. The Labour Party’s Facebook comment section, 1 January to 12 April, 2015 and 2016.

	2015 (Miliband)	2016 (Corbyn)
Posts	461	131
Comments retrieved	174,560	60,019
Comments reported by Facebook	235,781	72,721
Share retrieved	75%	87%
Median comments per post	324	427
Commenters	30,155	10,056
Share of comments by top 1% of commenters	48%	45%
Commenters who commented once	59%	57%
Replies to other comments	55%	67%
Toxic	10.9% [10.5, 11.3]	8.9% [8.3, 9.4]
Insult	3.7%	3.0%
Identity attack	0.17%	0.15%
Threat	0.11%	0.06%

Toxicity with 95% intervals from bootstrapping posts.

None of this rules out a pattern of deletion I cannot see, but it would have to run against the obvious one.

4.3 Part crowd, part period

Part of the fall came from a change in who was commenting (Figure 2). Comments by accounts seen only in the 2015 window were toxic 11.1% of the time. Comments by accounts seen only in 2016 were toxic 8.5% of the time. This gap is not a matter of activity: within every band of activity, from people who commented once to those who commented more than fifty times, the accounts seen only in 2015 were more toxic, and reweighting them to the activity mix of the 2016 newcomers leaves their rate at 11.1%.

The accounts seen in both windows also changed. Their comments were toxic 10.5% of the time in 2015 and 9.5% in 2016, a fall of 1.0 points (SE 0.45, $p = 0.02$). Averaged by person rather than by comment, the fall is 0.6 points and not distinguishable from zero (95% interval -0.6 to 1.9), and with account fixed effects and controls it is 0.45 points ($p = 0.32$). The safest reading is that the people who stayed became a little less toxic, by between half a point and a point.

How much of the overall fall is due to the change of crowd depends on what that smaller change among stayers represents. If it reflects the period, such as a quieter news agenda or a less heated election calendar, and would have applied to everyone, then applying it to the 2015 crowd predicts a 2016 rate of 9.9%. On that reading, about half of the 2.0-point fall (1.0 points) is due to the change of crowd and about half (1.0 points) to the period. If instead the stayers’ change reflects only their own behaviour, more of the fall is composition. Either way, a substantial part of the fall came from different people commenting, and the people who stayed did not become more toxic.

The most active commenters were less toxic than occasional ones in both years, the opposite of the pattern Kim et al. (2021) found in news comments. Among the most active 1%, 10.5% of comments were toxic in 2015 and 7.9% in 2016, against 12.2% and 9.0% among people who commented once. Within posts, toxic comments were slightly more likely than others to receive at least one like, by about 3 to 4 percentage points in both years.

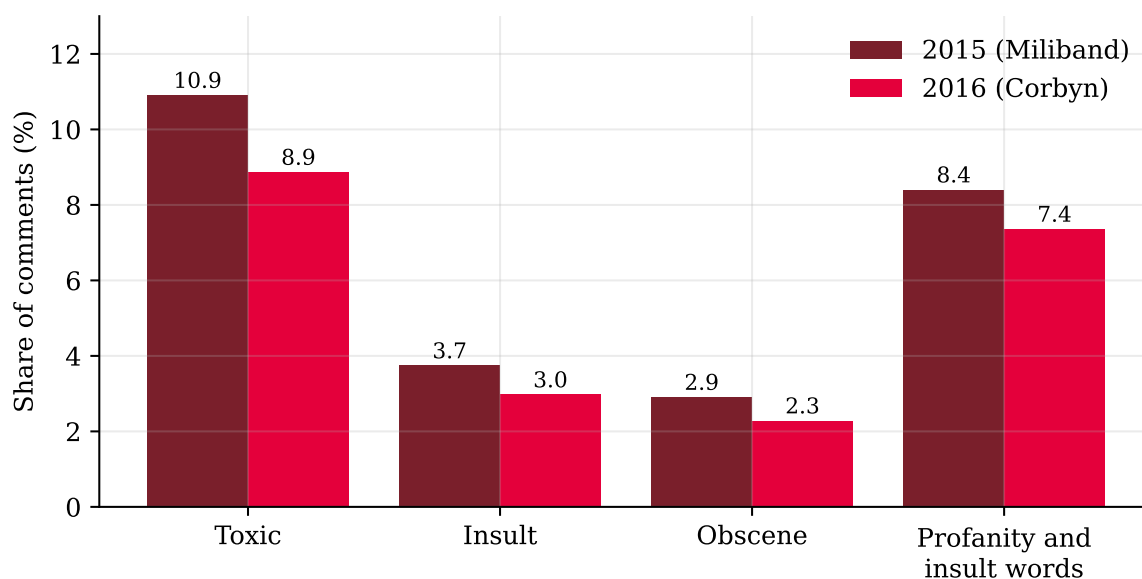


Figure 1. Share of comments classified as toxic, insulting or obscene by Detoxify, and containing words from a profanity list, 2015 and 2016.

4.4 What people argued about

The topics changed more than the tone (Figure 3). In 2015, 9.6% of comments mentioned UKIP or Nigel Farage, and comments about UKIP were among the most often toxic (13%). In 2016, only 3.0% did. Comments about the EU and the coming referendum rose from 3.2% to 10.8%. Comments mentioning the party leader rose from 6.6% about Miliband to 10.0% about Corbyn, and mentions of the Conservatives rose slightly, from 17.9% to 19.8%.

Pejorative labels for Labour’s internal factions rose but stayed rare. “Blairite”, “red Tory” and “Tory-lite” appeared in 0.39% of comments in 2015 (676 comments) and 0.67% in 2016 (401). “Trot”, “hard left” and “loony left” were almost absent in 2015 (0.02%, 41 comments) and appeared in 0.26% in 2016 (156). Neither kind of comment was more often toxic than average in 2016 (6.0% and 6.4%). Broader lists that include neutral terms show the same direction (appendix). In the coded sample, comments about the party’s internal factions appeared only in 2016 (9 of 150 comments, against none of 150 in 2015; $p = 0.003$).

Comments mentioning Jewish people, Israel or antisemitism rose from 0.16% to 0.59% of comments (286 comments in 2015 and 356 in 2016), and about 15% of them were toxic in both years. The rise falls in the weeks when antisemitism in Labour became a public story: the co-chair of Oxford University Labour Club resigned in February 2016 citing antisemitism in the club, and a former party candidate was suspended in March over social media posts. Among comments the classifier flagged as identity attacks, the share mentioning these terms rose from 4% to 20%, but the numbers are very small (12 comments in 2015, 18 in 2016), and in 2016 two accounts wrote 8 of the 18. Reading them shows what the classifier could and could not tell. In 2015, five of the twelve were antisemitic. In 2016, one of the eighteen was antisemitic and one was hostile to Palestinians; ten accused Labour or its leader of antisemitism, several of them while attacking Muslims, and five challenged antisemitism or Holocaust denial by other commenters. The classifier was often right that a comment attacked a group, but wrong about which group. These counts cannot estimate how common antisemitic comments were on the page: identity

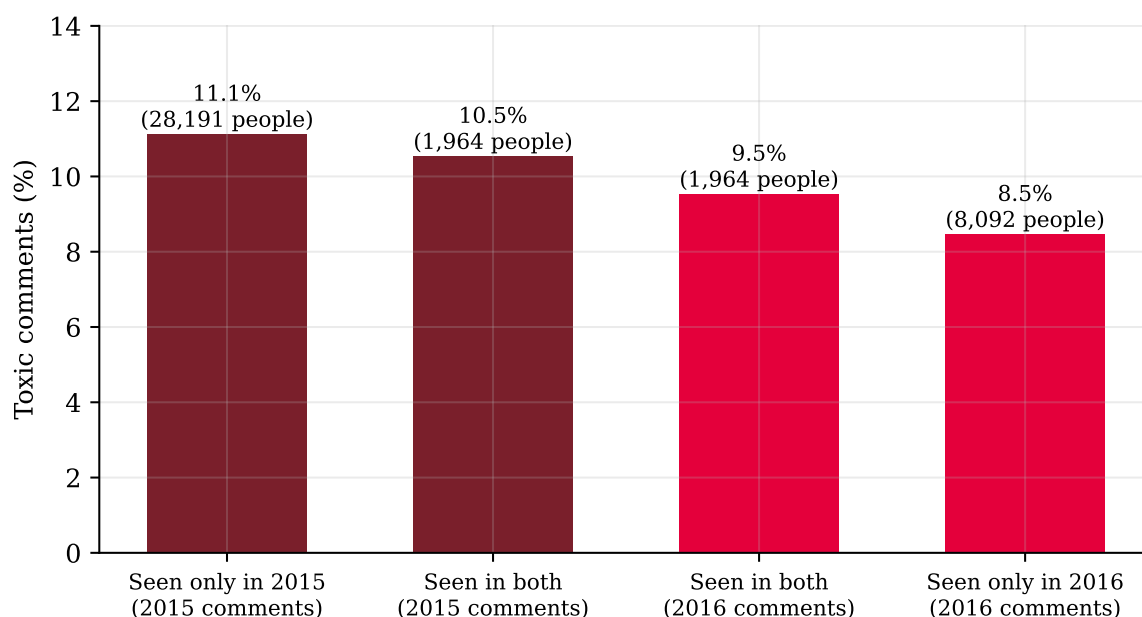


Figure 2. Share of comments that were toxic, by group of commenters. Accounts seen only in the 2015 window, accounts seen in both windows (shown for each year), and accounts seen only in the 2016 window. Groups are defined by the two observation windows.

attacks are rare, the classifier misses many, and the sample is tiny. What they show is that in these weeks the flagged comments were mostly part of an argument about antisemitism in Labour, which the Equality and Human Rights Commission later found to involve unlawful harassment and discrimination by the party (Equality and Human Rights Commission 2020).

4.5 Stance

In the AI-assisted reading of 300 comments, the share critical of Labour or its leader fell from 27% in 2015 to 17% in 2016 ($p = 0.03$), and the share supporting Labour rose slightly, from 27% to 31% ($p = 0.44$). About half the comments in each year did neither: they argued with other commenters, discussed policy or the news, or were off topic. Splitting the sample by group of commenter, 30% of comments by accounts seen only in 2015 were critical of Labour, against 22% by accounts seen in both windows in 2015, 17% by those accounts in 2016, and 17% by accounts seen only in 2016. The accounts that did not return were not simply Labour's own supporters of the Miliband years: a larger share of them were critics. These estimates come from small samples (51 to 99 comments per group), coded with AI assistance and reviewed by one person, and are less certain than the figures from the full data.

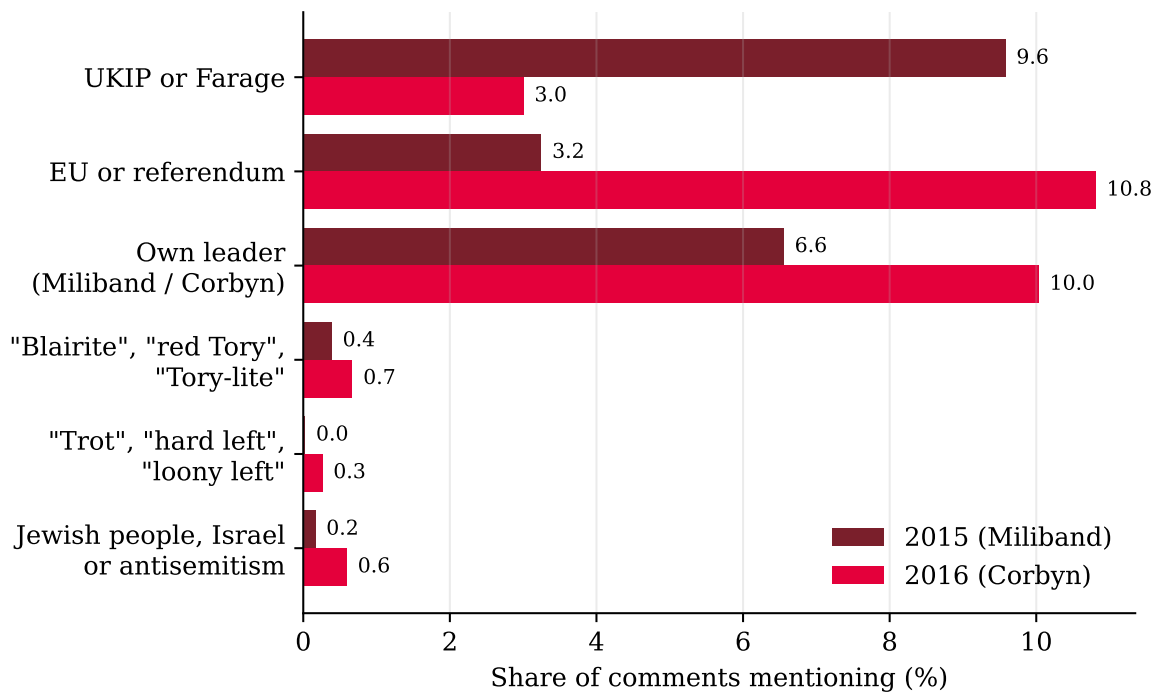


Figure 3. Share of comments mentioning selected topics, 2015 and 2016. Word lists are in the appendix.

5 Discussion

5.1 A partly new crowd, a slightly calmer tone

The comment section on Labour's page did change between Miliband and Corbyn. In the same weeks of the year, it became smaller, more conversational, more concentrated on the leader and the EU, less concerned with UKIP, a little less toxic and a little less critical of the party. About half of the fall in toxicity, and on some assumptions more, came from a change of crowd. The accounts not seen again after the 2015 window had been more toxic, and more often critical of Labour, than those that arrived in 2016, and the accounts seen in both windows became slightly less toxic.

Against the expectations set out earlier, the results are mixed. As expected, replies rose, comments from outside the party appear to have fallen, and the topics moved from the general election and UKIP to the EU and Labour itself. Supportive comments rose only slightly and not significantly. Pejorative factional labels rose, as expected, but stayed rare and were not more toxic than other comments.

This is consistent with what is known about Labour's membership after 2015 ([Whiteley et al. 2019](#); [Bale et al. 2020](#)): a large, committed and sympathetic audience arrived, and part of a more mixed audience, including critics drawn to the page in the run-up to a general election, was not seen again. It is not consistent with the claim that Corbyn's following made this space more hostile, at least for the party's main Facebook page in the first months of 2016. The higher toxicity of the accounts seen only in 2015 is a statement about who was on the page in an election year, many of them critics, and not about Miliband's supporters.

5.2 The arguments moved

What did change was the argument. In 2015 Labour's commenters argued with and about UKIP and the general election. In 2016 they argued about Europe, about Corbyn and, a little more than before, about Labour itself. Factional labels and arguments about antisemitism were still rare in these weeks, but both were rising. A comparison later in 2016, after the referendum, the antisemitism row and the second leadership contest, might find a different picture.

5.3 Leader or calendar?

These results compare two periods. They cannot separate the effect of a new leader from the effect of being three months from a general election rather than three months from local and devolved elections. The 2015 window included a general election campaign that would attract opponents to Labour's page, and the 2016 window included the start of the EU referendum campaign. The drop in UKIP mentions and the rise in EU mentions are probably calendar effects as much as leadership effects. The fall in toxicity is present in every month, including January, months before either campaign; and the comparison of January to March, outside the 2015 short campaign, gives a similar fall. That shows the result is not driven by the last weeks of the election campaign, but not that it is driven by the leader: the whole of early 2015 was a pre-election period. The page also posted far less under Corbyn, so each post gathered a larger and perhaps different audience. These differences are part of what changed between the two periods, and I cannot separate them from the change of leader.

6 Limitations

Beyond the comparison of two periods, four limitations matter. First, the 2015 comments were collected a year after they were written, and a quarter of them were missing. I find no relationship between retrieval and toxicity, and the missing comments would have to be implausibly toxic to reverse the result, but deleted comments cannot be scored. Second, Detoxify is an imperfect measure of incivility. It misreads sarcasm and over-flags heated discussion of identity, as the identity-attack audit shows; it is better at comparing rates across large groups than at classifying single comments. Third, the topic flags are word lists and the stance estimates come from a small AI-assisted sample reviewed by one person. Fourth, this is one page. Labour's Facebook presence included Corbyn's own page, Momentum and many local and supporters' pages, and the comment sections there may have changed differently.

7 Conclusion

Between Ed Miliband's last months and Jeremy Corbyn's first, the comment section of Labour's Facebook page did not become more hostile. It became slightly less toxic and slightly less critical of the party. About half of the fall in toxicity came from a change in who was commenting: the accounts that arrived were less toxic than those not seen again, and those seen in both years became a little less toxic too. What changed most was the argument: from UKIP and the general election to Europe, the leader and, increasingly, Labour itself.

Provenance and AI use

The data were collected with Netvizz for my doctoral research, before Netvizz closed in 2019. This paper builds on my doctoral thesis (Hotham 2021). The analysis and writing were completed in 2026.

Version 4 (October 2026) corrects the word list for mentions of the Conservatives, which had counted any use of “hunt” (including Tristram Hunt, a Labour MP, and phrases such as “witch hunt”); it now counts only Jeremy Hunt. Mentions of the Conservatives in 2016 change from 20.0% to 19.8% of comments, and the word list in the appendix now matches the code. The deposited topic shares are corrected to match. No other reported figure changes.

The analysis code and drafts of the text were prepared with the assistance of an AI system (Claude, by Anthropic), which was also used to review drafts. I checked the analysis and the text and take responsibility for both.

Ethics and data

The comments were collected through Netvizz during my doctoral research at the University of Bath. Commenters’ identities were replaced with one-way hashes at collection. I do not quote or publish comments. The AI-assisted reading of 300 sampled comments and of the 30 identity-attack comments was used only to produce counts. Aggregate data are deposited with this preprint: weekly counts of comments and of toxic, insulting and identity-attack comments, and topic shares by year. No comment text, commenter identifier or post-level identifier is included.

References

- Bale, T., Webb, P., and Poletti, M. (2020). *Footsoldiers: Political Party Membership in the 21st Century*. Routledge, London.
- Coe, K., Kenski, K., and Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4):658–679.
- Dennis, J. (2020). A party within a party posing as a movement? Momentum as a movement faction. *Journal of Information Technology & Politics*, 17(2):97–109.
- Dixon, L., Li, J., Sorensen, J., Thain, N., and Vasserman, L. (2018). Measuring and mitigating unintended bias in text classification. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 67–73.
- Dommett, K. and Temple, L. (2018). Digital campaigning: The rise of Facebook and satellite campaigns. *Parliamentary Affairs*, 71(suppl_1):189–202.
- Equality and Human Rights Commission (2020). Investigation into antisemitism in the Labour party. Technical report, Equality and Human Rights Commission, London.
- Gerbaudo, P. (2019). *The Digital Party: Political Organisation and Online Democracy*. Pluto Press, London.

- Gorrell, G., Bakir, M. E., Roberts, I., Greenwood, M. A., and Bontcheva, K. (2020). Which politicians receive abuse? Four factors illuminated in the UK general election 2019. *EPJ Data Science*, 9:18.
- Hanu, L. and Unitary team (2020). Detoxify. GitHub repository, <https://github.com/unitaryai/detoxify>.
- Hede, A., Agarwal, O., Lu, L., Mutz, D. C., and Nenkova, A. (2021). From toxicity in online comments to incivility in American news: Proceed with caution. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*.
- Hotham, T. (2021). *How Do Political Parties Use Facebook and What Does It Offer to Their Campaigns*. PhD thesis, University of Bath.
- Hotham, T. (2026a). Follow the leader? How British parties divided the work between leader pages and party pages on Facebook at the 2015 and 2017 general elections. Preprint, Zenodo, October 2026, <https://doi.org/10.5281/zenodo.23280205>.
- Hotham, T. (2026b). One city, two comment sections: Toxicity and identity on Sadiq Khan’s and Zac Goldsmith’s Facebook pages, 2016. Preprint, Zenodo, October 2026, <https://doi.org/10.5281/zenodo.23288552>.
- Hotham, T. (2026c). Setting the tone or supplying the words? Donald Trump’s Facebook posts and the hostility of his comment section in the 2016 Republican primaries. Preprint, Zenodo, October 2026, <https://doi.org/10.5281/zenodo.23288125>.
- Hotham, T. (2026d). Who carries the party on Facebook? Leaders, party pages and the polls in Britain, 2010 to 2020. Preprint, Zenodo, October 2026, <https://doi.org/10.5281/zenodo.23263764>.
- Hotham, T. (2026e). Who speaks in the candidate’s comment section? Participation, toxicity and reward on Donald Trump’s and Hillary Clinton’s Facebook pages in 2016. Preprint, Zenodo, October 2026, <https://doi.org/10.5281/zenodo.23287107>.
- Jigsaw and Conversation AI (2018). Toxic comment classification challenge. Kaggle, <https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge>.
- Kalsnes, B., Larsson, A. O., and Enli, G. (2017). The social media logic of political interaction: Exploring citizens’ and politicians’ relationship on Facebook and Twitter. *First Monday*, 22(2).
- Kim, J. W., Guess, A., Nyhan, B., and Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, 71(6):922–946.
- Rieder, B. (2013). Studying Facebook via data extraction: The Netvizz application. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 346–355.
- Rossini, P. (2022). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*, 49(3):399–425.
- Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, 18(2):121–138.
- Sap, M., Card, D., Gabriel, S., Choi, Y., and Smith, N. A. (2019). The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678.

- Southern, R. and Harmer, E. (2021). Twitter, incivility and “everyday” gendered othering: An analysis of tweets sent to UK members of parliament. *Social Science Computer Review*, 39(2):259–275.
- Stromer-Galley, J. (2014). *Presidential Campaigning in the Internet Age*. Oxford University Press, New York.
- Theocharis, Y., Barberá, P., Fazekas, Z., and Popa, S. A. (2020). The dynamics of political incivility on Twitter. *SAGE Open*, 10(2). <https://doi.org/10.1177/2158244020919447>.
- Ward, S. and McLoughlin, L. (2020). Turds, traitors and tossers: The abuse of UK MPs via Twitter. *The Journal of Legislative Studies*, 26(1):47–73.
- Whiteley, P., Poletti, M., Webb, P., and Bale, T. (2019). Oh Jeremy Corbyn! Why did Labour party membership soar after the 2015 general election? *British Journal of Politics and International Relations*, 21(1):80–98.

A Appendix

Table 2. Difference between 2016 and 2015 in percentage points. Linear probability models; controls are log comment length, reply and calendar month; standard errors clustered by post unless stated. The fixed-effects model clusters by account.

Outcome	Difference	SE	Comments
Toxic, no controls	-2.04***	0.35	234,579
Toxic, with controls	-1.83***	0.37	234,579
Toxic, January to March only	-1.72***	0.39	179,764
Toxic, posts with $\geq 90\%$ retrieval	-1.76	1.21	29,198
Toxic, with topic controls	-1.74***	0.36	234,579
Insult, with controls	-0.70***	0.16	234,579
Identity attack, with controls	-0.03	0.02	234,579
Toxic, with controls, two-way clustering (post and account)	-1.83***	0.55	234,579
Toxic, with controls, posts with $\geq 80\%$ retrieval	-2.71***	0.81	–
Toxic, accounts in both years, comment-weighted	-1.00*	0.45	–
Toxic, accounts in both years (account fixed effects)	-0.45	0.45	–

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3. Word lists used to flag comments (case-insensitive, whole words; * marks any ending).

Flag	Terms
Miliband	miliband, milliband, ed m, ed (not followed by ‘balls’), red ed
Corbyn	corbyn*, jeremy (not followed by ‘hunt’), jc
Conservatives	tory, tories, conservative(s), cameron, osborne, boris, gideon, ids, duncan smith, jeremy hunt
UKIP	ukip, farage, kipper(s)
EU	eu, europe*, brexit, referendum, vote leave, leave campaign, remain campaign, in or out
‘Blairite’ and similar, pejorative	blairite(s), red tory/tories, tory-lite
‘Trot’ and similar, pejorative	trot(s), trotskyist(s), hard left, loony left
‘Blairite’ and similar, broad	blair*, new labour, red tory/tories, tory-lite, centrist(s), progress
‘Trot’ and similar, broad	trot(s), trotsk*, momentum, hard left, loony left, communis*, marxis*, socialis*
Jewish people, Israel, antisemitism	jew(s), jewish, antisemit*, anti-semit*, zionis*, israel*
Immigration	immigra*, migrant(s), refugee(s), asylum, foreigner(s), border*