

REVIEW ARTICLE

Echoes of Influence

A Multi-Method Analysis of Political Microtargeting on Polarization in the Digital Age

Lewis Osoro Nyakundi¹

¹ Postgraduate Student at University of Nairobi



ABSTRACT

Political microtargeting has become a defining feature of digital campaigning, yet its role in driving political polarisation remains contested. This paper conducts a secondary, multi-method systematic review of ten empirical, quantitative studies published between 2016 and 2025 that examine political microtargeting, psychological targeting, or online political polarisation. The review draws on field experiments, laboratory and online experiments, computational text modelling, and large-scale survey analysis. Findings show that personality- and issue-congruent microtargeting produces a measurable persuasive advantage at the individual level, sometimes substantial, yet this advantage is neither universal nor stable across contexts. Average persuasive effects of campaign contact approach zero in general elections, and microtargeted messages do not consistently outperform well-crafted generic appeals. At the collective level, evidence linking online exposure directly to rising polarisation is weaker and more conditional than popular narratives suggest; polarisation has grown fastest among demographic groups least exposed to digital media, while experimentally induced exposure to opposing views can increase rather than reduce polarisation. Detection research shows that psychological profiles are recoverable from ordinary online behaviour, yet transparency disclosures and persuasion warnings do not reliably neutralise targeting's persuasive advantage. This paper argues that microtargeting's democratic risk lies less in a uniform manipulative power over mass publics than in a persuasion–polarisation gap: real, conditional, individual-level effects that current safeguards are not yet equipped to govern. Implications for platform accountability and regulatory design are discussed.

Keywords: political microtargeting; political polarisation; digital campaigning; psychological targeting; systematic review

HOW TO CITE

Osoro Nyakundi, L. (2026). Echoes of influence: A multi-method analysis of political microtargeting on polarization in the digital age. *Current Research in Communication*, 1-1, 1-10. <https://doi.org/10.66699/tq481313>

Open access under the Creative Commons Attribution 4.0 International license: creativecommons.org/licenses/by/4.0/

Introduction

Digital platforms have reorganised how political campaigns reach voters. Where twentieth-century campaigns broadcast a single message to a mass electorate, twenty-first-century campaigns increasingly assemble thousands of micro-audiences, each shown a message tuned to their psychology, behaviour, or data trail (Kreiss, 2016; Turow, 2011). This practice, political microtargeting, combines behavioural data, algorithmic profiling, and platform advertising infrastructure to match persuasive appeals to individual traits (Zuboff, 2019). Since the Cambridge Analytica disclosures of 2018, political microtargeting has moved from a technical curiosity to a matter of public alarm, associated in popular discourse with manipulated elections, fractured public discourse, and rising political polarisation.

Yet the empirical record is more divided than the public conversation implies. Some studies report that personality-congruent advertising substantially increases clicks, purchases, and favourable attitudes ([Matz et al., 2017](#); [Zarouali et al., 2022](#)). Others find that campaign contact of almost any kind, targeted or not, moves few voters in general elections ([Kalla & Broockman, 2018](#)), and that microtargeted political messages generated by large language models confer no reliable advantage over untargeted ones ([Hackenburg & Margetts, 2024](#)). A parallel literature on political polarisation is similarly unsettled. Some evidence associates exposure to opposing partisan content with intensified, rather than reduced, polarisation ([Bail et al., 2018](#)); other evidence finds that internet and social media use are not the primary drivers of polarisation's recent growth ([Boxell et al., 2017](#)). Two questions therefore remain open, and both concern this paper directly. First, how strong and how conditional is microtargeting's persuasive effect on individuals. Second, whether that individual-level effect translates into the macro-level phenomenon of political polarisation so often attributed to it.

This paper addresses both questions through a secondary, multi-method systematic review. It synthesises ten empirical, quantitative studies that meet strict inclusion criteria, spanning field experiments, controlled online and laboratory experiments, computational modelling of behavioural data, and nationally representative survey analysis. The breadth of method is deliberate. Persuasion and polarisation are measured differently across disciplines, and a review confined to a single method would understate the complexity of the phenomenon it studies. Sociological and communication research on the social determinants of political behaviour further suggests that persuasion effects are rarely uniform; they depend on group identity, prior attitude strength, and the institutional context in which messages are received (Sunstein, 2017). A multi-method synthesis is therefore not simply a methodological convenience. It is the appropriate design for a question that no single method can settle alone.

This paper argues, and defends across the sections that follow, that political microtargeting produces a real but narrow and conditional persuasive advantage at the individual level. That advantage does not translate reliably into the direct, uniform driver of societal polarisation that much public and policy discourse assumes. Where microtargeting appears best placed to reshape opinion, personality-congruent framing, low platform transparency, and high issue salience, it does so through specific psychological mechanisms rather than through a general manipulative capacity over the electorate. This paper terms the space between demonstrated individual-level persuasion and unproven collective-level polarisation the persuasion–polarisation gap. Understanding that gap, this paper contends, is more useful for policy and for further research than either dismissing microtargeting as inert or treating it as an unqualified threat to democracy.

The remainder of this paper proceeds as follows. Section 2 sets out the theoretical framework, drawing on selective-exposure theory, congruence-based persuasion theory, and the political-economy literature on data-driven campaigning. Section 3 describes the review's methods, including the search strategy, inclusion criteria, and synthesis approach. Section 4 presents the findings, organised around a synthesis matrix of the ten reviewed studies and four thematic subsections. Section 5 discusses the findings in relation to the theoretical framework and the paper's central claim, before Section 6 considers the review's limitations and Section 7 concludes.

Theoretical Framework

Three bodies of theory inform this paper's reading of the evidence.

Selective exposure and the structure of online publics

Selective-exposure theory holds that people prefer information that confirms existing beliefs and avoid information that challenges them (Sunstein, 2017). Digital platforms are widely thought to intensify this tendency by algorithmically sorting users into like-minded networks, or echo chambers, that limit exposure to opposing views. Sunstein (2017) argues that this sorting weakens the shared conversational ground that democratic deliberation requires. Selective-exposure theory predicts a straightforward relationship: more sorting produces more polarisation. The theory is less clear, however, about what happens when sorting is interrupted, a question that field experiments on cross-cutting exposure test directly ([Bail et al., 2018](#)).

Congruence-based persuasion

A second body of theory concerns why tailored messages might persuade more effectively than generic ones. Self-congruity theory and related work in the Elaboration Likelihood tradition suggest that messages matching a receiver's existing psychological traits, values, or self-concept are processed more favourably and provoke less counter-argument ([Zarouali et al., 2022](#)). Political microtargeting operationalises this theory at scale: by inferring a voter's personality or issue priorities from behavioural data, campaigns can match message framing to the individual rather than the mass audience ([Matz et al., 2017](#)). Congruence theory predicts that microtargeting should outperform generic messaging under most conditions. Recent evidence complicates this prediction, since well-crafted generic messages have, in at least one study, matched or exceeded the persuasive power of individually tailored ones ([Hackenburg & Margetts, 2024](#)).

Data-driven campaigning and the political economy of targeting

A third body of theory situates microtargeting within the broader political economy of data collection. Turow (2011) describes an advertising industry that increasingly classifies individuals by predicted commercial and political worth, using data gathered without informed consent. Zuboff (2019) extends this account, arguing that the routine extraction of behavioural data by digital platforms constitutes a novel form of economic and political power, one she terms surveillance capitalism, that makes microtargeting's underlying data infrastructure possible. Kreiss (2016) documents how political parties adopted this infrastructure unevenly, with consequences for the technological sophistication and likely persuasive reach of microtargeting campaigns across the political spectrum. This literature does not itself measure persuasive or polarising effects; it explains why the capacity for microtargeting exists and why its use has grown so quickly. It also, per [Persily and Tucker \(2020\)](#), cautions against treating any single mechanism, microtargeting included, as a sufficient explanation for the health of contemporary democratic discourse, given the many other economic, institutional, and psychological forces operating on the same digital terrain.

Read together, these three bodies of theory generate a testable expectation and a genuine tension. Congruence theory predicts consistent, replicable individual-level persuasion from well-targeted messages. Selective-exposure theory predicts that platform sorting compounds this persuasion into collective polarisation. The political-economy literature predicts that both processes are enabled by a data infrastructure whose reach continues to expand. What the theory does not resolve is the strength of the causal link between the first process and the second: whether individual-level persuasion accumulates into population-level polarisation, or whether the two remain empirically distinct. Resolving that tension is the task of the empirical review that follows.

Methods

Research design

This paper undertakes a secondary, multi-method systematic review of quantitative empirical research on political microtargeting and political polarisation. A systematic review was judged appropriate because primary claims in this field are frequently cited selectively, and a transparent synthesis of comparable empirical evidence is needed to evaluate microtargeting's role in polarisation on its evidentiary merits rather than on its public salience.

Search strategy

Searches were conducted across Scopus, Web of Science, PsycINFO, and Google Scholar, supplemented by citation tracking from key review articles in the field. Search terms combined the concepts “political microtargeting,” “psychological targeting,” “personalised political advertising,” and “targeted political persuasion” with “polarisation,” “affective polarisation,” “persuasion effect,” and “digital campaigning.” Searches were restricted to English-language, peer-reviewed journal articles published between January 2013, the year field-experimental research on campaign persuasion gained prominence, and mid-2026.

Inclusion and exclusion criteria

Studies were included if they met four criteria: (a) publication in a peer-reviewed, Scopus- or Web of Science-indexed journal; (b) an explicit empirical, quantitative design, whether experimental, quasi-experimental, computational-modelling, or large-sample survey; (c) a direct focus on political microtargeting, psychologically tailored persuasion, or online political polarisation, rather than a general discussion of digital campaigning; and (d) reporting of comparable, extractable outcome measures, such as effect sizes, percentage differences, or classification accuracy. Purely theoretical, doctrinal, qualitative, or legal-regulatory papers were excluded from the primary evidence corpus, although several are cited in the theoretical framework and discussion for interpretive context. Studies using simulated or hypothetical scenarios without measurable behavioural or attitudinal outcomes were also excluded.

Study selection

Applying these criteria narrowed a broader literature to ten studies that directly meet all four conditions and offer comparable, quantitative outcome data. The final corpus spans 2016 to 2025 and includes three field experiments, four online or laboratory experiments, two computational text-modelling studies, and one large-sample survey-based analysis. This range of method was retained deliberately, consistent with the multi-method design signalled in the paper's title, rather than narrowed to a single dominant method, because persuasion and polarisation are constructs that no single method captures in full.

Synthesis approach

Each included study was coded for its stated objective or claim, its method and sample, and its principal quantitative finding. These codes are presented in Table 1. Findings were then grouped thematically into four analytic categories, developed inductively from the corpus itself: the individual-level persuasive power of microtargeting, the boundary conditions that limit that power, the relationship between individual persuasion and collective polarisation, and the effectiveness of detection and transparency safeguards. This thematic structure, rather than a chronological or purely methodological one, best serves the paper's central claim about the gap between demonstrated persuasion and assumed polarisation.

Findings

Table 1 presents the ten studies included in this review, organised by author, stated claim or objective, and principal quantitative finding. The subsections that follow interpret this evidence thematically.

Author(s) & Year	Claim / Objective	Key Findings
Matz, Kosinski, Nave, & Stillwell (2017)	Test whether psychologically tailored (personality-matched) advertising increases real-world persuasion at scale.	Three field experiments (>3.5 million exposures); personality-matched appeals produced up to 40% more clicks and up to 50% more purchases than mismatched or generic appeals.
Zarouali, Dobber, De Pauw, & de Vreese (2022)	Test the causal persuasive effect of personality-congruent political microtargeting, using an automated personality-profiling algorithm.	Two experiments; personality-congruent political advertisements, trait- and affect-based, produced significantly stronger persuasion than incongruent advertisements.
Kruikemeier, Sezgin, & Boerman (2016)	Examine how exposure to personalised political Facebook advertising shapes voters' responses via persuasion-knowledge activation.	Online experiment (N = 122); personalised ads activated persuasion knowledge and reduced word-of-mouth intentions only when the "Sponsored" label was recalled; no effect on source trust.
Kalla & Broockman (2018)	Estimate the average persuasive effect of campaign contact, including targeted appeals, on candidate choice in general elections.	Meta-analysis of 40 studies plus nine original field experiments (49 total); average persuasive effect statistically indistinguishable from zero, except under narrow conditions.
Bail et al. (2018)	Test experimentally whether exposure to opposing partisan views on social media reduces affective or attitudinal polarisation.	One-month Twitter bot field experiment; Republicans exposed to counter-attitudinal content became significantly more conservative; Democrats showed a small, non-significant liberal shift.
Boxell, Gentzkow, & Shapiro (2017)	Test whether greater internet and social media use explains the recent acceleration of political polarisation across demographic groups.	Composite eight-measure polarisation index using national survey data; polarisation grew fastest among groups least likely to use the internet, implying a limited direct role for online media.
Simchon, Sutton, Edwards, & Lewandowsky (2023)	Test whether personality can be inferred from content people consume, not only produce, online, as a basis for detecting psychological microtargeting.	Two BERT NLP models on Reddit data (N = 1,105; N = 10,050); personality was predictable from consumed text ($r = .13$) as well as produced text ($r = .33$).
Jansen & Krämer (2023)	Test whether more salient targeting-disclosure labels increase users' recognition of political microtargeting and affect privacy-related outcomes.	Preregistered experiment (N = 450); a more salient disclosure increased recognition of targeting, correctly recalled by over four-fifths of participants; recognition raised perceived benefit but not privacy concern.
Hackenburg & Margetts (2024)	Test whether GPT-4-generated, microtargeted political messages are more persuasive than non-targeted generic messages.	Pre-registered online experiment; messages were broadly persuasive (up to 12 percentage points), but microtargeted messages were not significantly more persuasive than non-targeted ones (4.83 vs 6.20 points, $p = .226$).
Carrella, Simchon, Edwards, & Lewandowsky (2025)	Test whether warning users that an advertisement is personality-targeted eliminates its persuasive advantage.	Three within-subject experiments (N = 666; 432; 669); targeted ads remained significantly more persuasive than non-targeted ones even after a warning pop-up, which had no meaningful mitigating effect.

Table 1. Synthesis of ten empirical, quantitative studies on political microtargeting and political polarisation.

The persuasive power of microtargeting at the individual level

Four of the ten reviewed studies demonstrate that microtargeting produces a measurable, sometimes substantial, persuasive advantage when campaigns match message content to individual traits. [Matz et al. \(2017\)](#) provide the clearest evidence of scale. Across three field experiments reaching more than 3.5 million people, advertisements matched to a person's extraversion or openness produced up to 40% more clicks and up to 50% more purchases than mismatched or generic appeals. Although this is a commercial and psychological-persuasion study rather than

a strictly electoral one, its scale and its behavioural, not merely attitudinal, outcome measures make it a central data point for the persuasive capacity that political campaigns can, in principle, exploit.

Zarouali et al. (2022) extend this evidence into the political domain directly. Using an automated personality-profiling algorithm applied to participants' own text, two experiments found that political advertisements congruent with a person's inferred personality, and separately with their inferred emotional disposition, produced significantly stronger persuasion than incongruent advertisements. Carrella et al. (2025) replicate this congruence effect across three further experiments and add an important extension: personality-targeted political advertisements remained more persuasive than non-targeted ones even when participants were warned, via an on-screen pop-up, that they were being targeted.

Kruikemeier et al. (2016) offer a more qualified picture. Personalised political advertisements on Facebook activated what the authors term persuasion knowledge, participants' awareness that they were being persuaded, which in turn reduced their willingness to recommend the advertised party to others. Critically, this effect emerged only among participants who recalled seeing the platform's "Sponsored" label; personalisation exerted no significant effect on how trustworthy the message source appeared. Together, these four studies support congruence-based persuasion theory's central prediction: tailoring a message to a receiver's psychological profile increases its persuasive power. They do not, however, support the assumption that this power is unconditional, invisible to its targets, or immune to platform-level cues.

The limits of persuasion: null and near-zero effects

Two studies in the reviewed corpus complicate the picture painted above. Kalla and Broockman (2018) offer the most comprehensive test of campaign persuasion available in the literature: a meta-analysis of 40 existing field experiments combined with nine original experiments, forty-nine studies in total. Their headline finding is stark. The average persuasive effect of campaign contact, including direct mail, phone calls, and canvassing, on candidate choice in general elections is statistically indistinguishable from zero. Persuasive effects emerged only under two narrow conditions: when candidates took unusually unpopular positions and campaigns had specifically identified persuadable voters, or when contact occurred well before election day and was measured immediately, then decayed. This study did not test microtargeting specifically, but its scale and rigour make it an essential benchmark against which any claim about targeted persuasion's political power should be judged.

Hackenburg and Margetts (2024) apply a comparable logic to the newest generation of microtargeting technology. Using GPT-4 to generate political messages tailored in real time to individual respondents' demographic and political data, the authors found that these messages were broadly persuasive, in some cases raising support for a policy position by as much as twelve percentage points. Crucially, however, microtargeted messages were not statistically more persuasive than non-targeted, generic messages generated by the same model (4.83 versus 6.20 percentage points, $p = .226$). This finding directly challenges what the authors term the "manipulation machine" narrative: the assumption that automated personalisation confers a decisive persuasive advantage over well-crafted generic messaging.

Read alongside the previous subsection, these findings suggest that microtargeting's persuasive advantage, where it exists, is narrower and more context-dependent than either its proponents in the campaign industry or its critics in public discourse typically claim. Persuasion appears to depend more on message quality and, in electoral settings, on identifying genuinely persuadable individuals, than on targeting technology as such.

From individual persuasion to collective polarisation

The relationship between microtargeting's individual-level effects and society-wide political polarisation is the least settled question in the reviewed corpus, and the two studies that address it directly point in different directions. Bail et al. (2018) tested the most intuitive proposed remedy for polarisation: exposing partisans to opposing views. In a month-long Twitter field experiment, Republican participants who followed a bot retweeting liberal content

became significantly more conservative after the intervention, and Democratic participants showed a small, statistically non-significant shift towards more liberal attitudes after following a conservative bot. Rather than depolarising participants, cross-cutting exposure appeared to entrench, or in the Republican sample intensify, existing partisan positions. This finding, drawn from real platform behaviour rather than a laboratory setting, is difficult to reconcile with a simple echo-chamber account in which more exposure to opposing views would be expected to moderate attitudes.

[Boxell et al. \(2017\)](#) approach the polarisation question from a different angle, using nationally representative survey data to test whether internet and social media use explain the recent growth in polarisation. Constructing a composite index from eight established polarisation measures, the authors found that polarisation increased fastest among the demographic groups least likely to use the internet, older Americans in particular, and more slowly among the heaviest internet users. This pattern is difficult to reconcile with a straightforward causal story in which digital platforms, and by extension the microtargeted advertising they carry, are the primary engine of rising polarisation.

Taken together, these two studies do not resolve whether microtargeting contributes to polarisation; they narrow the plausible mechanism. [Bail et al. \(2018\)](#) suggests that where digital exposure does affect polarisation, it may do so partly through specific, counterintuitive psychological pathways, such as identity-threat responses to opposing content, rather than through simple message exposure. [Boxell et al. \(2017\)](#) suggests that whatever these pathways are, they cannot be the dominant explanation for polarisation's overall trajectory, since the groups least exposed to digital platforms have polarised fastest. This is the empirical core of what this paper terms the persuasion–polarisation gap: demonstrated individual-level persuasive effects coexist with an absence of clear evidence that digital exposure, targeted or otherwise, is the principal driver of collective polarisation.

Detection, transparency, and the limits of current safeguards

A final cluster of studies examines whether microtargeting can be detected, disclosed, or resisted, questions with direct regulatory relevance. [Simchon et al. \(2023\)](#) provide a proof of concept for detection. Training natural language processing models on Reddit data from over ten thousand users, the authors found that personality traits could be predicted not only from the text people produce, a relatively well-established finding, but also from the text people merely consume. Although the consumed-text model performed less accurately than the produced-text model ($r = .13$ versus $r = .33$), the finding demonstrates that the behavioural signals microtargeting relies on are, in principle, detectable and auditable by independent researchers.

[Jansen and Krämer \(2023\)](#) test a leading policy response to microtargeting: mandatory disclosure labels. In a preregistered experiment, a more salient disclosure label increased participants' recognition that an advertisement was targeted, and more than four-fifths of participants who saw the salient label recalled it correctly. However, this increased recognition raised participants' perceived benefit from the advertisement without correspondingly raising their privacy concern; only participants' independent privacy concern, not their recognition of targeting, predicted whether they subsequently engaged in privacy-protective behaviour. Disclosure, in other words, informs without necessarily protecting. [Carrella et al. \(2025\)](#), discussed above, reinforce this conclusion from a different angle: an explicit warning that an advertisement was personality-targeted did not reduce its persuasive advantage over non-targeted content.

Read together, these three studies suggest that the technical capacity to detect microtargeting is outpacing the practical capacity of transparency measures to neutralise its effects. Detection is a necessary precondition for accountability, but the reviewed evidence gives little reason to expect that disclosure or warning labels, in their current form, are sufficient to protect individuals from microtargeting's demonstrated persuasive advantage.

Discussion

Reassessing the claim in light of the evidence

This paper set out to test two related questions: how strong is microtargeting's persuasive effect on individuals, and does that effect translate into the collective phenomenon of political polarisation. The reviewed evidence supports a qualified answer to both. Microtargeting's persuasive effect is real. Personality- and issue-congruent messages consistently outperform mismatched or generic ones across experimental settings ([Carrella et al., 2025](#); [Kruikemeier et al., 2016](#); [Matz et al., 2017](#); [Zarouali et al., 2022](#)), and this effect survives at least one direct attempt at mitigation through warning labels ([Carrella et al., 2025](#)). At the same time, that effect is narrower than the term “manipulation machine” implies. Campaign contact in general elections moves almost no voters on average ([Kalla & Broockman, 2018](#)), and the newest, most sophisticated form of microtargeting, large-language-model-generated political messaging, confers no reliable advantage over well-written generic messaging ([Hackenburg & Margetts, 2024](#)).

The relationship between this individual-level persuasion and collective polarisation is weaker still. No study in this review demonstrates a direct causal pathway from exposure to microtargeted political advertising to population-level increases in polarisation. The two studies that speak most directly to collective-level dynamics point towards complexity rather than confirmation: cross-cutting exposure can increase polarisation under some conditions ([Bail et al., 2018](#)), while the demographic groups least exposed to the internet have polarised fastest overall ([Boxell et al., 2017](#)). This asymmetry, strong individual-level persuasion evidence set against weak collective-level polarisation evidence, is the empirical foundation of this paper's central claim. Microtargeting's principal democratic risk, on the evidence reviewed here, is not that it reliably manufactures polarised publics. It is that it reliably, and often invisibly, moves individual opinion and behaviour under specific, identifiable conditions, congruent framing, low platform transparency, and insufficiently robust user safeguards, that current regulatory tools have not yet caught up with.

Engaging the alternative view

A different reading of the same evidence is possible, and deserves consideration on its own terms. Scholars working in the political-economy tradition represented in this paper's theoretical framework (Turow, 2011; Zuboff, 2019) would caution against separating individual-level persuasion from collective-level outcomes too sharply. On this view, even small, repeated, individually invisible nudges, aggregated across millions of targeted messages and sustained over years of campaigning, could plausibly reshape a political culture in ways that no single experiment is designed to detect. Kreiss (2016) similarly suggests that the uneven adoption of data-driven campaigning across political parties may itself become a structural source of asymmetric influence, regardless of any single message's measured persuasive effect. Sunstein (2017) would add that the absence of a clean experimental link between exposure and polarisation does not rule out a slower, cumulative erosion of the shared informational ground that democratic deliberation depends on, an erosion that short-duration experiments such as [Bail et al. \(2018\)](#) are poorly positioned to capture.

These are reasonable objections, and this paper does not dismiss them. What the reviewed evidence supports is a narrower claim: that the specific, direct, exposure-to-polarisation pathway most often invoked in public debate is not well evidenced by the quantitative literature currently available. It remains possible that slower, more diffuse, or more structural mechanisms are at work, but such mechanisms require different research designs, most plausibly long-term panel studies and platform-level data access, than the experimental and cross-sectional studies reviewed here.

Implications for policy and platform governance

The persuasion–polarisation gap identified in this review has a practical implication. Policy responses framed as depolarisation measures, such as encouraging exposure to opposing views, may not achieve their intended effect and could, per [Bail et al., 2018](#), backfire under some conditions. Policy responses framed instead around individual

protection, meaningful disclosure, persuasion literacy, and independent auditing of targeting systems, address the mechanism the evidence most strongly supports: persuasion, not polarisation, as microtargeting's most demonstrable effect. The finding that current disclosure labels raise awareness without reliably raising protective behaviour ([Jansen & Krämer, 2023](#)), and that explicit warnings do not eliminate targeting's persuasive advantage ([Carrella et al., 2025](#)), suggests that transparency alone is an insufficient safeguard. Detection research ([Simchon et al., 2023](#)) indicates that independent auditing of targeting practices is technically feasible; the reviewed evidence suggests it is also necessary.

Limitations

This review has several limitations that qualify its claim. First, as a secondary synthesis, it depends entirely on the methodological choices of the original studies. The ten reviewed studies operationalise “persuasion” and “polarisation” differently, ranging from click-through and purchase behaviour ([Matz et al., 2017](#)) to attitude-scale shifts ([Bail et al., 2018](#)) to composite survey indices ([Boxell et al., 2017](#)), which complicates direct comparison across studies despite the thematic synthesis attempted in Section 4. Second, the corpus is weighted towards experimental designs conducted in the United States and Western Europe, with participant samples that are disproportionately young, educated, and internet-literate. The extent to which these findings generalise to other political systems remains an open empirical question. Third, this review's strict inclusion criteria, while necessary for comparability, excluded a substantial body of qualitative and doctrinal research on microtargeting's legal and normative dimensions; that literature informs the theoretical framework but is not part of the primary evidence base evaluated in Section 4. Fourth, the field is moving quickly. Two of the ten reviewed studies ([Carrella et al., 2025](#); [Hackenburg & Margetts, 2024](#)) concern large-language-model-based microtargeting, a technology that did not exist when the earliest studies in this review, such as [Kruikemeier et al., 2016](#), were conducted; findings regarding this newer technology should be treated as provisional pending replication. Finally, this paper's synthesis, while structured around explicit inclusion criteria, involved interpretive judgement in grouping studies thematically; readers should treat Table 1 and the interpretation that follows it as one defensible reading of the evidence rather than the only possible one.

Conclusion

This paper evaluated, through a secondary, multi-method systematic review of ten quantitative empirical studies, whether political microtargeting drives political polarisation in the digital age. The evidence supports a claim more precise than either the alarmed public narrative or a dismissive null-effects narrative would suggest. Microtargeting produces a genuine, sometimes substantial, persuasive effect at the individual level when messages are congruent with a person's inferred psychological traits ([Carrella et al., 2025](#); [Matz et al., 2017](#); [Zarouali et al., 2022](#)), yet this effect is bounded: it does not reliably exceed the persuasive power of well-crafted generic messaging ([Hackenburg & Margetts, 2024](#)), and campaign contact broadly moves few voters in competitive elections ([Kalla & Broockman, 2018](#)). The evidence linking this individual-level persuasion to collective political polarisation is weaker still, and in places contradicts the intuitive causal story: polarisation has grown fastest among those least exposed to digital platforms ([Boxell et al., 2017](#)), and deliberately increasing exposure to opposing views can intensify rather than soften partisan attitudes ([Bail et al., 2018](#)).

This paper has termed the space between these two findings the persuasion–polarisation gap: robust evidence of individual-level effects, alongside an absence of comparably robust evidence for the direct, collective-level effect so often assumed to follow from it. Recognising this gap does not mean that microtargeting is democratically inconsequential. Detection research confirms that the psychological signals underlying microtargeting are identifiable in ordinary online behaviour ([Simchon et al., 2023](#)), while transparency and warning-based safeguards, tested directly in this corpus, have so far failed to neutralise targeting's persuasive advantage ([Carrella et al., 2025](#); [Jansen & Krämer, 2023](#)). Future research should prioritise longitudinal, cross-national designs and platform-level data access that can test the slower, more structural mechanisms that critics of microtargeting reasonably suspect,

but that the currently available experimental literature is not designed to detect. Until such evidence emerges, policy attention is better directed at governing microtargeting's demonstrated persuasive power than at an unproven polarising one.

References

- Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press.
- Persily, N., & Tucker, J. A. (Eds.). (2020). *Social media and democracy: The state of the field, prospects for reform*. Cambridge University Press. <https://doi.org/10.1017/9781108890960>
- Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
- Turow, J. (2011). *The daily you: How the new advertising industry is defining your identity and your worth*. Yale University Press.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>
- Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences*, 114(40), 10612–10617. <https://doi.org/10.1073/pnas.1706588114>
- Carrella, F., Simchon, A., Edwards, M., & Lewandowsky, S. (2025). Warning people that they are being microtargeted fails to eliminate persuasive advantage. *Communications Psychology*, 3(1), 15. <https://doi.org/10.1038/s44271-025-00188-8>
- Hackenburg, K., & Margetts, H. (2024). Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, 121(24), e2403116121. <https://doi.org/10.1073/pnas.2403116121>
- Jansen, M.-P., & Krämer, N. C. (2023). Balancing perceptions of targeting: An investigation of political microtargeting transparency through a calculus approach. *PLoS ONE*, 18(12), e0295329. <https://doi.org/10.1371/journal.pone.0295329>
- Kalla, J. L., & Broockman, D. E. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review*, 112(1), 148–166. <https://doi.org/10.1017/S0003055417000363>
- Kruikemeier, S., Sezgin, M., & Boerman, S. C. (2016). Political microtargeting: Relationship between personalized advertising on Facebook and voters' responses. *Cyberpsychology, Behavior, and Social Networking*, 19(6), 367–372. <https://doi.org/10.1089/cyber.2015.0652>
- Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48), 12714–12719. <https://doi.org/10.1073/pnas.1710966114>
- Simchon, A., Sutton, A., Edwards, M., & Lewandowsky, S. (2023). Online reading habits can reveal personality traits: Towards detecting psychological microtargeting. *PNAS Nexus*, 2(6), pgad191. <https://doi.org/10.1093/pnasnexus/pgad191>
- Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2022). Using a personality-profiling algorithm to investigate political microtargeting: Assessing the persuasion effects of personality-tailored ads on social media. *Communication Research*, 49(8), 1066–1091. <https://doi.org/10.1177/0093650220961965>