

Ideological Alignment, Cognitive Biases, and Media Exposure in Political Outgroups and Polarization: A Systematic Review

Jon Robbie C. Canque*

Davao Oriental State University, Mati City, Davao Oriental, Philippines

*Corresponding Author

DOI: <https://doi.org/10.51244/IJRSI.2026.1308000209>

Received: 24 August 2026; Accepted: 29 August 2026; Published: 15 September 2026

ABSTRACT

Political communication research increasingly theorizes ideological alignment, cognitive biases, and media exposure as a sequential causal chain shaping attitudes toward political outgroups and polarization, yet whether the field empirically tests this three-variable model or only invokes it rhetorically has not been systematically examined. This mixed-methods systematic review synthesizes empirical evidence published between January 2000 and 2026 across four databases (Scopus, Web of Science, PsycINFO, PubMed) using an Antecedent–Mechanism–Outcome (AMO) framework and PRISMA 2020 reporting standards, integrating quantitative gap-mapping with qualitative thematic synthesis of how each construct in the chain is conceptualized. A Full-Chain AMO Inclusion Rule retained 158 studies for primary synthesis from 784 records identified; 412 partial-pathway studies were retained as a contextual pool for gap-mapping. Of 158 included studies, 65 (41.1%) satisfied the rule through cognitive-bias mechanisms alone without measuring media exposure; 50 (31.6%) measured media exposure only by self-report; 36 (22.8%) manipulated it experimentally; and 8 (5.1%) used behavioral or trace-data measurement. Only 22.8% formally tested mediation, and 68.4% sampled exclusively from the United States, raising questions about cross-national generalizability. Thematic synthesis revealed that apparent convergence on shared construct labels masks substantive disagreement about what each construct is and where it sits in the information-processing stream. The field’s integrated theoretical model is not the empirical model it predominantly tests. Closing this gap — through behavioral media-exposure measurement, formal mediation estimation, longitudinal designs, and cross-national replication including in hybrid-media contexts such as the Philippines — is the field’s most actionable priority.

Keywords: affective polarization, motivated reasoning, meta-perception, mixed-methods synthesis, Philippine politics

INTRODUCTION

Political polarization has become one of the defining concerns of contemporary democratic life. Across political science, communication studies, and social psychology, three explanatory frameworks dominate scholarly accounts of how mass publics come to dislike, distrust, and dehumanize their political opponents. The first is ideological alignment and partisan sorting, which structure citizens’ political worldviews and locate them in increasingly distinct social and political camps (Iyengar et al., 2012; Mason, 2018; Westwood et al., 2017). The second is cognitive biases — particularly motivated reasoning, selective exposure, identity-protective cognition, and meta-perception — which shape how partisan information is processed once encountered (Derreumaux et al., 2022; Moore-Berg et al., 2020; Taber and Lodge, 2006). The third is media exposure, which in an era of partisan outlets, social platforms, and algorithmically curated feeds supplies the informational environment in which the first two processes operate (Bail et al., 2018; Levendusky, 2013; Piccardi et al., 2024; Stroud, 2010).

These three families are routinely theorized as a sequential causal chain: ideological identity selects citizens into congenial information environments; congenial environments confirm prior beliefs through motivated processing; confirmed beliefs strengthen identity. The chain is the field’s reigning conceptual architecture, underwriting major theoretical claims about echo chambers, filter bubbles, affective polarization, and the

democratic costs of partisan media (Garrett, 2019; Stroud, 2010; Wojcieszak and Garrett, 2018). Yet whether the field actually tests this three-variable architecture — or only invokes it rhetorically while testing simpler two-variable alternatives — has not been addressed by prior systematic syntheses. Existing reviews of affective polarization (Iyengar et al., 2019), selective exposure (Stroud, 2017), and motivated reasoning (Kahan, 2016) have each illuminated one segment of the chain, but none has synthesized whether the segments fit together as the integrated theory requires. Three further limitations motivate this synthesis: prior reviews have generally taken shared construct labels at face value; have under-examined methodological asymmetries within the field; and have largely been confined to U.S. samples.

The Philippine case illustrates why these limitations matter beyond the United States. Despite a rapidly evolving partisan-media ecosystem, mass mobilizations on competing sides, and active social-media-mediated political identity formation, the country has not been studied through an integrated AMO framework — making it both a substantive priority for cross-national replication and a revealing test of the field's three-variable theory in a multiparty, personalist, hybrid-media context.

This review addresses these limitations through a mixed-methods systematic synthesis of empirical evidence published between January 2000 and 2026. The review adopts an Antecedent–Mechanism–Outcome (AMO) framework in which ideological alignment (A) is the antecedent, cognitive biases (M_1) and media exposure (M_2) operate as parallel mechanisms, and outgroup attitudes and polarization (O) constitute the downstream outcome. A Full-Chain AMO Inclusion Rule retained only studies operationalizing at least one A, at least one M (M_1 or M_2), and at least one O construct within a single empirical paper; the rule's permissive treatment of the mechanism stage is deliberate, allowing the field's actual measurement choices to be observed rather than assumed. The synthesis is guided by three research questions: (RQ1) what the empirical evidence shows at each stage of the AMO chain and how each stage has been operationalized and conceptualized; (RQ2) how the three variables function together as a sequential causal chain across national contexts, regime types, media systems, and time periods; and (RQ3) how the full-chain evidence base compares with the broader contextual literature of partial-pathway studies and what directions for future research follow.

METHODS

Protocol and Reporting Standards

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021) for design, conduct, and reporting. Search documentation followed the PRISMA-S checklist (Rethlefsen et al., 2021). The protocol was developed prior to formal screening to anchor analytic decisions and prevent post hoc selection. This review was not pre-registered; no protocol was separately deposited in a public registry. Synthesis methodology followed the Synthesis Without Meta-analysis (SWiM) reporting guideline (Campbell et al., 2020), with conditional meta-analytic pooling. Quality appraisal followed the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018) adapted to the design types represented in the corpus.

Conceptual Framework

The review adopted an Antecedent–Mechanism–Outcome (AMO) framework rather than PEO or PICO because the constructs of interest form a mediation-structured causal chain. Ideological alignment (A) is defined as individual-level ideological predisposition (partisan identification, ideological self-placement, ideological extremity, partisan strength, identity centrality). Cognitive biases (M_1) are systematic biases in political information processing (motivated reasoning, identity-protective cognition, selective exposure, hostile media perception, biased assimilation, disconfirmation bias, meta-perception). Media exposure (M_2) is individual-level exposure to political media content (partisan media, social media feeds, algorithmically curated environments, cross-cutting and incidental exposure, mediated interpersonal contact). Outgroup attitudes and polarization (O) include affective evaluations, social-distance preferences, dehumanization, intolerance, support for political violence, and ideological or affective polarization. Population scope was restricted to adults (18+) in mediated political environments.

Eligibility Criteria

Eligibility criteria were defined a priori and refined following pilot screening of 40 records. Inclusion criteria were: peer-reviewed empirical studies (or recognized preprints from PsyArXiv, SocArXiv, SSRN, OSF, or APSA Preprints) published between January 2000 and 2026; sampling adults in mediated political environments; quantitative, qualitative, or mixed-methods designs; and English, Spanish, German, French, or Portuguese language. The publication-window choice reflected the onset of large-scale internet-mediated political communication and the beginning of systematic scholarly attention to online echo chambers and partisan media fragmentation (Pariser, 2011; Sunstein, 2001). Table 1 presents the full eligibility framework.

Table 1 Eligibility criteria for inclusion and exclusion.

Domain	Inclusion criteria	Exclusion criteria
Study focus	Empirical studies examining relationships among ideological alignment, cognitive biases in political information processing, and media exposure in relation to outgroup attitudes or polarization.	Studies unrelated to ideology, media effects, political cognition, or polarization. Studies on non-political outgroups with no partisan-political dimension.
Population	Adults (18+): voters, citizens, media users, political actors, online communities. Nationally or regionally representative, or non-probability samples where sampling is described.	Studies exclusively on children or adolescents. Non-human or purely computational populations without human-participant data.
Study design	Peer-reviewed empirical studies: quantitative, qualitative, or mixed-methods. Surveys, experiments (laboratory, survey, field), panel and longitudinal designs. Recognized preprints with complete methods and results.	Editorials, opinion pieces, theoretical essays without primary data. Book reviews, commentaries. Conference abstracts with insufficient detail. Content analyses with no human-participant linkage. Retracted articles.
Full-Chain AMO Rule	Must operationalize within a single paper: (i) at least one A construct; (ii) at least one M construct (M_1 OR M_2); and (iii) at least one O construct.	Studies operationalizing only one or two AMO stages. Partial-pathway studies excluded from primary synthesis; retained as contextual pool.
Time frame	Studies published January 2000 to 2026.	Studies published before January 2000.
Publication type	Peer-reviewed journal articles; conference proceedings with complete findings; recognized preprints; doctoral dissertations formally defended where no peer-reviewed version exists.	Blogs, op-eds, non-scholarly publications. Undergraduate or master-level theses.
Language	English (primary). Spanish, German, French, Portuguese (contingent on translation capacity).	Other languages.
Outcome measures	Political polarization (affective or ideological). Outgroup attitudes (hostility, prejudice, dehumanization). Ideological extremity. Intergroup perceptions or partisan stereotyping.	Outcomes unrelated to political identity, partisan groups, or polarization.
Methodological reporting	Full-text accessible. Sample size, sampling approach, measurement instruments, and analytic approach identifiable from full text.	Abstract-only records where full text cannot be retrieved. Studies where basic design cannot be identified.

Note. A = antecedent (ideological alignment); M₁ = cognitive bias mechanism; M₂ = media exposure mechanism; O = outcome (outgroup attitudes/polarization).

Information Sources and Search Strategy

Searches were executed in four databases selected for disciplinary triangulation: Scopus (multidisciplinary social sciences), Web of Science (SSCI, SCIE, and ESCI tiers), PsycINFO (psychology), and PubMed (biomedical and behavioral sciences). No single database adequately covers the AMO chain's constituent disciplines, and using four in combination minimizes single-database coverage bias while ensuring representation of indexed peer-reviewed venues and high-impact behavioral-science outlets. Searches were conducted in early 2026. The master Boolean string combined four conceptual blocks: Block 1 AND (Block 2 OR Block 3) AND Block 4, mapping directly onto the AMO framework. Block 1 retrieved Antecedent terms (e.g., "ideological alignment," "partisan identification"). Block 2 retrieved Cognitive Bias terms (e.g., "motivated reasoning," "selective exposure," "meta-perception"). Block 3 retrieved Media Exposure terms (e.g., "partisan media," "social media," "echo chamber"). Block 4 retrieved Outcome terms (e.g., "affective polarization," "outgroup hostility"). Blocks 2 and 3 were joined by OR because pilot testing showed that requiring both mechanisms simultaneously yielded near-zero recall. Prior to full execution, the master string was piloted against a 15-study benchmark set and validated when it retrieved at least 13 of 15 benchmark studies. A supplementary mediation-signal query was AND-joined in Scopus and Web of Science to prioritize likely full-chain papers.

Study Selection

Records from all four databases were deduplicated, then screened in two stages. At the title and abstract stage, records manifestly irrelevant to the AMO framework were excluded. At the full-text stage, eligibility was assessed against the protocol criteria with the Full-Chain AMO Inclusion Rule as the decisive criterion. Identification yielded 784 records: Scopus (n = 390), Web of Science (n = 159), PsycINFO (n = 175), and PubMed (n = 60). Removal of 102 duplicates left 682 records for screening. At title and abstract, 49 records were excluded for manifest irrelevance, leaving 633 reports sought for retrieval. Of these, 158 satisfied the Full-Chain AMO Inclusion Rule and were retained for primary synthesis; 412 partial-pathway studies were retained as the contextual literature pool for gap-mapping; and 63 records were excluded outright, primarily for failing publication-type criteria. The PRISMA 2020 flow diagram is presented as Fig. 1.

Data Extraction and Quality Appraisal

Data were extracted using a 30-column standardized template recording, for each full-text record: full citation; publication venue and peer-review status; country and political system; study design; sample size and sampling method; population; presence and operationalization of A, M₁, M₂, and O constructs; chain segment tested; whether mediation was formally tested through an indirect-effect model; results and effect direction; eligibility decision and reason for exclusion; reviewer initials; adjudication flag; and analytic notes. A separate qualitative-extraction sheet recorded verbatim definitional and theoretical language used by primary authors for each AMO construct, providing the input to the thematic synthesis. Included studies were appraised using the MMAT (Hong et al., 2018), rated as Met, Partially Met, Not Met, or Cannot Determine. Appraisal was used to contextualize and weight findings rather than to exclude studies; studies rated Not Met on three or more critical domains were flagged for sensitivity analysis.

Synthesis Design

The synthesis integrated two analytic layers in a sequential complementarity design (Pluye and Hong, 2014; Sandelowski et al., 2007). The first was a quantitative gap-mapping and distributional analysis: study counts, percentages of studies operationalizing each AMO stage, distributions of measurement strategies, proportions formally testing mediation, geographic and temporal distributions, and effect-direction tabulations. The second was a qualitative thematic synthesis following Thomas and Harden (2008): verbatim definitional language was extracted from primary sources, descriptive codes were generated, and analytical themes were developed to capture the conceptual logic underlying each code cluster. Coding was conducted by the sole author across three passes with verbatim primary-author language preserved for verification; the absence of a second coder is

acknowledged as a limitation of the review. Integration followed Pluye and Hong's (2014) sequential logic: quantitative patterns identified locations of empirical concentration and absence; thematic synthesis interpreted what those patterns meant in conceptual terms. Narrative synthesis followed SWiM guidelines (Campbell et al., 2020). Meta-analytic pooling was conditional on at least 10 studies reporting comparable effect-size estimates; given heterogeneity of measurement and design, narrative synthesis was the analytic approach used.

Treatment of Contextual Literature and Moderator Analyses

Partial-pathway studies were excluded from primary synthesis per the Full-Chain AMO Inclusion Rule but retained as the contextual literature pool. Permitted uses included background framing, gap identification, and explanation of inconsistencies. Contextual studies were never used for effect-size pooling, never used to support causal claims, and were not counted in the PRISMA total. Where data permitted, the following moderators were examined: country and political system; media type; type of cognitive bias; study design; time period (pre- vs. post-2016); and mechanism-pathway emphasis (M₁-anchored vs. M₂-anchored designs).

RESULTS

Profile of the Included Evidence Base

The 158 included studies are distributed across political-science and political-psychology venues: approximately 43% political science, 21% psychology, and 6% communication and media studies, with the remainder spread across general-science journals, sociology, and specialist outlets. Approximately 83% were dual-indexed in WoS Core (SSCI/SCIE) and Scopus. Table 2 summarizes the distributional profile across operationalization, methodological, geographic, and temporal dimensions.

Table 2 Distributional profile of the 158 included studies.

Dimension	n	%	Notes
Mechanism operationalization			
M ₁ (cognitive bias) only — no M ₂	66	41.3	No media-exposure measure
M ₂ self-report only	50	31.3	Survey-based exposure measures
M ₂ experimental manipulation	36	22.5	Stimulus or feed-rerank designs
M ₂ behavioral / trace data	8	5.0	Logs, browser histories, on-platform
Mediation testing			
Formal indirect-effect model	36	22.5	Structural equation, path, PROCESS
No formal mediation test	108	67.5	Associations only
Insufficient information	16	10.0	—
Effect direction			
Polarizing	76	47.5	Increased outgroup hostility
Depolarizing	24	15.0	Interventions, contact effects
Mixed / conditional	12	7.5	Direction moderator-dependent

Null	1	0.6	—
Unclassifiable	47	29.4	Conceptual ambiguity in A vs. O
Geographic scope			
United States only	110	68.8	Two-party, asymmetric media market
Israel	6	3.8	—
United Kingdom	6	3.8	—
Spain	4	2.5	Multiparty
Germany	4	2.5	Multiparty
Other / cross-national	30	18.8	Long tail; <10 cross-national designs
Temporal distribution			
2000–2009	2	1.3	Pre-affective-polarization era
2010–2014	22	13.8	—
2015–2019	49	30.6	Rise of meta-perception paradigm
2020–2024	64	40.0	Intervention studies concentrate
2025–2026	20	12.5	—

Note. Percentages may sum to slightly more or less than 100 due to rounding or unclassifiable cases. The 3:1 polarizing-to-depolarizing ratio in directional studies is consistent with the field's framing of polarization as a default trajectory that interventions must work against.

Two patterns from Table 2 are particularly notable. First, mechanism operationalization is sharply asymmetric: 41.1% of full-chain-eligible studies satisfied the rule through M_1 alone, and behavioral M_2 measurement is rare (5.1%). Second, geographic concentration is severe: 68.4% of included studies sampled exclusively from the United States, and cross-national designs accounted for fewer than ten studies.

Fig. 1 presents the PRISMA 2020 flow diagram and Fig. 2 displays the thematic structure of the 158 included studies organized into five categories — ideological alignment (A), cognitive biases (M_1), media exposure (M_2), outgroup attitudes and polarization (O), and integrated chain and contextual patterns — and into themes and sub-themes within each category.

Findings on Ideological Alignment (A)

Quantitative profile

Across the 158 included studies, ideological alignment was operationalized most frequently as partisan identification — typically a binary or trichotomous indicator of Democrat/Republican affiliation in U.S. samples, or a multiparty equivalent in non-U.S. contexts. Self-placement on a left–right or liberal–conservative continuum was the second most common operationalization, often paired with measures of ideological extremity. Identity-strength measures (partisan strength, importance of party identity to self-concept) were common in studies grounded in social-identity theory. A smaller subset operationalized alignment in derived ways: issue-extremity scales, ideological congruence with party position, or feeling-thermometer differentials. The pattern of evidence

is consistent: ideological alignment is a robust antecedent of cognitive biases and outgroup attitudes, but its effect size and form depend on how alignment is measured. Identity-strength operationalizations tend to report larger and more robust associations with outgroup hostility (Comellas and Torcal, 2023; Marchal, 2022; Moore-Berg et al., 2020); ideological-extremity operationalizations tend to report quadratic or asymmetric patterns in which both extremes show stronger negative emotions and outgroup derogation than moderates (Krosch et al., 2013; van Prooijen et al., 2015).

Thematic synthesis

Thematic synthesis surfaces three distinct conceptions that the shared label conceals: alignment as categorical group affiliation (Iyengar et al., 2012; Westwood et al., 2017); as a position on a continuous ideological dimension with extremity capturing distance from the center (Cely, 2025; van Prooijen et al., 2015); and as the strength of psychological attachment to a political category (Huddy and Yair, 2020; Mason, 2018). These conceptions are not interchangeable: a respondent can score identically on a partisan-affiliation item and very differently on an identity-strength item. Across the corpus, identity-strength operationalizations produced the largest associations with outgroup hostility; categorical-affiliation operationalizations produced smaller but more replicable associations; and continuous-extremity operationalizations produced nonlinear and asymmetric patterns. The boundary between ideological alignment and outgroup attitude is not stable across the literature, with affective polarization sometimes treated as a mediator (Druckman et al., 2021; Wojcieszak and Warner, 2020) and sometimes as the outcome (Landry et al., 2023; Moore-Berg et al., 2020).

Findings on Cognitive Biases (M_1)

Cognitive biases are the most heavily represented mediator construct in the included evidence base. Quantitative tabulation reveals four mechanism families: motivated reasoning and identity-protective cognition; social-identity and intergroup-bias processes; meta-perceptions of outgroup hostility and dehumanization; and selective evaluation, including biased assimilation and hostile media perception.

Motivated reasoning and identity-protective cognition

Studies grounded in motivated-reasoning theory consistently find that partisans process political information in ways that protect prior beliefs and group identity, yet the construct is invoked at different points in the information-processing stream. Park (2024) treats motivated reasoning as a cluster of biases interacting with cross-cutting exposure to produce perceived polarization, locating the bias in evaluation. Gvirsman (2014) frames motivated reasoning through reinforcing-spiral theory, locating it in differential learning from congenial versus uncongenial content. Derreumaux et al. (2022) reformulate motivated reasoning as biased information sampling, locating it upstream of evaluation. These three accounts place the cognitive bias at three different points in the stream — selection, learning, and evaluation.

Social-identity and intergroup-bias processes

A second, partially overlapping family operationalizes M_1 as social-identity salience, ingroup favoritism, and intergroup bias. Marchal (2022) infers ideological leanings from users' prior contributions to liberal- and conservative-leaning online communities and shows that online discussions are characterized by more positive engagement with like-minded users and more negative engagement with opponents — identity expressed through interaction patterns. Wojcieszak and Garrett (2018) show that priming national identity increases affective polarization among immigration opponents and indirectly increases selective exposure to pro-attitudinal news — identity as a state activated by stimulus. Amira et al. (2019) report that under symbolic moral threat, partisans shift from ingroup help toward outgroup harm — identity as a defensive structure that flips valence under specific moderators.

Meta-perceptions and exaggerated outgroup hostility

Among the most consistently replicated findings is the meta-perceptual exaggeration of outgroup hostility. Moore-Berg et al. (2020) report that Democrats and Republicans dislike and dehumanize each other at similar

levels but believe the outgroup dislikes and dehumanizes them roughly twice as much as the outgroup actually reports; these exaggerated negative meta-perceptions are stable and independently associated with intergroup hostility. Druckman et al. (2021) show that informing respondents that out-partisans are less extreme and less engaged than assumed reduces partisan animus. Landry et al. (2023) demonstrate that correcting exaggerated meta-dehumanization reduces respondents' own dehumanization of out-partisans, with effects persisting at one-week follow-up. Overgaard (2024) shows that unifying media content reduces affective polarization through improved meta-perceptions. Meta-perceptions are the conspicuous exception to conceptual divergence under shared vocabulary: they are consistently defined as beliefs about the outgroup's beliefs about one's ingroup, consistently measured through estimated-versus-actual differentials, and consistently theorized through a dyadic-misperception logic.

Selective evaluation, biased assimilation, and hostile media perception

A fourth family involves selective evaluation of political information. Shi et al. (2022) find that exposure to news about an ingroup politician produces greater perceived credibility and partisan ambivalence relative to outgroup-politician exposure, providing evidence of identity-protective evaluation in a scandal-news context. Popan et al. (2019) report that uncivil internet political discussion lowers perceptions of outgroup rationality, with lower perceived rationality inversely related to evaluations of the political outgroup. Across these studies, selective evaluation is consistently defined as biased credibility judgment under partisan stakes; conceptual boundaries are tighter than for motivated reasoning.

Findings on Media Exposure (M_2)

Media exposure shows the sharpest measurement asymmetry of any construct in the chain. Of the 158 included studies, 50 (31.6%) measure M_2 only by self-report (typically retrospective survey items about media diet); 36 (22.8%) manipulate M_2 experimentally through stimulus presentation, framed news clips, or feed-rerank designs (Bail et al., 2018; Baliotti et al., 2021; Levendusky, 2013); and only 8 (5.0%) measure M_2 through behavioral or trace data such as browser histories, on-platform logs, or computational social-science designs (Marchal, 2022; Piccardi et al., 2024). The remaining 41.3% omit M_2 entirely. Thematic synthesis surfaces four distinct construals that the term “media exposure” conceals: media diet (cumulative content profile), stimulus exposure (one-shot manipulated content), environmental exposure (algorithmically curated feeds), and contact exposure (mediated cross-partisan interaction).

Findings on Outgroup Attitudes and Polarization (O)

The outcome construct shows the broadest operational range. Affective-polarization measures dominate — feeling thermometers, social-distance preferences, trait-stereotype gaps — but the included corpus also contains dehumanization scales, intolerance and civil-liberties items, support for political violence and anti-democratic action, voting and turnout outcomes, and willingness to engage in cross-partisan contact. Empirical correlations among these outcome measures are imperfect: a respondent can show large feeling-thermometer differentials but small social-distance preferences, or vice versa. Klar et al. (2018) demonstrate that part of what feeling-thermometer items measure as outparty dislike is dislike of partisanship per se rather than of the outparty as a group. The thematic synthesis identifies four phenomena — feeling-thermometer differentials, social-distance preferences, trait-stereotype gaps, and dehumanization — that the single label “affective polarization” conflates.

The Three Variables as a Sequential Causal Chain

Bringing the three constructs together, the integrated chain analysis produces three findings. First, the dominant tested sub-pathway is $A \rightarrow M_1 \rightarrow O$: ideological alignment activates a cognitive-bias mechanism (most often motivated reasoning, social-identity salience, or meta-perception), and the mechanism produces an outgroup-attitude or polarization outcome. This sub-pathway is over-represented in the included corpus. Second, the $A \rightarrow M_2 \rightarrow O$ sub-pathway is comparatively under-tested: only 50 studies measured M_2 by self-report and only 8 behaviorally, and the integrated $A \rightarrow M_2 \rightarrow O$ pathway with formal mediation testing is performed in a small minority of these. Third, the full $A \rightarrow M_1 \cdot M_2 \rightarrow O$ pathway — in which both mechanisms are measured in parallel — is rare; where implemented (e.g., Lee, 2025; Wojcieszak and Garrett, 2018), the mediating role of media

exposure operates partly through cognitive biases and partly independently. Studies before 2016 emphasized traditional partisan-media exposure and attitude-extremity outcomes (Garrett et al., 2014; Levendusky, 2013; Stroud, 2010), while studies from 2016 onward shifted toward affective polarization and meta-perceptual mechanisms (Druckman et al., 2021; Iyengar et al., 2019; Landry et al., 2023; Moore-Berg et al., 2020), with intervention studies concentrated in 2020–2025.

Comparison with the Contextual Literature

The 412 partial-pathway studies retained as the contextual literature pool were concentrated in two chain-stage patterns. The most common pattern operationalized A, M₁, and O but omitted M₂ (119 studies, 28.9% of the contextual pool). The second most common pattern operationalized all four constructs but did not satisfy full-chain eligibility because integration was incomplete or mediation logic was not articulated (95 studies, 23.1%). Together these two patterns accounted for over half of the contextual pool. Smaller subsets operationalized only A and M₁ (44 studies, 10.7%); A, M₁, and M₂ without an outcome (32 studies, 7.8%); M₁→O without A (17 studies); A→O without an explicit mechanism (15 studies); none of the four stages in a directly testable empirical form (33 studies, 8.0%); or only A (14 studies). Across both pools, integrated A→M₁·M₂→O designs, behavioral M₂ measurement, and longitudinal full-chain studies were uncommon.

DISCUSSION

The Three-Variable Model and the Two-Variable Practice

The most consequential structural finding concerns what the field operationalizes rather than what it concludes. Although the AMO chain is routinely presented as a three-variable model (Iyengar et al., 2019; Stroud, 2010), the corpus reveals an empirical practice in which media exposure functions more often as a theoretical placeholder than as a fully operationalized construct. The pattern is consistent with Guess's (2015) demonstration that self-report measures systematically overestimate exposure to congenial content, and with the methodological tradition — exemplified in the motivated-reasoning paradigms reviewed by Kahan (2016) — of testing M₁ mechanisms without parallel M₂ measurement; together, these conditions create structural incentives to omit M₂. The empirical consequence is that theoretical accounts emphasizing media-exposure mechanisms (echo chambers, filter bubbles, partisan diets, algorithmic feeds) rest on a thinner evidence base than their theoretical prominence suggests, while the field's strongest empirical results — including the meta-perceptual exaggeration finding (Moore-Berg et al., 2020) — are produced by an A→M₁→O sub-pathway that does not require M₂ measurement. The gap between theory and practice on M₂ is the field's most actionable empirical priority, and closing it through behavioral, trace-data, and platform-log measurement (Marchal, 2022; Piccardi et al., 2024) is where the next wave of designs should concentrate.

The Underuse of Formal Mediation Testing and the Conceptual Diagnosis

A parallel diagnosis follows from the mediation-testing pattern. The AMO framework is fundamentally a mediation theory: antecedent activates mechanism activates outcome. Yet the majority of included studies report associations consistent with mediation without estimating the indirect effect that the theory implies. Where M₂ measurement is the field's first absence, formal mediation testing is its second — and the two absences interact, since estimating an indirect effect through a poorly measured mediator yields biased estimates even when the structural model is correctly specified (Pluye and Hong, 2014). The qualitative thematic synthesis adds a third structural problem: at every stage of the chain, apparent empirical convergence is partly an artifact of vocabulary. “Ideological alignment” refers to categorical affiliation, continuous ideological position, and identity strength — three constructs that produce different effect sizes. “Motivated reasoning” refers to information-sampling bias, learning bias, and evidence-evaluation bias — three operations at three different points in the information-processing stream. “Media exposure” refers to media diet, stimulus exposure, environmental exposure, and contact exposure — four constructs that imply different causal structures. “Affective polarization” refers to feeling-thermometer differentials, social-distance preferences, trait-stereotype gaps, and dehumanization. Meta-perceptions are the conspicuous exception: across diverse samples and methods, the construct produces replicable descriptive findings and replicable intervention effects with relatively little conceptual drift.

The Philippine Case: Activism, Partisan Bloggers, and the AMO Chain in a Hybrid Media Environment

The geographic concentration of the included evidence in U.S. samples (68.8%) is the single largest limitation on the generalizability of current findings. The United States is structurally distinctive: a two-party system with no proportional alternative, an asymmetric partisan-media market with a long-established conservative media infrastructure, and a partisan identity unusually fused with social identity, race, and religion. The most urgent cross-national replication priorities are multiparty proportional-representation systems, competitive-authoritarian and hybrid regimes, the Global South, and contexts where partisan identity competes with ethnic, religious, or dynastic cleavages. The Philippines sits at the intersection of several of these features and offers an instructive case for examining whether the AMO chain reproduces, transforms, or breaks down in a hybrid-media, personalist-partisan environment.

Three features of contemporary Philippine politics map directly onto the AMO chain. First, ideological alignment (A) is structured less by left–right ideology than by dynastic and personalist identification, with the 2016 Duterte and 2022 Marcos Jr.–Sara Duterte UniTeam coalitions both reorganizing partisan identity around personalities rather than parties (Arugay and Slater, 2023); the post-2024 UniTeam breakup has further activated affective-polarization dynamics within what was a single coalition, and Agonosa (2025) documents an associated affective outcome — “fear of friendship over,” or anticipatory anxiety about politically motivated friendship dissolution — that parallels but is not identical to the social-distance measures dominant in the U.S. literature. Second, cognitive-bias mechanisms (M_1) operate on identity structures with unusually salient moral content, characterized by Feliciano (2024) as a “moral chasm” consistent with the motivated-reasoning and identity-protective-cognition mechanisms documented in the U.S. literature (Derreumaux et al., 2022; Park, 2024); meta-perceptual exaggeration of outgroup hostility (Moore-Berg et al., 2020) is empirically untested in the Philippines but theoretically expected to be substantial. Third, media exposure (M_2) is structured by a hybrid traditional-and-social media environment in which partisan vloggers and political bloggers have become institutionalized political actors — campaigning for candidates, receiving government appointments, and obtaining presidential press accreditations equivalent to professional journalists (Ong and Cabañes, 2018; Sinpeng et al., 2024) — and in which the post-2022 fragmentation of these networks provides observable behavioral evidence of motivated-reasoning mechanisms operating in real time (Arugay and Slater, 2023; Sinpeng et al., 2024).

Recent mobilizations make the chain’s operation visible: Iglesia ni Cristo–led rallies organized through social media (Sarao, 2025), pro-Duterte demonstrations following the former president’s surrender to the International Criminal Court in March 2025, and the 2025–2026 anti-corruption protests around the flood-control scandal organized by student networks and youth coalitions (Human Rights Watch, 2025) each constitute a real-world instance in which ideological alignment activates motivated-reasoning and identity-protective mechanisms and platform-mediated exposure both transmits and amplifies the activation — and each remains untested through the integrated three-variable framework that this synthesis identifies as the field’s theoretical aspiration.

Three implications follow specifically for Philippine political-communication research. First, the dominant U.S. operationalization of A as left–right ideology will under-perform in the Philippine case; identity-strength and dynastic-affiliation operationalizations are more likely to capture the relevant variation. Second, the partisan-blogger ecosystem provides a natural setting for behavioral M_2 measurement (subscription patterns, engagement traces, cross-platform mobility) that is methodologically rare in the global literature; Filipino researchers are well-positioned to contribute behavioral M_2 evidence that the field as a whole is missing. Third, longitudinal designs that follow respondents across the rapid reconfigurations of Philippine partisan identity — UniTeam formation, UniTeam breakup, ICC arrest, ongoing anti-corruption mobilizations — would directly test chain ordering in a way that U.S. cross-sectional designs cannot.

Methodological and Theoretical Implications

Three methodological patterns warrant particular attention beyond the M_2 asymmetry. The first is the relative scarcity of formal mediation testing (22.8% of included studies). The second is over-reliance on cross-sectional designs: the chain is theorized as a temporal process, but most evidence is synchronic. Longitudinal designs that measure A, M_1 , M_2 , and O at repeated waves are the only way to test chain ordering rigorously, and they are rare

in both evidence pools. The third is the asymmetry between cognitive-bias measurement (often experimentally manipulated) and media-exposure measurement (often self-reported); closing this asymmetry through trace data, on-platform logs, or experimentally controlled feeds is a field priority. Theoretically, the relationships among the three variables are best understood not as a single linear pathway but as a family of pathways with at least four parallel sub-mechanisms operating through M_1 (motivated reasoning, social-identity processes, meta-perceptions, selective evaluation) and at least four through M_2 (media diet, stimulus exposure, environmental exposure, contact exposure). The interventional implications of these M_1 sub-mechanisms vary by locus in the information-processing stream: sampling-stage biases call for algorithmic-feed correction, learning-stage biases for balanced-content interventions over time, and evaluation-stage biases for inoculation or critical-thinking primes. Meta-perceptions deserve a more central place in chain models: the strength and consistency of meta-perceptual evidence makes them one of the most replicable findings in the field, connecting cognitive-bias and intervention literatures productively. The chain's reversibility through M_1 interventions has normative implications: polarization is not inevitable, and depolarization is empirically tractable through identity- and meta-perception-correction designs.

Implications for Future Research, Policy, and Practice

Five recommendations follow for future research. First, prioritize integrated $A \rightarrow M_1 \cdot M_2 \rightarrow O$ designs that combine cognitive-bias and media-exposure mechanisms in parallel within the same empirical study (Lee, 2025; Wojcieszak and Garrett, 2018). Second, supplement self-report measures of partisan-media use with on-platform logs, browser histories, controlled feed reranking, or pre-registered field experiments (Baliotti et al., 2021; Piccardi et al., 2024). Third, implement longitudinal full-chain designs in which A , M_1 , M_2 , and O are measured at repeated waves; three-wave or longer panels with embedded experimental components are particularly valuable. Fourth, declare the conceptual specification of each construct used in any new study: which conception of ideological alignment, which locus of motivated reasoning, which conception of media exposure, which phenomenon of affective polarization. Fifth, replicate the chain outside the United States, especially in multiparty proportional systems, competitive-authoritarian and hybrid regimes, and the Global South — with the Philippines a methodologically advantaged priority case. For policy and practice, two implications follow from the chain's reversibility through M_1 interventions: meta-perceptual correction interventions are among the most empirically tractable depolarizing tools available (Druckman et al., 2021; Landry et al., 2023); and contact-based interventions are effective but conditional, requiring program designers to distinguish between contact opportunities and contact quality (Wojcieszak and Warner, 2020).

Limitations

Six limitations should be acknowledged. First, the review was restricted to four databases and operationally to English-language sources, which may have excluded relevant non-English studies particularly from Latin America, East Asia, and continental Europe. Second, the publication window of January 2000 to 2026 excluded earlier foundational works. Third, the Full-Chain AMO Inclusion Rule is intentionally strict and excluded a large body of partial-pathway studies from primary synthesis. Fourth, conceptual heterogeneity in the operationalization of A , M_1 , M_2 , and O — itself a finding of the thematic synthesis — makes cross-study comparison challenging and contributes to the 29.1% of included studies whose effect direction could not be cleanly classified. Fifth, the breadth of the inclusion rule means that observed heterogeneity partly reflects the protocol's design choice. Sixth, the thematic synthesis was conducted by a single coder; the three-pass coding procedure was systematic and verbatim primary-author language is preserved for verification, but multi-coder replication with formal inter-rater reliability would strengthen the thematic findings. The Philippine-context analysis draws on grey-literature, news-media, and recent peer-reviewed sources rather than on full-chain AMO studies of Philippine samples, which do not yet exist; it is framed as a research-priority case rather than a synthesis result.

CONCLUSION

This systematic review's contribution lies less in cataloguing what the field has found than in diagnosing how the field has tested its own theory. The integrated three-variable model underwriting major claims about

polarization, echo chambers, and partisan media is widely endorsed in principle but unevenly implemented in practice; and the field's apparent empirical convergence on shared construct labels coexists with substantive divergence about what those constructs are. Recognizing this gap reframes the research agenda. Future work should prioritize designs that operationalize both mechanism families within the same study, supplement self-report with behavioral and trace-data measurement of media exposure, estimate mediation formally rather than rhetorically, and follow respondents longitudinally so that chain ordering can be tested rather than assumed. Equally important is replication outside the United States, especially in hybrid-media and multiparty contexts where the chain's invariance has yet to be established; the Philippine case, with its institutionalized partisan-blogger ecosystem and rapidly reconfiguring personalist coalitions, is a methodologically advantaged setting for that work. The chain is empirically reversible through meta-perceptual and identity-correction interventions (Druckman et al., 2021; Landry et al., 2023), so polarization is contingent on the joint operation of identity, cognition, and information, and interventions that interrupt the chain at the M_1 stage are currently its most tractable depolarizing levers. Sustained progress will depend on whether the field's empirical practice catches up to its theoretical aspiration.

ACKNOWLEDGEMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration Of Interest Statement

The author declares no competing interests.

AI use disclosure

The author acknowledges the use of artificial intelligence-assisted tools during the preparation of this manuscript. During manuscript preparation, we used OpenAI ChatGPT (version GPT-4) to assist in drafting literature summaries and ensuring clarity of expression. These tools were used only for grammar checking, language refinement, clarity improvement, and formatting support. No artificial intelligence tool was used to generate the study's research design, data extraction, analysis, interpretation of findings, or conclusions.

Data Availability

All materials supporting this systematic review, including the PRISMA checklist and flow diagram, full search strategies for all databases, extracted records, screening logs, inclusion/exclusion criteria, and data extraction tables, are available for verification. These materials have been uploaded as supplementary files with this submission. Editors and reviewers can access them directly through the supplementary material section. There are no restrictions on access for the purposes of peer review.

Ethics Declarations

This systematic review synthesized previously published research and did not involve direct collection of data from human participants. Formal Institutional Review Board (IRB) approval was therefore not required. No informed consent was required because this review did not involve direct data collection from human participants. The review adhered to standards of research integrity: the protocol was developed prior to formal screening; the synthesis followed PRISMA 2020 (Page et al., 2021), PRISMA-S (Rethlefsen et al., 2021), and SWiM (Campbell et al., 2020) reporting guidelines; quality appraisal followed the Mixed Methods Appraisal Tool (Hong et al., 2018); and the dual-layer treatment of primary and contextual literature was specified a priori.

Author Contributions

The sole author performed all roles in the research process in accordance with the Contributor Roles Taxonomy (CRediT): Conceptualization; Methodology; Investigation; Data curation; Formal analysis; Writing — original draft; Writing — review and editing; Visualization; Project Administration.

REFERENCES

1. Agonosa MJM (2025). Exploring fear of friendship over (FOFO) in social media political participation during the 2022 Philippine presidential elections. *Asian J Commun* 35(6):558–581
2. Ahler DJ (2014). Self-fulfilling misperceptions of public polarization. *J Polit* 76(3):607–620
3. Ahler DJ, Sood G (2018). The parties in our heads: misperceptions about party composition and their consequences. *J Polit* 80(3):964–981
4. Amira K, Wright JC, Goya-Tocchetto D (2019). In-group love versus out-group hate: which is more important to partisans and when? *Polit Behav* 43(2):473–494
5. Amsalem E, Merkley E, Loewen PJ (2022). Does talking to the other side reduce inter-party hostility? Evidence from three studies. *Polit Commun* 39(1):61–78
6. Arceneaux K, Vander Wielen R (2025). An open mind or a big heart: possible routes to reducing in-group bias. *Polit Commun* 42(1):1–22
7. Arugay AA, Slater JKA (2023). Mobilized and polarized: social media and disinformation narratives in the 2022 Philippine elections. *Pac Aff* 96(3):549–571
8. Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MBF, Lee J, Mann M, Merhout F, Volfovsky A (2018). Exposure to opposing views on social media can increase political polarization. *Proc Natl Acad Sci USA* 115(37):9216–9221
9. Baldassarri D, Park B (2019). Was there a culture war? Partisan polarization and secular trends in U.S. public opinion. *J Polit* 82(3):809–827
10. Baliatti S, Getoor L, Goldstein DG, Watts DJ (2021). Reducing opinion polarization: effects of exposure to similar people with differing political views. *Proc Natl Acad Sci USA* 118(52):e2112552118. <https://doi.org/10.1073/pnas.2112552118>
11. Banda KK, Cluverius J (2018). Elite polarization, party extremity, and affective polarization. *Elect Stud* 56:90–101
12. Bantel I (2023). Camps, not just parties: the dynamic foundations of affective polarization in multi-party systems. *Elect Stud* 83:102614
13. Brienza JP, Kung FYH, Chao MM (2021). Wise reasoning, intergroup positivity, and attitude polarization across contexts. *Nat Commun* 12(1):3313
14. Bruneau E, Kteily N (2017). The enemy as animal: symmetric dehumanization during asymmetric warfare. *PLOS ONE* 12(7):e0181422
15. Campbell M, McKenzie JE, Sowden A et al (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ* 368:l6890. <https://doi.org/10.1136/bmj.l6890>
16. Cassese EC (2019). Partisan dehumanization in American politics. *Polit Behav* 43(1):29–50
17. Cely T (2025). One more constrained than the other: asymmetrical ideological alignment and its implications for polarization. *Eur J Polit Res*. <https://doi.org/10.1111/1475-6765.12742>
18. Comellas JM, Torcal M (2023). Ideological identity, issue-based ideology and bipolar affective polarization in multiparty systems. *Elect Stud* 83:102615
19. Costello K, Hodson G (2012). Explaining dehumanization among children: the interspecies model of prejudice. *Br J Soc Psychol* 53(1):175–197
20. Derreumaux Y, Bergh R, Hughes BL (2022). Partisan-motivated sampling: re-examining politically motivated reasoning across the information processing stream. *J Pers Soc Psychol* 123(2):316–336
21. Druckman JN, Klar S, Krupnikov Y, Levendusky M, Ryan JB (2021a). (Mis-)estimating affective polarization. *J Polit* 84(2):1106–1117
22. Druckman JN, Klar S, Krupnikov Y, Levendusky M, Ryan JB (2021b). Affective polarization, local contexts and public opinion in America. *Nat Hum Behav* 5(1):28–38
23. Enders AM, Armaly MT (2018). The differential effects of actual and perceived polarization. *Polit Behav* 41(3):815–839
24. Feliciano MPJ (2024). Bridging the moral chasm: understanding the increasing polarization in the Philippines. In: *Democratic Erosion*. Democratic Erosion Consortium. <https://democratic-erosion.org/2024/bridging-the-moral-chasm/>. Accessed 1 June 2026
25. Garrett KN, Bankert A (2018). The moral roots of partisan division: how moral conviction heightens affective polarization. *Br J Polit Sci* 50(2):621–640

26. Garrett RK (2019). Social media's contribution to political misperceptions in U.S. presidential elections. *PLOS ONE* 14(3):e0213500
27. Garrett RK, Weeks BE, Neo RL (2014). Driving a wedge between evidence and beliefs: how online ideological news exposure promotes political misperceptions. *J Comput Mediat Commun* 21(5):331–348
28. Górska P, Bulska D, Górski M (2025). The more, the merrier: the effect of social network heterogeneity on attitudes toward political opponents. *Soc Incl* 13(1)
29. Guess AM (2015). Measure for measure: an experimental test of online political media exposure. *Polit Anal* 23(1):59–75
30. Gvirsman SD (2014). It is not that we do not know, it is that we do not care: explaining why selective exposure polarizes attitudes. *Mass Commun Soc* 17(1):74–97
31. Han J, Federico CM (2017). Conflict-framed news, self-categorization, and partisan polarization. *Mass Commun Soc* 20(4):455–480
32. Harel TO, Nir N, Vandermeulen D, Maoz I, Halperin E (2025). A threat to cohesion: intragroup affective polarization in the context of intractable intergroup conflict. *J Confl Resolut* 69(4):689–718. <https://doi.org/10.1177/00220027241247033>
33. Harteveld E (2021). Ticking all the boxes? A comparative study of social sorting and affective polarization. *Elect Stud* 72:102337
34. Hodson G, Costello K (2007). Interpersonal disgust, ideological orientations, and dehumanization as predictors of intergroup attitudes. *Psychol Sci* 18(8):691–698
35. Holliday DE, Lelkes Y, Westwood SJ (2025). Why depolarization is hard: evaluating attempts to decrease partisan animosity in America. *Proc Natl Acad Sci USA* 122(39):e2508827122. <https://doi.org/10.1073/pnas.2508827122>
36. Hong QN, Pluye P, Fàbregues S et al (2018). Mixed Methods Appraisal Tool (MMAT) version 2018. Canadian Intellectual Property Office
37. Huddy L, Yair O (2020). Reducing affective polarization: warm group relations or policy compromise? *Polit Psychol* 42(2):291–309
38. Human Rights Watch (2025). World report 2025: Philippines. In: World Report 2025. Human Rights Watch. <https://www.hrw.org/world-report/2025/country-chapters/philippines>. Accessed 1 June 2026
39. Iyengar S, Lelkes Y, Levendusky M, Malhotra N, Westwood SJ (2019). The origins and consequences of affective polarization in the United States. *Annu Rev Polit Sci* 22:129–146
40. Iyengar S, Sood G, Lelkes Y (2012). Affect, not ideology: a social identity perspective on polarization. *Public Opin Q* 76(3):405–431
41. Iyengar S, Westwood SJ (2015). Fear and loathing across party lines: new evidence on group polarization. *Am J Polit Sci* 59(3):690–707
42. Jenke L (2024). Affective polarization and misinformation belief. *Polit Behav* 46(2):825–884. <https://doi.org/10.1007/s11109-022-09851-w>
43. Kahan DM (2016). The politically motivated reasoning paradigm, part 1. In: Scott RA, Kosslyn SM (eds) *Emerging trends in the social and behavioral sciences*. Wiley, pp 1–16. <https://doi.org/10.1002/9781118900772.etrds0417>
44. Kane JV, Mason L, Wronski J (2021). Who's at the party? Group sentiments, knowledge, and partisan identity. *J Polit* 83(4):1783–1799
45. Klar S, Krupnikov Y, Ryan JB (2018). Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship. *Public Opin Q* 82(2):379–390
46. Koetke J, Conrique BG, Schumann K (2023). Bridging the divide: the effect of individuating information on attitudes toward political outgroup members. *J Soc Polit Psychol* 11(2):562–581. <https://doi.org/10.5964/jsp.9836>
47. Krosch AR, Berntsen L, Amodio DM, Jost JT, Van Bavel JJ (2013). On the ideology of hypodescent: political conservatism predicts categorization of racially ambiguous faces as Black. *J Exp Soc Psychol* 49(6):1196–1203
48. Kubin E, Gray KJ, von Sikorski C (2023). Reducing political dehumanization by pairing facts with personal experiences. *Polit Psychol* 44(5):1119–1140
49. Landry AP, Schooler JW, Willer R, Seli P (2023). Reducing explicit blatant dehumanization by correcting exaggerated meta-perceptions. *Soc Psychol Personal Sci* 14(4):407–418

50. Lee J (2025). Mediating roles of cognitive bias and media exposure: a parallel mediation test. *Polit Commun*. <https://doi.org/10.1080/10584609.2025.2449318>
51. Levendusky M (2013). *How partisan media polarize America*. University of Chicago Press, Chicago
52. Marchal N (2022). “Be nice or leave me alone”: an intergroup perspective on affective polarization in online political discussions. *Commun Res* 49(3):376–398
53. Mason L (2018). *Uncivil agreement: how politics became our identity*. University of Chicago Press, Chicago
54. Moore-Berg SL, Ankori-Karlinsky L-O, Hameiri B, Bruneau E (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proc Natl Acad Sci USA* 117(26):14864–14872
55. Ong JC, Cabañes JVA (2018). Architects of networked disinformation: behind the scenes of troll accounts and fake news production in the Philippines. *The Newton Tech4Dev Network*
56. Overgaard CSB (2024). Perceiving affective polarization in the United States: how social media shape meta-perceptions and affective polarization. *Soc Media Soc* 10(1). <https://doi.org/10.1177/20563051241232662>
57. Page MJ, McKenzie JE, Bossuyt PM et al (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. <https://doi.org/10.1136/bmj.n71>
58. Pariser E (2011). *The filter bubble: what the Internet is hiding from you*. Penguin Press, New York
59. Park B (2024). Cross-cutting exposure, motivated reasoning, and perceived polarization. *Polit Commun* 41(5):729–748. <https://doi.org/10.1080/10584609.2024.2310549>
60. Piccardi T, Saveski M, Jia C, Hancock JT, Tsai JL, Bernstein MS (2024). Social media algorithms can curb misinformation, but do users want them to? *Science* 386(6724):991–996
61. Pluye P, Hong QN (2014). Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. *Annu Rev Public Health* 35:29–45
62. Popan JR, Coursey L, Acosta J, Kenworthy J (2019). Testing the effects of incivility during internet political discussion on perceptions of rational argument and evaluations of a political outgroup. *Comput Hum Behav* 96:81–90
63. Rethlefsen ML, Kirtley S, Waffenschmidt S et al (2021). PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev* 10(1):39
64. Sandelowski M, Voils CI, Barroso J (2007). Defining and designing mixed research synthesis studies. *Res Sch* 13(1):29
65. Sarao Z (2025). INC’s anti-corruption rally draws over 300,000 in Manila. *Philippine Daily Inquirer*. <https://newsinfo.inquirer.net/>
66. Shi R, Messaris P, Cappella JN (2022). Partisan ambivalence in the processing of news about ingroup and outgroup politicians’ scandals. *Mass Commun Soc* 25(2):230–253
67. Sinpeng A, Tapsell R, Cabañes JVA, Ong JC (2024). Posting for profit: social media influencers in Philippine politics. *FULCRUM (ISEAS – Yusof Ishak Institute)*
68. Stroud NJ (2010). Polarization and partisan selective exposure. *J Commun* 60(3):556–576
69. Stroud NJ (2017). Selective exposure theories. In: Kenski K, Jamieson KH (eds) *The Oxford handbook of political communication*. Oxford University Press, Oxford, pp 531–548
70. Sunstein CR (2001). *Republic.com*. Princeton University Press, Princeton
71. Taber CS, Lodge M (2006). Motivated skepticism in the evaluation of political beliefs. *Am J Polit Sci* 50(3):755–769
72. Thomas J, Harden A (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol* 8:45
73. van Prooijen J-W, Krouwel APM, Boiten M, Eendebak L (2015). Fear among the extremes: how political ideology predicts negative emotions and outgroup derogation. *Pers Soc Psychol Bull* 41(4):485–497
74. Westwood SJ, Iyengar S, Walgrave S, Leonisio R, Miller L, Strijbis O (2017). The tie that divides: cross-national evidence of the primacy of partyism. *Eur J Polit Res* 57(2):333–354
75. Wojcieszak M, Garrett RK (2018). Social identity, selective exposure, and affective polarization: how priming national identity shapes attitudes toward immigrants via news selection. *Hum Commun Res* 44(3):247–273
76. Wojcieszak M, Warner BR (2020). Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact. *Polit Commun* 37(6):789–811

Figure Legends

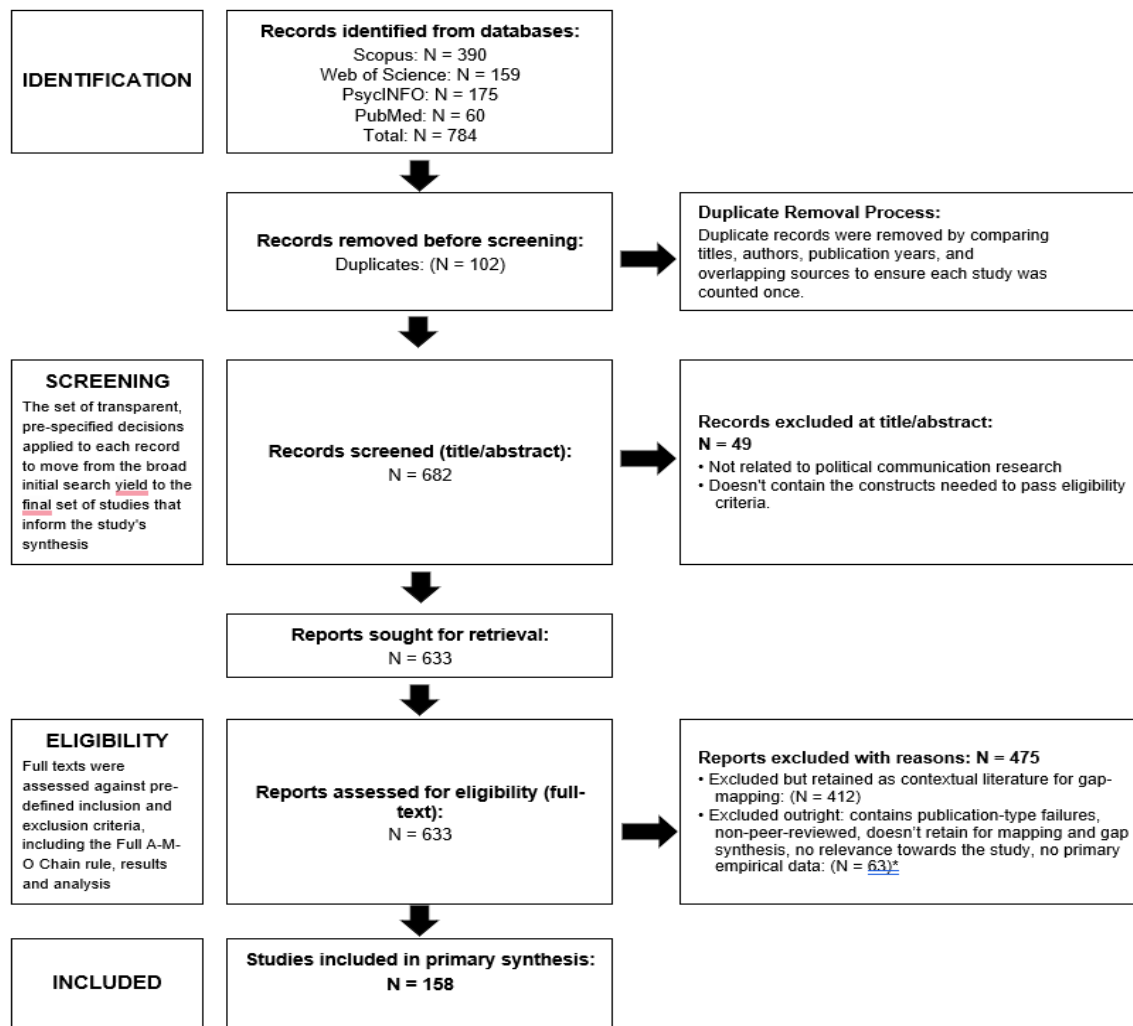


Fig. 1 | PRISMA 2020 flow diagram for study selection. Records were identified from four electronic databases: Scopus (n = 390), Web of Science (n = 159), PsycINFO (n = 175), and PubMed (n = 60); total N = 784. Following removal of 102 duplicates, 682 records were screened at the title and abstract stage, of which 49 were excluded for manifest irrelevance to the AMO framework. All 633 full-text records were assessed for eligibility. Of these, 158 satisfied the Full-Chain AMO Inclusion Rule and were retained for primary synthesis. A further 412 partial-pathway studies were retained as the contextual literature pool for gap-mapping. Sixty-three records were excluded outright, primarily for failing publication-type criteria (working papers from non-recognized repositories, an undergraduate thesis, edited book chapters without primary data, manuscripts with no identifiable peer-reviewed venue, and two studies removed during post-hoc eligibility re-verification for lacking primary empirical data). Source: author's own diagram following PRISMA 2020 (Page et al., 2021).

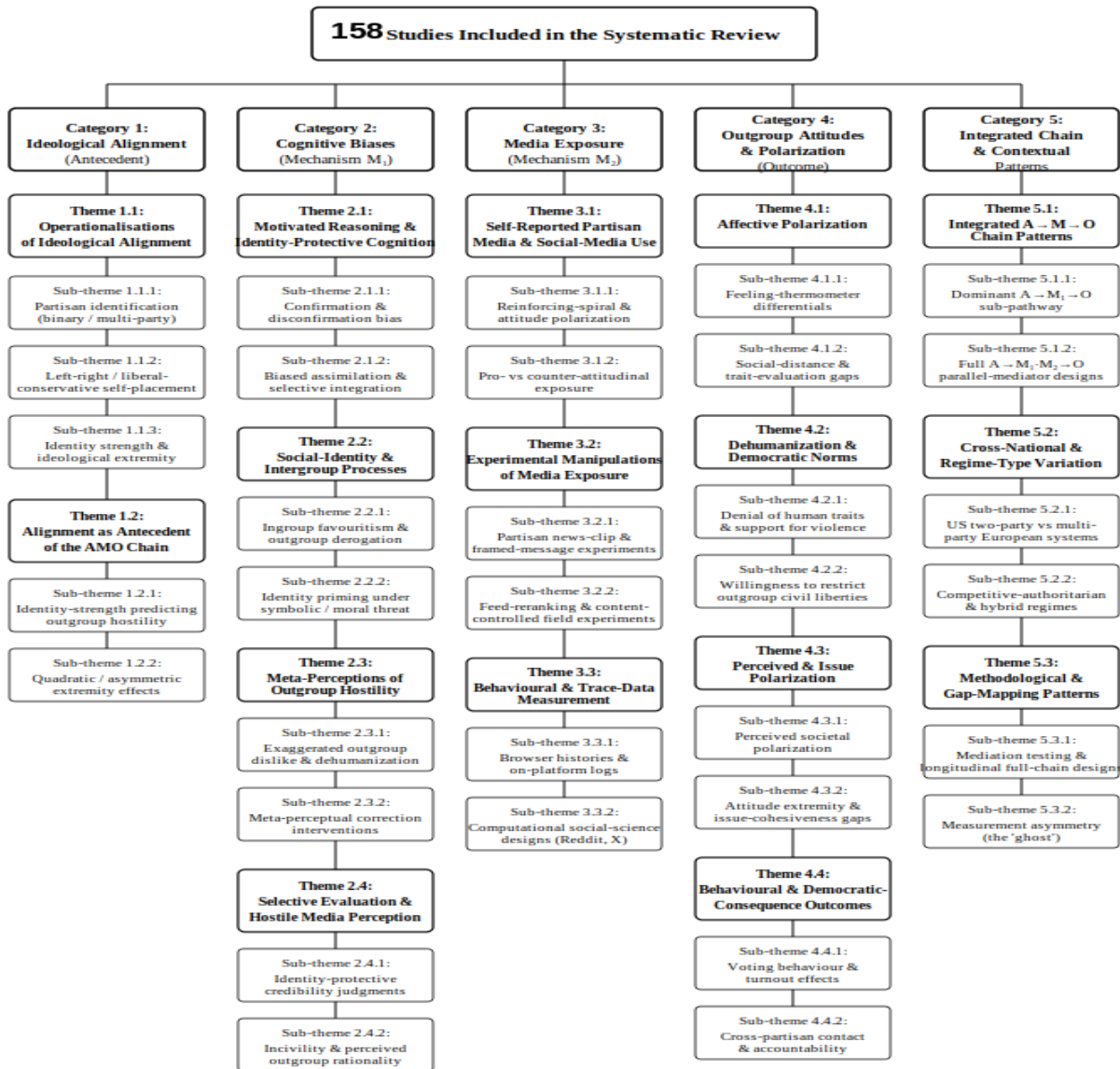


Fig. 2 | Display of SLR results: thematic structure of the 158 included studies organized by AMO category, theme, and sub-theme.