Supplementary information for

From form to function: an ontology of communicative influence beyond truth and intent

Anonymised for peer review

Abstract

This supplement provides the full operational specification and supporting validation and verification materials for *From form to function: An ontology of communicative influence beyond truth and intent*.

Appendices A–F contain the normative human-readable specification: TargetConstruct foundations and fixed Tactic mappings (A), the Method Tactic catalogue (B), Narrative devices and archetype derivation (C), Polarization motifs and mappings (D), TopicDomain groupings (E), and AboutObject vocabularies and evidence rules (F).

Appendix G documents formal verification and reproducibility artefacts; Appendix H reports the framework-coverage assessment; Appendix I documents the development-stage annotation exercise and specification-revision cases; and Appendix J provides the conceptual provenance and decision ledger linking literature constructs to the final canonical vocabularies.

The operational appendices remain authoritative for annotation. These materials support independent inspection, implementation, and reuse of the specification; empirical annotation reliability and computational annotation performance are not established here.

How to use this supplement

This supplement is designed to be entered by task rather than read linearly. For message annotation or computational implementation, Appendices A–F constitute the operational specification: Appendix B governs Method/Tactic assignment; Appendix C Narrative; Appendix D Polarization; Appendix E TopicDomain; and Appendix F AboutObject. Appendix A provides the theoretical basis and fixed mappings for TargetConstruct, which is derived from assigned Tactics rather than annotated independently.

The remaining appendices document supporting evidence and development materials rather than additional annotation requirements. Appendix G records the formal verification of the OWL and SHACL reference representation; Appendix H supports assessment of framework coverage; Appendix I

documents the development-stage annotation exercise and specification-revision cases; and Appendix J provides the conceptual provenance and decision trail underlying the canonical vocabularies.

Where a compact table, provenance entry, or explanatory summary abbreviates an operational rule, the full definition, minimum-evidence condition, boundary rule, and canonical mapping in Appendices A–F govern annotation.

Guide to the supplemental appendices

App.	**Primary purpose**	**Use when...**
A	TargetConstruct foundations	interpreting the six TargetConstructs or inspecting fixed Tactic mappings.
B	Method Tactic codebook	annotating or implementing the 71 canonical communicative influence Tactics.
C	Narrative specification	annotating NarrativeDevices, attributes, configurations, or inspecting NarrativeArchetype derivation.
D	Polarization specification	annotating group-positioning motifs or inspecting their fixed directions and axes.
E	TopicDomain	indexing the substantive subject matter of a message.
F	AboutObject	annotating broader influence objects and deriving their domain and sphere.
G	Formal verification	inspecting or reproducing the OWL, SHACL, and derivation checks.
H	Framework coverage	examining how adjacent frameworks map onto the ontology and how apparent gaps were adjudicated.
I	Development-stage exercise	examining annotation cost, revision cases, and applied boundary adjudications.
J	Conceptual provenance	inspecting the literature basis and design decisions underlying the canonical vocabularies.

Contents

How to use this supplement **1**

A Method dimension: TargetConstruct foundations and derivation **8**
A.1 Perception / Attention . . . 9
A.2 Belief . . . 9
A.3 Attitude . . . 9
A.4 Identity . . . 9
A.5 Agency / Choice . . . 10
A.6 Behavior . . . 10

B Method dimension: Extended cue–tactic catalogue **12**
B.1 Method families . . . 12
B.2 Cue annotation and Tactic assignment . . . 16
B.3 Operational quick reference . . . 17
B.4 Persuasion . . . 24
B.5 Framing . . . 44
B.6 Manipulation . . . 57
B.7 Coercion . . . 68
B.8 Deception (form-level; no intent inference or fact-checking) . . . 77
B.9 Distraction . . . 87

C Narrative devices and archetypes **96**
C.1 Narrative devices palette (for reference) . . . 96
C.2 Narrative device attributes (coding fields) . . . 97
C.3 Narrative binding and archetype composition . . . 101
C.4 Narrative archetypes and deterministic signature derivation . . . 104

D Polarization **106**
D.1 Polarization scope and axes . . . 106
D.2 Polarization axes (coder-facing glossary) . . . 108
D.3 Relational interpretation of polarization directions . . . 108

E TopicDomain groupings **121**
E.1 Cultural/Jurisdictional adaptation . . . 121

F AboutObject: Construct and controlled vocabularies **122**
F.1 L1: Sphere tags (`aboutObjectFamily`) . . . 125
F.2 L2: Domain taxonomy (A–L; closed list) . . . 125
F.3 L3 label catalogue (L1–L2–L3 mapping) . . . 126
F.4 AboutObject: Evidence templates (L4) and selection rules . . . 162
F.5 AboutObject: Evidence map (construct to role to sources) . . . 176

G Computational Verification and Reproducibility Package 179
G.1 Formal verification strategy . . . 180
G.2 Verification artefacts . . . 180
G.3 Validation procedure . . . 181
G.4 Results . . . 181

H Coverage check: cited frameworks against ontology architecture 182
H.1 SemEval propaganda and persuasion shared-task lineage . . . 185
H.2 SemEval-2020 Task 11: 14-class technique crosswalk . . . 185
H.3 SemEval-2021 Task 6 and SemEval-2024 Task 4: meme-persuasion crosswalk . . . 188
H.4 SemEval-2023 Task 3: 23-technique persuasion-subtask crosswalk . . . 190
H.5 ENISA/EEAS FIMI (2022): scope comparison . . . 193
H.6 Tudela et al. (2025): Influence Operation Ontology crosswalk 194
H.7 DISARM: object-level scope crosswalk . . . 195
H.8 Lenk (2026): multilevel framing–narrative correspondence . . 197
H.9 Institute for Propaganda Analysis (1939): seven-device crosswalk . . . 198
H.10 Entman (1993): framing-function crosswalk . . . 200
H.11 Benford & Snow (2000): core collective-action framing tasks . 202
H.12 Campaign- and operation-level frameworks: scope notes . . . 203
H.13 Coverage synthesis . . . 205

I Development-Stage Annotation Exercise and Specification-Revision Cases 206
I.1 Data and procedure . . . 207
I.2 Descriptive annotation-effort profile . . . 207
I.3 Development-stage annotation interface . . . 208
I.4 Specification-revision cases . . . 208
I.5 Lessons from specification revision . . . 217

J Ontology provenance and decision ledger 218
J.1 Method and Tactic provenance . . . 220
J.2 TargetConstruct provenance and mapping verification . . . 272
J.3 Narrative provenance and derivation . . . 276
J.4 Polarization provenance and deterministic axis derivation . . 283
J.5 TopicDomain provenance . . . 293
J.6 AboutObject provenance and vocabulary derivation . . . 297

List of Tables

S1 Reverse index of the fixed TargetConstruct mappings specified in Appendix B. TargetConstruct values are inherited deterministically from Tactic assignments and are not independently annotated. 11
S2 Operational interpretation of the six Method families. 13
S3 Operational quick reference for Method Tactics. Full Appendix B entries govern boundary cases and overlapping assignments. 18
S4 Narrative devices palette. 98
S5 Narrative device attributes and allowed values. 99
S6 Narrative device attribute glossary (coder-facing). 100
S7 Coder-facing rules for narrative binding and archetype composition. 103
S8 Narrative archetypes: definitions, devices, and spotting cues. . 104
S9 Narrative archetype signatures from device attributes. 106
S10 Polarization motifs (Divide): operational definitions and coding boundaries. 110
S11 Polarization motifs (Cohere): operational definitions and coding boundaries. 113
S12 Polarization motifs (Bridge): operational definitions and coding boundaries. 117
S13 Deterministic motif → direction and axis-set mapping (PolarizationEffect lookup). 120
S14 TopicDomain groupings with illustrative subtopics. 122
S15 L3 label catalogue with deterministic L3-to-L2 domain and L2-to-L1 sphere inheritance. 137
S16 L4 evidence templates: selection guidance and typical cue families. 163
S17 Traceability of literature use in the AboutObject specification. 177
S18 Coverage check: summary of framework verdicts. 183
S19 SemEval-2020 Task 11 propaganda-technique classes crosswalked to the canonical ontology. 185
S20 SemEval-2021/2024 meme-persuasion techniques crosswalked to the canonical ontology. 188
S21 SemEval-2023 Task 3 persuasion techniques crosswalked to the canonical ontology. 190
S22 Influence Operation Ontology domains crosswalked to the message-internal ontology architecture. 194
S23 DISARM object types crosswalked to the message-internal ontology. 196
S24 Lenk's four analytical levels crosswalked to the message-internal ontology architecture. 197

S25 Institute for Propaganda Analysis devices crosswalked to canonical Method/Tactic operations. 199
S26 Entman's framing functions crosswalked to the ontology architecture. 201
S27 Benford and Snow's core framing tasks crosswalked to the ontology architecture. 202
S28 Scope comparison with campaign- and operation-level frameworks. 203
S29 Descriptive effort profile for the 34-item development-stage annotation exercise. 207
S30 Case 1: schema migration of the development-stage analyst record and final-codebook adjudication. 210
S31 Case 2: schema migration of the development-stage analyst record and final-codebook adjudication. 211
S32 Case 3: schema migration of the development-stage analyst record and final-codebook adjudication. 215
S33 Provenance of canonical Persuasion Tactics. 220
S34 Principal Persuasion decision ledger. 226
S35 Provenance of canonical Framing Tactics. 229
S36 Principal Framing inclusion, merge, split, and boundary decisions. 233
S37 Provenance of canonical Manipulation Tactics. 236
S38 Principal Manipulation inclusion, distinction, and decomposition decisions. 240
S39 Provenance of canonical Coercion Tactics. 245
S40 Principal Coercion inclusion, distinction, and structural decisions. 247
S41 Provenance of canonical Deception Tactics. 250
S42 Principal Deception inclusion, distinction, and boundary decisions. 253
S43 Provenance of canonical Distraction Tactics. 258
S44 Principal Distraction inclusion, distinction, and boundary decisions. 261
S45 Disposition of selected non-canonical Method/Tactic candidates. 266
S46 Conceptual provenance and inclusion rules for TargetConstructs. 272
S47 Principal TargetConstruct boundary decisions. 274
S48 Deterministic TargetConstruct mapping audit. 275
S49 Conceptual provenance of canonical NarrativeDevices. 277
S50 Principal Narrative cross-layer boundary decisions. 279
S51 Provenance and deterministic signatures of NarrativeArchetypes. 281
S52 Provenance of Polarization directions. 284
S53 Conceptual provenance of Polarization axes. 285

S54 Conceptual provenance and deterministic axis assignment of canonical Polarization motifs. 286
S55 Principal Polarization cross-layer boundary decisions. 289
S56 Source-taxonomy provenance of TopicDomain. 294
S57 Principal TopicDomain adaptation decisions. 294
S58 Principal TopicDomain cross-layer boundary decisions. 296
S59 Principal construct-design decisions underlying AboutObject. 298
S60 AboutObject domain provenance and canonical L3 inventory counts. 300
S61 Principal AboutObject inclusion, consolidation, and boundary decisions. 302
S62 Global AboutObject cross-layer boundary rules. 306

A Method dimension: TargetConstruct foundations and derivation

Appendices A and B jointly specify the *Method* dimension of the ontology. Observable message-internal cues support the annotation of *Tactics*; each Tactic belongs to one of six Method families: Persuasion, Framing, Manipulation, Coercion, Deception, or Distraction. *TargetConstruct* is not an independently annotated dimension. Instead, one or more TargetConstruct values are inherited deterministically from the annotated Tactic according to the fixed mappings specified in Appendix B.

The Method dimension therefore distinguishes interpretive assignment from deterministic derivation. Observable message-internal cues support Tactic assignment under the operational definitions and boundary rules. Once a Tactic is assigned, its Method family and fixed TargetConstruct set are derived deterministically. Consequently, the same Tactic always yields the same Method family and TargetConstruct set wherever it is annotated. TargetConstruct is not inferred separately from the surrounding message and does not represent an empirically observed audience effect.

When a Tactic maps to multiple TargetConstruct values, the set represents the stable theoretically supported target profile of the Tactic class. This does not imply that every individual instance is directed equally or simultaneously at every member of that set. The fixed mapping records the construct profile associated with the Tactic class rather than an instance-specific inference about audience effects.

The TargetConstruct vocabulary comprises six operational categories: *Perception/Attention*, *Belief*, *Attitude*, *Identity*, *Agency/Choice*, and *Behavior*. These categories characterise the audience-state dimensions constitutively associated with the ontology's Tactics. They are not proposed as an exhaustive taxonomy of psychological processes or as a temporal or causal sequence of influence effects.

Because TargetConstruct is inherited deterministically from Tactic assignment, the definitions below provide the conceptual interpretation and theoretical basis of the fixed mappings rather than a separate annotation procedure. Annotators do not select additional TargetConstruct values from the message.

A TargetConstruct is mapped only where the corresponding audience-state dimension is part of the Tactic's constitutive influence operation. A psychological mechanism through which a Tactic may operate, a feature that makes its leverage effective, or an audience state that may arise downstream does not by itself license that TargetConstruct. Fixed mappings therefore represent what a Tactic characteristically operates on rather than the full causal pathway through which an instance may exert influence.

A.1 Perception / Attention

Perception/Attention captures attentional access, allocation, selection, processing priority, and relative salience (Posner, 1992; Lang, 2000; Entman, 1993; Mack and Rock, 1998; Simons and Chabris, 1999). It is mapped where the defining operation of a Tactic requires a recoverable change in the relative accessibility, allocation, priority, or salience of message content—for example, by foregrounding one element while backgrounding another, redirecting attention from one focal object to another, obscuring otherwise available information, or burdening processing with competing material.

Mere presentation, emphasis, vividness, emotionality, evaluative force, or importance of a substantive warrant does not by itself yield a *Perception/Attention* mapping. Where the claimed salience relation depends only on imagining an alternative wording or unspecified competing considerations rather than on a contrast or attentional relation recoverable from the defining operation, the construct is not inherited.

A.2 Belief

Belief captures propositional representations concerning states, relations, causes, consequences, and normative expectations (Fishbein and Ajzen, 1975; Ajzen, 1991; Kunda, 1990). It is mapped where a Tactic constitutively operates on what is represented or accepted as being the case, including the establishment, support, dispute, reframing, distortion, or undermining of a proposition or relation. Belief is distinct from *Attitude*, which concerns evaluative stance rather than propositional acceptance.

A.3 Attitude

Attitude captures favourable or unfavourable evaluative orientation toward an object, actor, issue, policy, behaviour, or outcome (Ajzen, 1991; Petty and Cacioppo, 1986). It is mapped where changing evaluation—for example approval or disapproval, attraction or aversion, desirability or undesirability—is intrinsic to the Tactic rather than a possible consequence of changing beliefs or behaviour.

A.4 Identity

Identity captures influence directed at personal self-concept or social group identification (Tajfel and Turner, 1979; Bem, 1972; Cialdini and Goldstein, 2004). It is mapped where activation, affirmation, threat, strengthening, weakening, or redefinition of who the recipient is, aspires to be, or belongs with is constitutive of the Tactic. Identity relevance arising only as a possible consequence of another operation does not yield the mapping.

A.5 Agency / Choice

Agency/Choice is an operational umbrella for influence directed at the recipient's capacity or perceived capacity for self-directed judgement and action. Its epistemic component concerns the capacity to monitor, comprehend, evaluate, and scrutinise information independently; its practical component concerns perceived freedom, capability or control, and the availability, feasibility, or viability of courses of action (Bandura, 2001; Proust, 2008; Brehm, 1966; Ajzen, 1991; Bandura, 1997).

For the practical component, a Tactic must constitutively alter, constrain, scaffold, or represent a material change in the availability, feasibility, controllability, or practical viability of courses of action. Merely supplying reasons for preferring one available option over another, changing its evaluative attractiveness, or representing different expected benefits or costs does not by itself yield an *Agency/Choice* mapping. An option may remain physically possible while being represented as practically non-viable relative to a relevant protected goal, as where a fear appeal represents a non-protective course as exposing the recipient to serious threatened harm.

A.6 Behavior

Behavior captures observable conduct that an influence Tactic is structured to elicit, inhibit, maintain, or alter. Behavioural theories distinguish action from its psychological antecedents: in the Theory of Planned Behavior, behavioural intention is the immediate antecedent of behaviour, while perceived behavioural control can additionally contribute to behaviour where performance is not fully under volitional control (Ajzen, 1991). Social-influence research likewise treats *compliance* and *conformity* as behavioural outcomes of influence, including procedures designed to increase agreement with or performance of a requested act (Cialdini and Goldstein, 2004; Freedman and Fraser, 1966; Cialdini et al., 1975).

Within the ontology, *Behavior* is inherited from Tactics whose constitutive operation is directed toward observable conduct, for example, performing, choosing, joining, sharing, complying, refraining, or otherwise changing an action. The mapping is stable at the Tactic level: it does not require independent identification of a behavioural target from each message, nor does it imply that the audience actually performed the targeted action.

Behavior is distinguished from *Agency/Choice*: Behavior concerns the action toward which the Tactic is directed, whereas Agency/Choice concerns perceived freedom, capability, control, or available alternatives for acting. A Tactic should therefore map to *Behavior* only when influencing conduct is intrinsic to that Tactic, not merely because changes in belief, evaluation, identity, or attention could ultimately affect behaviour.

Supplementary Table S1: Reverse index of the fixed TargetConstruct mappings specified in Appendix B. TargetConstruct values are inherited deterministically from Tactic assignments and are not independently annotated.

TargetConstruct	Tactics mapping to construct (see App. B)
Perception / Attention	Attribute framing; Affective associative cueing; Induced hypocrisy / self-inconsistency pressure; Salience-based concealment; Whataboutism; Red herring; Straw man; Obfuscation (jargon/complexity); Agenda shifting / evasive refocusing; Information flooding / overload; Humor as deflection; Ad hominem / personal-credibility attack;
Belief	Appeal to authority; Social proof (bandwagon); Appeal to tradition; Appeal to novelty; Storytelling / narrative packaging; Two-sided with refutation; Preemptive inoculation / prebunking; Doubt-seeding; Appeal to anger / indignation; Presuppositional embedding; Repetition; Statistical evidence; Source / channel credibility; Associative transfer / guilt-or-honor by association; Unity (shared identity); False dilemma; Causal oversimplification; Slippery-slope / consequential escalation; Scapegoating; Moral framing; Metaphor / analogy; Dehumanizing representation; Inevitability framing; Collective nostalgic framing; Threat framing; Guilt trip; Love-bombing; Appeal to pity / sympathy; Pride / ego appeals; Gaslighting / epistemic destabilization; Exaggeration / overstatement; Minimization / understatement; Strategic vagueness / equivocation; Selective evidence presentation (cherry-picking); False balance / distorted evidential parity; False equivalence / distorted comparability; Salience-based concealment; Whataboutism; Straw man; Ad hominem / personal-credibility attack.
Attitude	Ingratiation; Appeal to anger / indignation; Appeal to tradition; Appeal to novelty; Storytelling / narrative packaging; Preemptive inoculation / prebunking; Unity (shared identity); Associative transfer / guilt-or-honor by association; Scarcity appeal; Loaded language; Name-calling / labeling; Ridicule / mockery; Slippery-slope / consequential escalation; Scapegoating; Dehumanizing representation; Moral framing; Collective nostalgic framing; Attribute framing; Guilt trip; Appeal to pity / sympathy; Love-bombing; Pride / ego appeals; Exaggeration / overstatement; Minimization / understatement; Selective evidence presentation (cherry-picking); False equivalence / distorted comparability; Humor as deflection; Ad hominem / personal-credibility attack.
Identity	Unity (shared identity); Altercasting; Collective nostalgic framing; Shaming / derogation; Pride / ego appeals; Gaslighting / epistemic destabilization.
Agency / Choice	Appeal to fear; Preemptive inoculation / prebunking; Scarcity appeal; False dilemma; Inevitability framing; Manipulative choice architecture; Gaslighting / epistemic destabilization; Explicit threats; Implied threats / coercive intimidation; Blackmail / disclosure leverage; Coercive ultimatum; Collective punishment threats; Obfuscation (jargon/complexity); Information flooding / overload.
Behavior	Foot-in-the-door (FITD); Door-in-the-face (DITF); Authority-based directive; Social proof (bandwagon); Appeal to fear; Appeal to anger / indignation; Reciprocity; Commitment / consistency; Low-ball; Unity (shared identity); Altercasting; Scarcity appeal; Gain/loss goal framing; Risky-choice framing; Guilt trip; Induced hypocrisy / self-inconsistency pressure; Appeal to pity / sympathy; Love-bombing; Shaming / derogation; Pride / ego appeals; Dangling contingent rewards; Manipulative choice architecture; Explicit threats; Implied threats / coercive intimidation; Blackmail / disclosure leverage; Coercive ultimatum; Collective punishment threats.

B Method dimension: Extended cue–tactic catalogue

Together with Appendix A, this appendix specifies the *Method* dimension of the ontology. It provides the canonical catalogue of Tactics within the six Method families (Persuasion, Framing, Manipulation, Coercion, Deception, and Distraction), together with the observable cues used to identify them.

Each entry specifies: (i) a Tactic definition, (ii) key spotting cues, (iii) the fixed TargetConstruct mapping inherited whenever that Tactic is annotated, and (iv) the literature supporting the Tactic and its mapping. Annotators identify observable cues and assign the corresponding Tactic; Method family and TargetConstruct values are then derived through fixed ontology mappings rather than independently selected from the message.

The boldface Tactic names in this appendix are the canonical annotation labels. Table S1 and all cross-references to Tactics use these labels verbatim. Informal descriptions, abbreviations, or synonyms may appear in explanatory prose but do not constitute separate annotation labels.

B.1 Method families

Each canonical Tactic belongs to exactly one of six Method families: *Persuasion*, *Framing*, *Manipulation*, *Coercion*, *Deception*, or *Distraction*. The families group Tactics by the constitutive communicative operation through which they perform influence work. They classify message form and function, not communicator motive, truth status, or achieved effect.

Method family is not independently annotated. Annotators determine whether message-internal evidence satisfies the operational definition and minimum-evidence conditions of a canonical Tactic. Once a Tactic is assigned, its Method family is inherited deterministically from the canonical mapping. The family definitions in Table S2 therefore provide the conceptual interpretation of that mapping and a basis for inspecting Tactic placement; they do not introduce a second annotation decision.

Neighbouring families may share surface features, psychological mechanisms, or downstream consequences. Family placement therefore follows the *constitutive operation* specified by the Tactic definition rather than isolated lexical features, presumed communicator intent, anticipated audience response, or the presence of a particular TargetConstruct. Where neighbouring Tactics remain plausible, the Tactic-level include–exclude and boundary rules in this appendix govern assignment.

Supplementary Table S2: Operational interpretation of the six Method families.

Method family	**Distinguishing communicative operation**
Persuasion	Supplies a represented feature of a claim, source, situation, social relation, or available response as a reason for accepting, qualifying, or withholding acceptance of a proposition, adopting or revising an evaluation, making a choice, or performing a course of action. The influence operation works through an evidential, argumentative, social, identity-based, motivational, or compliance-relevant warrant represented in the message (Toulmin, 2003; Petty and Cacioppo, 1986; Chaiken, 1980; Cialdini, 2007; O'Keefe, 2016). Persuasion includes both substantive warrants and non-substantive influence cues where those cues themselves provide the represented basis for judgement or response. Under the canonical mapping, no Persuasion Tactic inherits *Perception/Attention*.
Framing	Structures how a target, event, option, cause, consequence, attribute, or relation is construed or evaluated by selecting, foregrounding, organising, categorising, or representing it in a particular way. The defining operation lies in the representation through which the focal matter is selectively represented, construed, or evaluatively positioned, including lexical, categorical, causal, temporal, moral, metaphorical, gain–loss, and attribute-based forms (Entman, 1993; Chong and Druckman, 2007; Levin et al., 1998). Framing does not require an alternative wording to be present in the message unless the individual Tactic definition specifies a recoverable contrast, reference point, or complementary representation.
Manipulation	Uses an affective, self-evaluative, relational, motivational, epistemic, or decision-environment condition as leverage over judgement or conduct. The relevant condition or relation itself performs the steering operation: for example guilt or obligation, sympathy, shame, self-image, attachment to a source, affective salience, a source-controlled reward, asymmetric access to available responses, or destabilisation of the recipient's confidence in their own judgement (Noggle, 1996; Coons and Weber, 2014; Susser et al., 2019; Mathur et al., 2019). The family classification concerns the observable form of leverage and does not require a judgement that the communicator acted deceptively, improperly, or with manipulative intent.

Continued on next page

Supplementary Table S2: Operational interpretation of the six Method families (continued)

Method family	**Distinguishing communicative operation**
Coercion	Uses a prospective adverse consequence as contingent leverage over identifiable conduct. The message establishes or makes recoverable a relation between an act, omission, continuation, cessation, or other course of conduct and a prospective sanction, harm, deprivation, exposure, loss, or comparable adverse consequence, and communicatively deploys that consequence to compel or deter behaviour (Schelling, 1966; Raven, 1993; George, 1991; Lamond, 1996). The defining relation is the conduct–consequence contingency, not merely frightening content, hostile tone, moral condemnation, or a representation of danger. Under the canonical mapping, all Coercion Tactics inherit *Agency/Choice* and *Behavior*.
Deception	Presents content in a potentially misleading relation that is recoverable from the message itself without requiring external fact-checking or inference of deceptive intent. The constitutive operation may concern, for example, a recoverable comparison or evidential baseline, the completeness of a represented evidence set, the relative standing of competing evidence, the comparability of represented cases, or the presentation and perceptual availability of a material qualification. The family is therefore defined at the level of misleading communicative form rather than lying: falsity and intent may be relevant to lying, but neither is required for annotation here (Chisholm and Feehan, 1977; Saul, 2012; Mahon, 2016; Rogers et al., 2017). Mere falsity, unsupported assertion, ambiguity, omission, impersonation, or insincerity is insufficient where the relevant misleading relation cannot be established from message-internal evidence.

Continued on next page

Supplementary Table S2: Operational interpretation of the six Method families (continued)

Method family	**Distinguishing communicative operation**
Distraction	Redirects, substitutes, obscures, or burdens engagement with an identifiable focal matter recoverable from the available discourse. The operation may introduce comparative or tangential material, substitute another issue or position for the matter under scrutiny, evade or refocus an established question, overload processing with competing material, obscure focal information through complexity, or displace substantive engagement through humorous or person-directed material (Walton, 2004; Walton, 2005; King et al., 2017; Roberts, 2018). The defining relation is displacement or obstruction of engagement with the focal matter, not simply the presence of additional information, complexity, humour, criticism, or topic change. Under the canonical mapping, all Distraction Tactics inherit *Perception/Attention*; no other Method family has *Perception/Attention* as a TargetConstruct for every one of its members.

Cross-family placement. A structural feature that occurs in more than one family does not by itself determine family membership. Canonical placement follows the operation that constitutes the Tactic as defined by its full evidence and boundary rules. For example, *False dilemma* may restrict a recoverable set of alternatives, but its defining operation is the framing of the available choice set; it therefore belongs to Framing rather than Deception. *Straw man* may misrepresent a recoverable position, but its defining operation is the substitution of a materially different position for the one under scrutiny and the redirection of attack toward that substitute; it therefore belongs to Distraction.

Persuasion and Manipulation may be adjacent where social, affective, identity-relevant, or motivational material contributes to a response. Their distinction does not turn on whether that material can be described as a reason or warrant in ordinary language. Canonical placement instead follows the particular influence relation constituted by the Tactic. Persuasion Tactics characteristically make a represented feature of the claim, source, situation, social relation, prior exchange, or available response part of the basis for judgement or conduct. Manipulation Tactics characteristically make an affective, self-evaluative, relational, motivational, epistemic, or decision-environment condition itself the leverage through which judgement or conduct is steered.

Several Tactic-level boundaries illustrate this distinction. *Commitment / consistency* invokes a prior choice, pledge, promise, or commitment as a reason for follow-through, whereas *Induced hypocrisy / self-inconsistency pres-*

sure confronts the recipient with a discrepancy between an endorsed standard and their own represented conduct and uses that discrepancy as pressure. *Ingratiation* cultivates favourable evaluation or liking of the source, whereas *Love-bombing* uses unusually intense or accelerated relational attachment as leverage. *Altercasting* assigns or activates a social or relational role whose associated expectation provides the influence basis, whereas *Pride / ego appeals* use a positively valued self-image, competence, or status as leverage for the promoted response. These examples illustrate the family-level distinction; the full Tactic definitions and their explicit boundary rules remain authoritative for annotation.

The same principle separates neighbouring influence relations involving authority, benefits, and sanctions. Where represented authority or standing itself functions as the basis for compliance, the operation is *Authority-based directive* within Persuasion. Where a prospective benefit controlled or offered by the source is made conditional on compliance, the operation is *Dangling contingent rewards* within Manipulation. Where a prospective adverse consequence is made contingent on conduct and used as leverage, the operation belongs to Coercion (Raven, 1993). These placements do not imply that the respective Tactics cannot co-occur; co-coding is permitted where the message independently satisfies the operational criteria for each Tactic.

The family definitions are consequently descriptive of the canonical Tactic architecture rather than independent coding rules. Annotation proceeds from message-internal evidence to canonical Tactic assignment; Method family and the fixed TargetConstruct profile are then inherited from that assignment.

B.2 Cue annotation and Tactic assignment

A *Cue* is the observable message-internal evidence that licenses a Tactic assignment under the corresponding "How to spot" rule. Annotators mark the shortest contiguous span that contains sufficient evidence for the Tactic. Where the defining evidence is relational or sequential—for example, two requests, a claim and its rebuttal, or a prior commitment and a later request—the selected span should be the shortest span containing the elements required to establish that relation.

Tactic assignment is based on communicative function evidenced in the message, not on isolated lexical markers. A title, number, group pronoun, scarcity term, gift term, or other surface feature is therefore insufficient unless it performs the operation specified by the Tactic definition and spotting rule. Phrases such as "in order to" in the definitions describe the observable rhetorical function of the message and do not require an inference about the communicator's private intention.

Emotionality alone does not license a Tactic. Fear, guilt, pity, pride, affection, anger, sadness, disgust, enthusiasm, or other affective content should

be assigned a Tactic only where the message independently satisfies the defining communicative relation specified for that Tactic; annotators do not infer a generic "appeal to emotion" from emotional intensity or likely audience affect.

For sequential Tactics, a defining stage may have occurred before the annotated message only when that prior stage is explicitly recoverable from the message itself. Annotators must not supply an unobserved prior request, commitment, rejection, or audience response from assumption or external context.

Cue spans and Tactics are not mutually exclusive. The same span may support multiple Tactic assignments when each Tactic's defining conditions are independently satisfied. Overlapping assignments should not be suppressed merely because the Tactics share lexical material or inherit the same TargetConstruct.

Quoted, reported, or mentioned material is assigned a Tactic only when the message deploys that material as part of its own influence operation. Mere description, criticism, negation, or metalinguistic mention of a tactic is insufficient. Annotators assess whether authority, credibility, evidence, scarcity, consensus, or another warrant is invoked; they do not verify the truth of the claim, the actual credibility of the source, the communicator's mental state, or the audience's empirical response.

After a Tactic is assigned, its Method family and TargetConstruct set are derived from the fixed mappings in this appendix and are not independently selected from the cue span. TargetConstruct mappings follow the constitutive-operation criterion specified in Appendix A: mechanisms, rhetorical levers, likely psychological reactions, and downstream effects do not independently add TargetConstruct values unless the corresponding operation is constitutive of the Tactic class.

Tactic assignment characterises the communicative operation evidenced in the message and does not by itself imply deceptive, malicious, exploitative, or otherwise ill intent. A communicator may sincerely regard the influence operation as appropriate, justified, beneficial, or necessary. Method-family labels such as *Manipulation* and *Deception* are therefore used as operational classifications of communicative form and function within the ontology, not as independent judgements about the communicator's motive, moral status, or state of mind. Where a Tactic conventionally carries an intentional or normative connotation in the source literature, annotation requires only the observable operation specified by its definition and spotting rule unless intent is itself explicitly represented in the message.

B.3 Operational quick reference

Table S3 provides a compact lookup surface for the canonical Tactics specified below. It is a derived annotation aid rather than a separate codebook.

The *Core operation* column summarises the defining communicative function, while *Minimum evidence* states the minimum message-internal relation that should be recoverable before the Tactic is considered. Full definitions, exclusions, boundary rules, and overlap conditions in the corresponding entries below remain authoritative. TargetConstruct values in the final column are inherited automatically from Tactic assignment and are not independently annotated. For compactness, P/A = Perception/Attention, Bel = Belief, Att = Attitude, Id = Identity, A/C = Agency/Choice, and Beh = Behavior.

Supplementary Table S3: Operational quick reference for Method Tactics. Full Appendix B entries govern boundary cases and overlapping assignments.

Tactic	Core operation	Minimum evidence	Targets
Persuasion			
Foot-in-the-door (FITD)	Use completed small-request compliance to promote a larger request.	Prior compliance with a smaller request + subsequent larger target request; the sequence must be recoverable.	Beh
Door-in-the-face (DITF)	Retreat from a rejected large request to a smaller target request.	Larger request + rejection/noncompliance + smaller request presented as a concession or retreat.	Beh
Appeal to authority	Use an actor's expertise or recognised authority as an epistemic warrant.	Expertise, credentials, or recognised authority + accompanying claim, judgement, or recommendation + authority functions as a reason to accept it.	Bel
Authority-based directive	Use represented authority to direct conduct.	Source authority relevant to the recipient or conduct + identifiable directive + the authority relation itself warrants compliance.	Beh
Social proof (bandwagon)	Use others' beliefs, approval, choices, or behaviour as social evidence.	Consensus, prevalence, endorsement, or normative behaviour + corresponding judgement or conduct promoted through it.	Bel; Beh
Appeal to tradition	Use longstanding or customary status as a warrant.	Tradition or historical persistence + used as a reason for correctness, legitimacy, preference, or retention.	Bel; Att
Appeal to novelty	Use newness or modernity as a warrant.	Novelty, modernity, or innovation + used as a reason for superiority, correctness, credibility, or desirability.	Bel; Att
Ingratiation	Cultivate favourable evaluation of the source.	Flattery, agreement, similarity, favourable self-presentation, or comparable rapport cue + source liking/favourability is the operative function.	Att
Appeal to fear	Use threatened harm to motivate a protective response.	Threat or harmful consequence + identifiable protective response + the threat functions as a reason for that response.	A/C; Beh
Appeal to anger / indignation	Use an anger-relevant offence or injustice as persuasive leverage.	Identifiable provocation or offence + anger/indignation relation + a claim, evaluation, judgement, position, or conduct promoted through it.	Bel; Att; Beh

Table S3 (continued)

Tactic	Core operation	Minimum evidence	Targets
Storytelling / narrative packaging	Use story form as a persuasive vehicle for claims or evaluations.	Observable event/actor sequence or story structure + story form carries an identifiable persuasive claim or evaluation.	Bel; Att
Two-sided with refutation	Present and rebut an opposing claim or argument.	Identifiable opposing proposition or argument + substantive rebuttal supporting the advocated position or weakening the opposition.	Bel
Preemptive inoculation / prebunking	Prepare resistance to anticipated future influence.	Recoverable future-influence orientation + forewarning + counterargument, diagnostic cue, refutational material, or recognition strategy supplied in advance.	Bel; Att; A/C
Doubt-seeding	Use uncertainty or limitations to weaken confidence in an epistemic target.	Identifiable claim, source, evidence base, consensus, or inference + uncertainty, disagreement, or limitations + reduced confidence or suspended judgement promoted through them.	Bel
Presuppositional embedding	Introduce unestablished content as presupposed background.	Recoverable proposition not already established + encoded as taken for granted or presupposed rather than as the focal assertion.	Bel
Repetition	Reiterate the same proposition as reinforcement.	Same or substantially equivalent propositional content recurs + recurrence functions as reinforcement rather than ordinary restatement.	Bel
Statistical evidence	Use quantitative information as evidential support.	Number, statistic, quantitative comparison, or data display + functions as evidence for a separate proposition or inference.	Bel
Source / channel credibility	Use source or channel provenance as a credibility warrant.	Identifiable outlet, publication channel, or institutional provenance + its reputation or reliability functions as a reason to trust the accompanying claim.	Bel
Associative transfer / guilt-or-honor by association	Transfer represented qualities or standing through association.	Focal target + associated entity + recoverable association + the associate's commitments, qualities, credibility, prestige, stigma, or standing are transferred as a warrant.	Bel; Att
Reciprocity	Use a prior benefit or concession as a warrant for a return act.	Benefit, favour, gift, service, or concession already provided + reciprocal return relation + corresponding act expected or requested.	Beh
Commitment / consistency	Invoke a prior commitment as a reason for follow-through.	Prior explicit choice, pledge, promise, or commitment + corresponding conduct promoted because consistency with it is invoked.	Beh
Low-ball	Worsen terms after commitment while seeking continued compliance.	Prior agreement + later removal of an inducement, addition of cost, or other worsening of terms + continued compliance with the same decision sought.	Beh
Unity (shared identity)	Use shared ingroup identity as an influence warrant.	Recoverable shared membership or "weness" + that identity relation supports a claim, evaluation, or requested conduct.	Bel; Att; Id; Beh

Table S3 (continued)

Tactic	Core operation	Minimum evidence	Targets
Scarcity appeal	Use limited availability to increase value or accelerate action.	Limited time, quantity, supply, access, or continuing availability + scarcity itself functions as a reason to value or act before loss.	Att; A/C; Beh
Altercasting	Assign or activate a recipient role carrying expected conduct.	Recipient role, category, or identity position + associated behavioural expectation + role-consistent response promoted through it.	Id; Beh
Framing			
Loaded language	Frame a target through evaluatively or affectively charged lexical choice.	Marked connotative wording + substantially the same referent could be expressed more neutrally + lexical choice itself supplies the evaluation.	Att
Name-calling / labeling	Characterize a target through a condensed evaluative label.	Person, group, idea, institution, or other target + compact epithet or categorical label + label frames who or what the target is.	Att
Ridicule / mockery	Frame a target as contemptible, absurd, or laughable through mockery.	Identifiable target + derisive, humiliating, or comic treatment + ridicule itself supplies the negative evaluation.	Att
False dilemma	Represent a restricted alternative set as exhaustive.	Restricted alternatives + message forecloses intermediate, combined, neutral, or additional possibilities recoverable from the available context.	Bel; A/C
Causal oversimplification	Represent one cause as sufficient, overriding, or exhaustive.	Outcome or problem + single factor + message treats that factor as the sole, decisive, or comprehensive causal explanation.	Bel
Slippery-slope / consequential escalation	Evaluate an initial step through a projected downstream escalation.	Identifiable initial act or condition + represented progression, cascade, or precedent + later consequential endpoint structures judgement of the initial step.	Bel; Att
Scapegoating	Concentrate blame for a broader negative condition on a target.	Broader harm or problem + identifiable person or group + causal or responsibility attribution concentrates blame on that target.	Bel; Att
Moral framing	Interpret or evaluate a target through moral values or obligations.	Identifiable moral value, right, duty, virtue, vice, injustice, or prohibition + functions as a criterion for interpreting or evaluating the target.	Bel; Att
Metaphor / analogy	Organize interpretation through relational mapping from another domain.	Source or comparison domain + focal target + mapped relational structure used to understand or draw an inference about the target.	Bel
Dehumanizing representation	Represent people or groups through denial or degradation of human standing.	Human target + subhuman, animalistic, mechanistic, contaminating, or comparable dehumanizing representation + representation structures judgement of that target.	Bel; Att
Collective nostalgic framing	Use an idealised shared past to frame the present or future.	Collective past + positively valued or idealised representation + present or future interpreted/evaluated through contrast, loss, continuity, or restoration.	Bel; Att; Id

Table S3 (continued)

Tactic	Core operation	Minimum evidence	Targets
Inevitability framing	Represent an outcome or course as unavoidable or without viable alternatives.	Identifiable development, outcome, or course + necessity, irreversibility, predetermination, or no-alternative representation.	Bel; A/C
Threat framing	Organize interpretation by representing a phenomenon as a danger to a valued target.	Focal phenomenon + identifiable valued object, group, institution, interest, or condition + danger/risk/harm relation functions as the framing operation.	Bel
Gain/loss goal framing	Promote the same focal behaviour through valenced consequences.	Identifiable behaviour + gain/benefit or loss/cost tied to performing or not performing that behaviour.	Beh
Risky-choice framing	Structure choice under risk through gain/loss representation.	Choice alternatives + uncertain or probabilistic outcomes + gain/loss reference framing structures selection among those alternatives.	Beh
Attribute framing	Frame an attribute through a valenced description that structures its salience and evaluation.	Identifiable attribute + recoverable positive/negative framing relation + description makes one aspect of the attribute relatively focal and evaluatively marked.	P/A; Att
Manipulation			
Guilt trip	Use guilt-relevant responsibility, obligation, or indebtedness as pressure on conduct.	Recipient or relevant ingroup is represented as responsible, obligated, indebted, blameworthy, implicated, or in violation + corrective or compliant conduct is pressured through that relation.	Bel; Att; Beh
Induced hypocrisy / self-inconsistency pressure	Foreground self-discrepancy and use it as pressure for alignment.	Recipient-endorsed standard + recipient's inconsistent conduct + discrepancy foregrounded + subsequent aligned or corrective conduct pressured through it.	P/A; Beh
Appeal to pity / sympathy	Use another's suffering or vulnerability as a warrant for a favourable response.	Suffering, hardship, vulnerability, misfortune, or need + pity/sympathy relation + judgement, evaluation, leniency, aid, or conduct promoted through it.	Bel; Att; Beh
Love-bombing	Use unusually intense or accelerated affiliation to establish attachment for influence.	Dense, repeated, intense, or accelerated affection, praise, attention, gifts, or commitment cues + relational closeness/trust is accelerated + subsequent influence or commitment leverage is recoverable.	Bel; Att; Beh
Shaming / derogation	Use negative evaluation of recipient identity or status as pressure on conduct.	Derogatory self/social judgement + linked to requested, expected, or disfavoured conduct + negative classification supplies behavioural pressure.	Id; Beh
Pride / ego appeals	Tie a promoted response to a positively valued self-image or status.	Valued trait, competence, status, or self-image + believing, evaluating, or acting in the promoted way is presented as diagnostic of that positive identity.	Bel; Att; Id; Beh
Dangling contingent rewards	Offer a prospective benefit conditional on conduct.	Identifiable reward or benefit + recoverable conduct–reward contingency + reward functions as instrumental leverage for compliance.	Beh

Table S3 (continued)

Tactic	Core operation	Minimum evidence	Targets
Affective associative cueing	Use an affectively significant association to alter salience or attention.	Focal representation + affective, moral, social, or identity-relevant cue or association + association is activated or foregrounded to change relative salience.	P/A
Manipulative choice architecture	Steer choice through asymmetric access, effort, default, sequence, or reversibility.	Identifiable alternatives + structurally unequal action paths, effort, defaults, sequencing, or reversal conditions + asymmetry privileges one response.	A/C; Beh
Gaslighting / epistemic destabilization	Undermine recipient self-trust as an independent perceiver or judge.	Recipient perception, memory, reasoning, or judgement challenged + the recipient's epistemic competence or reliability itself is discredited.	Bel; Id; A/C
Coercion			
Explicit threats	Use an explicitly stated adverse consequence as contingent leverage over conduct.	Target conduct + explicit prospective harm or sanction + explicit or directly recoverable contingency + harm communicatively used as leverage.	A/C; Beh
Implied threats / coercive intimidation	Use an indirectly communicated adverse consequence as leverage over conduct.	Target conduct + prospective harm + coercive contingency/leverage recoverable, with one or more components conveyed indirectly.	A/C; Beh
Blackmail / disclosure leverage	Use prospective disclosure or non-disclosure of information as coercive leverage.	Information + prospective disclosure/non-disclosure + target conduct + contingency linking the information leverage to that conduct.	A/C; Beh
Coercive ultimatum	Add finality or bounded compliance to a sanction-backed demand.	Target conduct + prospective sanction used as leverage + explicit or recoverable final, non-negotiable, last-chance, or deadline condition.	A/C; Beh
Collective punishment threats	Use distributed sanctions against associated others as leverage over conduct.	Target conduct + prospective sanction extends beyond the immediately responsible actor to additional group members + distributed liability is used as leverage.	A/C; Beh
Deception			
Exaggeration / overstatement	Inflate represented degree relative to a recoverable baseline.	Target quantity, frequency, severity, likelihood, certainty, effect, importance, or other degree + recoverable baseline + communicated degree materially exceeds it.	Bel; Att
Minimization / understatement	Downscale represented degree relative to a recoverable baseline.	Target quantity, severity, harm, risk, responsibility, importance, or other degree + recoverable baseline + communicated degree materially reduces it.	Bel; Att
Strategic vagueness / equivocation	Leave a materially relevant matter underdetermined.	Context makes greater determinacy relevant + wording leaves a proposition, referent, quantity, scope, agent, responsibility, stance, condition, or commitment materially unresolved.	Bel
Selective evidence presentation (cherry-picking)	Use a favourable subset from a recoverable broader evidential set.	Broader evidence or consideration set recoverable + systematically supportive subset selected + countervailing members excluded, omitted, or denied comparable evidential use.	Bel; Att

Table S3 (continued)

Tactic	Core operation	Minimum evidence	Targets
False balance / distorted evidential parity	Represent recoverably unequal evidential standing as more balanced.	Competing positions, claims, sources, or evidence + recoverable asymmetry in support, prevalence, consensus, or credibility + subsequent parity/balance representation.	Bel
False equivalence / distorted comparability	Represent materially different comparands as equivalent on a relevant dimension.	Two or more comparands + communicated parity/equivalence dimension + recoverable material difference on that same dimension that is ignored, collapsed, or overridden.	Bel; Att
Salience-based concealment	Keep qualifying information present while making it markedly less perceptually accessible.	Focal representation + materially qualifying information + marked presentation asymmetry reducing the qualification's salience or perceptual accessibility.	P/A; Bel
Distraction			
Whataboutism	Deflect scrutiny through a comparative or counteraccusatory case that also challenges the focal judgement.	Established criticism, accusation, question, or scrutiny + countercase or comparison + attention redirected + countercase used to qualify or undermine the focal judgement.	P/A; Bel
Red herring	Divert attention from an identifiable focal issue with a non-resolving consideration.	Established focal issue or question + introduced consideration + displacement of engagement + consideration does not address the focal matter in the role for which it is introduced.	P/A
Straw man	Attack a materially altered substitute for an opponent's recoverable position.	Original position recoverable + materially different representation attributed to the same actor + criticism or rebuttal targets the substitute representation.	P/A; Bel
Obfuscation (jargon/complexity)	Make focal content materially harder to recover or scrutinise through linguistic opacity.	Identifiable focal content + unexplained jargon, technical density, abstraction, nominalisation, syntactic complexity, or comparable opacity + material processing/comprehension burden.	P/A; A/C
Agenda shifting / evasive refocusing	Replace the focal issue or criterion with a different agenda.	Established question, criticism, issue, or dimension + response substitutes a different actor, time frame, criterion, comparison, level, or question as the focal matter.	P/A
Information flooding / overload	Overwhelm selection and scrutiny with dense competing information.	Identifiable focal material + unusually dense or high-volume competing content + substantially increased selection, prioritisation, or processing burden.	P/A; A/C
Humor as deflection	Use humour or another non-serious mode to displace substantive engagement.	Established criticism, question, accusation, or challenge + humorous, ironic, sarcastic, or playful response + humour substitutes for or materially displaces substantive engagement.	P/A; Att

Table S3 (continued)

Tactic	Core operation	Minimum evidence	Targets
Ad hominem / personal-credibility attack	Use negative person-characterisation to discount their contribution.	Identifiable claim, argument, criticism, testimony, question, or contribution + negative characterisation of the contributor + characterisation is used to reduce the credibility or standing of that contribution.	P/A; Bel; Att

B.4 Persuasion

- **Foot-in-the-door (FITD).** Secure compliance with an initial small request to increase the likelihood of compliance with a subsequent, larger target request (Freedman and Fraser, 1966; Burger, 1999; Cialdini and Goldstein, 2004). *How to spot:* message-internal evidence establishes compliance with an initial small request and a subsequent, larger target request, with the prior compliance functioning as the step toward the larger act. The initial compliance may have occurred before the annotated message, but it must be explicitly recoverable from message-internal evidence. Merely presenting a small and a large request together before initial compliance is insufficient. *Target constructs: Behavior*—the defining operation is to use initial compliance to increase subsequent compliance with the larger target request.

- **Door-in-the-face (DITF).** Present an initially extreme request and, after the recipient rejects it, retreat to a smaller target request, thereby increasing the likelihood of compliance with the smaller request (Cialdini et al., 1975; O'Keefe and Hale, 1998; Cialdini and Goldstein, 2004). *How to spot:* message-internal evidence establishes an initial larger request, its rejection or noncompliance, and a subsequent smaller target request presented as a concession or retreat. The initial request or its rejection may have occurred before the annotated message, but the defining sequence must be explicitly recoverable from message-internal evidence. Merely presenting a large and a smaller request together before rejection, or simply lowering a request without evidence of prior rejection, is insufficient. *Target constructs: Behavior*—the defining operation is to use the retreat from a rejected larger request to increase compliance with the smaller target request.

- **Appeal to authority.** Invoke the expertise, credentials, professional status, leadership position, or recognised authority of a person or actor as a warrant for accepting a claim, judgement, or recommendation (Cialdini, 2007; Petty and Cacioppo, 1986; Chaiken, 1980). *How to spot:* titles, credentials, professional roles, specialist expertise, leadership positions, or expert endorsements are used as a reason to accept the accompanying claim or recommendation; the authority cue must

function as a warrant rather than merely identify the speaker. *Target constructs: Belief*—the defining operation is to use the expertise or recognised authority of an actor as evidence supporting acceptance of a proposition, judgement, or recommendation.

- **Authority-based directive.** Invoke the source's represented institutional, hierarchical, legal, organisational, professional, or other recognised authority to direct the recipient's conduct, such that the authority relation itself functions as a warrant for compliance (Raven, 1993; Tyler, 2006; Milgram, 1963; Cialdini and Goldstein, 2004). *How to spot:* the message establishes or invokes a source as occupying an authority position relevant to the recipient or the conduct at issue and uses that position as a basis for directing, requiring, permitting, prohibiting, continuing, terminating, or otherwise altering identifiable conduct (e.g., "as your supervisor, I am directing you to submit this by five", "by order of the chair, leave the chamber", "as the attending physician, I am instructing you to stop the procedure now"). A title, office, rank, professional role, or imperative alone is insufficient unless the source's represented authority to direct the recipient functions as the warrant for compliance. Annotators do not determine whether the source actually possesses the claimed authority, whether the directive is lawful, legitimate, justified, appropriate, or binding, or whether the recipient accepts the authority or complies. *Target constructs: Behavior*—the defining operation uses the source's represented authority to direct the recipient toward performing, continuing, refraining from, terminating, permitting, or otherwise altering identifiable conduct (Tyler, 2006; Milgram, 1963). *Boundary:* Code *Appeal to authority* when expertise, credentials, status, or recognised authority functions as an epistemic warrant for accepting a claim, judgement, or recommendation. Code *Authority-based directive* when the source's represented standing or authority to direct the recipient itself functions as the warrant for compliance with identifiable conduct. Code an applicable Coercion Tactic additionally when noncompliance is linked to a prospective adverse consequence; authority alone does not constitute coercion. Code *Dangling contingent rewards* additionally when compliance is linked to a prospective benefit controlled or offered by the source. Code *Altercasting* when the influence warrant instead consists in assigning or activating a role or identity position for the recipient and invoking conduct expected of that role. These Tactics may co-occur only where their respective operations are independently evidenced.

- **Social proof (bandwagon).** Use information about others' beliefs, judgements, approval, choices, or behaviour as social evidence for how the audience should judge or act (Deutsch and Gerard, 1955; Cialdini

et al., 1990; Schultz et al., 2007; Goldstein et al., 2008; Cialdini, 2007). *How to spot:* claims of consensus, widespread endorsement, prevalence, or normative behaviour (e.g., "most people agree", "9 out of 10 recommend", "millions have joined", "most guests reuse their towels") are presented as a reason for corresponding judgement or conduct. *Target constructs: Belief*—others' judgements or endorsements are used as social evidence bearing on what should be accepted as correct, credible, or appropriate; *Behavior*—others' behaviour or approval is used to promote corresponding or norm-consistent conduct. *Boundary:* Code *Social proof (bandwagon)* when information about others' beliefs, approval, choices, or behaviour functions as social evidence for corresponding judgement or conduct. Code *Unity (shared identity)* when shared membership or collective "we-ness" itself functions as the influence warrant. A reference to what fellow ingroup members believe or do may instantiate both only where both the prevalence or normative-information warrant and the shared-identity warrant are independently recoverable.

- **Appeal to tradition.** Invoke the longstanding, inherited, or customary status of a belief, practice, institution, or arrangement as a warrant for accepting or favouring it (Michaud, 2018). *How to spot:* historical persistence, inherited custom, established practice, or continuity across generations is presented as a reason that a claim, practice, institution, or position is correct, reliable, legitimate, preferable, or worthy of retention (e.g., "we have always done it this way", "this tradition has stood for generations", "time-tested"). Mere mention or description of history, heritage, age, or tradition is insufficient unless longstandingness or customary status functions as an influence warrant. *Target constructs: Belief*—traditional persistence or historical acceptance is used as a warrant bearing on the correctness, reliability, legitimacy, or acceptance of a proposition, practice, or institution; *Attitude*—traditional status is used to position the target as preferable, proper, valuable, or worthy of approval (Michaud, 2018). *Boundary:* Code *Appeal to tradition* when longstandingness, customary status, or historical persistence itself functions as a reason for accepting or favouring the target. Code *Collective nostalgic framing* when a positively valued shared past instead provides an interpretive and evaluative frame for understanding the present or future. The two may co-occur only where the communication both invokes traditional status as a warrant and independently constructs a valued shared past as a temporal frame.

- **Appeal to novelty.** Invoke the newness, modernity, recency, innovation, or technological advancement of a belief, practice, product, policy,

institution, or course of action as a warrant for accepting or favouring it (Biržietienė and Gabrėnaitė, 2014). *How to spot:* terms such as "new", "modern", "latest", "next-generation", "cutting-edge", "state-of-the-art", or comparable claims of novelty or advancement are presented as a reason that the target is more correct, reliable, credible, desirable, preferable, or otherwise superior. Mere description of something as new or recent is insufficient unless its newness or modernity functions as an influence warrant. *Target constructs: Belief*—newness, recency, or modernity is used as a warrant bearing on the correctness, reliability, credibility, or acceptance of a claim, practice, or position; *Attitude*—newness, innovation, or modernity is used to position the target as preferable, desirable, advanced, or superior (Biržietienė and Gabrėnaitė, 2014). *Boundary:* Code *Appeal to novelty* when newness, modernity, innovation, or recency itself functions as a reason that a target should be accepted or favoured. Code *Inevitability framing* when a development or course is represented as unavoidable, predetermined, necessary, or effectively without viable alternatives. The two may co-occur only where novelty or modernity independently functions as a warrant and the development is also represented as inevitable.

- **Ingratiation.** Use strategic interpersonal behaviours to increase the source's attractiveness, likability, or favourability in the eyes of the target (Jones, 1964; Gordon, 1996; Cialdini and Goldstein, 2004). *How to spot:* flattery or other-enhancement, strategic agreement or opinion conformity, favourable self-presentation, expressions of similarity, or comparable rapport-building cues are used to cultivate a more favourable evaluation of the source. *Target constructs: Attitude*—the defining operation is to increase favourable evaluation, interpersonal attraction, liking, or approval of the source; subsequent agreement or compliance may result from that favourable evaluation but is not constitutive of the Tactic. *Boundary:* So-called "plain-folks" or common-person self-presentation is coded as *Ingratiation* where ordinariness, relatability, similarity, informality, or an "I am like you" presentation functions to cultivate a more favourable evaluation of the source. Code *Unity (shared identity)* instead, or additionally where independently satisfied, when shared membership in an identifiable collective, social category, or "we" functions as the warrant for accepting a claim, adopting an evaluation, or acting accordingly. Ordinary dress, informal vocabulary, familiar settings, everyday activities, or claims of humble background do not by themselves license either Tactic unless the corresponding source-favourability or shared-identity influence operation is recoverable.

- **Appeal to fear.** Present a threatening outcome, danger, or risk as a

reason for adopting, maintaining, or otherwise performing a recommended protective response (Rogers, 1975; Witte and Allen, 2000; Tannenbaum et al., 2015). *How to spot:* the message establishes a threatening or harmful consequence and links that consequence to failure to adopt, or continuation of conduct contrary to, an identifiable recommended protective response. Efficacy information may additionally indicate that the protective response is effective or feasible but is not required for the Tactic. Mere description of danger or harm without an identifiable protective-response relation is insufficient. *Target constructs: Agency/Choice*—the defining threat–response structure represents one or more non-protective courses of action as materially unsafe or practically non-viable relative to a protected goal, thereby altering the perceived viability of available responses; *Behavior*—the represented threatening consequence is used to motivate adoption, maintenance, or performance of an identifiable protective response (Rogers, 1975; Witte and Allen, 2000; Tannenbaum et al., 2015). *Boundary:* Code *Threat framing* when representing a person, group, event, condition, development, or other phenomenon as dangerous, risky, hazardous, or threatening is itself the focal framing operation, without requiring an identifiable protective response. Code *Appeal to fear* when a represented threat or harmful consequence functions as a reason for adopting, maintaining, or performing an identifiable protective response. Do not automatically add *Threat framing* to every *Appeal to fear* merely because representation of danger is a necessary component of the fear appeal. Code both only where the message independently uses the threat characterisation to organise interpretation of the focal phenomenon in addition to using threatened harm as a warrant for the protective response. Code an applicable Coercion Tactic when a prospective adverse consequence is instead communicatively deployed as contingent leverage over identifiable target conduct. These Tactics may co-occur only where their respective operations are independently evidenced.

- **Appeal to anger / indignation.** Invoke or construct an identifiable offence, mistreatment, injustice, violation, betrayal, culpable act, obstruction, or comparable provocation so that anger, indignation, or outrage associated with that appraisal functions as persuasive leverage for accepting or rejecting a claim, adopting an evaluative stance, supporting a judgement or position, or performing or supporting an identifiable response (Walter et al., 2019; Skurka, 2019; Turner et al., 2020; Lerner and Keltner, 2001). *How to spot:* the message establishes an identifiable object or basis of anger or indignation and communicatively recruits that anger-relevant appraisal in support of a proposition, evaluation, judgement, position, or course of conduct. Observable forms

may explicitly invite anger or outrage (e.g., "You should be furious that they concealed this from the public"), present culpable conduct in a way that makes indignation itself part of the warrant (e.g., "After treating those families with such contempt, how can anyone still regard this administration as acting in their interests?"), or connect represented wrongdoing to a promoted response (e.g., "They knowingly cheated thousands of workers. Be outraged—demand their removal"). Explicit emotion words such as "angry", "outraged", "furious", or "indignant" are not required where the message makes an anger-relevant appraisal and its persuasive use recoverable. Conversely, angry tone, criticism, condemnation, insulting language, hostile expression, or the source's own display of anger does not by itself license the Tactic. The anger or indignation must function as part of the influence relation rather than merely accompanying the message. Annotators do not determine whether the represented offence actually occurred, whether anger is warranted or proportionate, whether the communicator sincerely experiences anger, whether the audience actually becomes angry, or whether the appeal is beneficial, harmful, rational, or effective. The Tactic records the observable use of anger or indignation as persuasive leverage (Walter et al., 2019; Demetriades et al., 2024). Heightened arousal, appraisal of certainty or control, approach motivation, changes in elaboration, faster judgement, heuristic processing, or increased or decreased scrutiny may mediate the effects of anger in particular contexts, but these are treated as possible psychological mechanisms or downstream consequences rather than as fixed TargetConstructs of the Tactic. Research does not support a general assumption that anger necessarily reduces careful processing: under some conditions anger has instead increased systematic or argument-sensitive processing (Moons and Mackie, 2007; Calanchini et al., 2016; Turner et al., 2020). *Target constructs: Belief*—anger or indignation associated with the represented offence, injustice, violation, betrayal, culpability, or comparable provocation is used as a warrant bearing on acceptance, rejection, qualification, or confidence in an identifiable proposition, interpretation, judgement, or position. Anger can influence judgement through emotion-specific appraisal tendencies, including judgements concerning certainty, control, and risk (Lerner and Keltner, 2001); *Attitude*—the anger-relevant appraisal is used to position an identifiable actor, action, institution, policy, event, practice, or other target toward condemnation, disapproval, aversion, opposition, or another unfavourable evaluative stance. Persuasive anger effects on evaluations are context-dependent rather than guaranteed, but evaluative influence is a recognised target of anger-based persuasion (Walter et al., 2019; Calanchini et al., 2016); *Behavior*—anger or indignation may function as motivational leverage for supporting, performing, withhold-

ing, opposing, confronting, punishing, protesting, voting, demanding accountability, mobilising, or otherwise altering identifiable conduct (Skurka, 2019; Turner et al., 2020). *Boundary:* Do not code *Appeal to anger / indignation* merely because a message is hostile, condemnatory, emotionally intense, insulting, or likely to make some audience members angry. Require a recoverable anger-relevant appraisal whose anger or indignation functions as persuasive leverage for a claim, evaluation, judgement, position, or response. Code *Moral framing* when an issue, actor, policy, outcome, or action is construed or evaluated through a moral value, principle, right, obligation, virtue, vice, injustice, betrayal, or prohibition without independently recruiting anger or indignation as persuasive leverage. A message may therefore morally frame a violation without constituting an anger appeal. Co-code only where the moral criterion and the anger-based persuasive relation are independently recoverable. Code *Scapegoating* when blame or responsibility for a broader negative condition, misfortune, or set of harms is concentrated on an identifiable person or group. Blame can contribute to anger, but *Scapegoating* does not automatically license *Appeal to anger / indignation*; require the anger or indignation relation itself to perform persuasive work. Code *Loaded language* when evaluatively or affectively charged lexical choice frames a target through connotation without independently using anger or indignation as a persuasive warrant. Angry or inflammatory wording alone is therefore insufficient for the present Tactic. Code *Affective associative cueing* when the preexisting affective, moral, social, or identity significance of an association is used to alter the relative salience or attentional accessibility of a focal representation. Possible attentional capture caused by anger does not itself license that Tactic; the associative–attentional relation must be independently recoverable. Code *Appeal to fear* when represented danger, threatened harm, or risk functions as a reason for adopting or maintaining an identifiable protective response. Code *Appeal to anger / indignation* when an offence, injustice, violation, betrayal, culpable act, or comparable provocation is instead used through anger or indignation as persuasive leverage for a claim, evaluation, judgement, position, or response. Fear and anger may co-occur psychologically, but co-code the Tactics only where both communicative relations are independently evidenced.

- **Storytelling / narrative packaging.** Use story form—for example, temporally or causally connected events involving actors, situations, and outcomes—as a persuasive vehicle through which claims or evaluations are conveyed (Green and Brock, 2000; Braddock and Dillard, 2016; Green and Appel, 2024). *How to spot:* the persuasive content is organised through an observable sequence of events, actors, situa-

tions, or consequences rather than merely presented as expository assertions; common cues include temporal progression, conflict and resolution, character-centred episodes, and before–after sequences. *Target constructs: Belief*—narrative presentation can support acceptance of story-consistent propositions and interpretations; *Attitude*—narrative presentation can shape favourable or unfavourable evaluation of actors, objects, issues, or outcomes (Green and Brock, 2000; Braddock and Dillard, 2016). Transportation, including attentional focus, imagery, and affect, is treated as a mechanism of narrative influence rather than as a separate fixed TargetConstruct. *Boundary:* Narrative structure alone does not license this Persuasion Tactic. Code *Storytelling / narrative packaging* only when story form functions as a vehicle for a claim or evaluation; narrative structure may otherwise be represented solely in the Narrative dimension.

- **Two-sided with refutation.** Present an opposing or anticipated opposing claim or argument and explicitly rebut it in support of the advocated claim or position or in order to weaken acceptance of that opposing claim if encountered (Allen, 1991; O'Keefe, 1999; Banas and Rains, 2010). *How to spot:* an identifiable opposing or anticipated opposing proposition or argument is acknowledged and accompanied by a substantive counterargument, correction, or rebuttal (e.g., "some argue X; however, ..." or "you may hear the claim that X; here is why that argument fails"). The opposing argument need not already have been encountered by the recipient: preemptive presentation and refutation of an anticipated counterargument also satisfies the Tactic. Mere acknowledgement, warning that persuasion or misinformation may occur, presentation of another side without answering it, or general instruction about manipulation techniques is insufficient unless an identifiable opposing proposition or argument is substantively rebutted. *Target constructs: Belief*—the defining operation is to weaken acceptance of an opposing proposition or argument and strengthen acceptance of the advocated claim, reasons, or conclusion. *Boundary:* Content-based inoculation or "prebunking" is coded as *Two-sided with refutation* where an anticipated opposing proposition or argument is presented and substantively rebutted before possible later exposure (Banas and Rains, 2010). Inoculation is not treated as a separate Tactic solely because the rebuttal is preemptive: prospective timing or an intervention-level objective of increasing resistance to later persuasion does not constitute a distinct message-internal influence operation. Technique-based prebunking, media-literacy instruction, games, simulations, or warnings about general manipulation strategies do not by themselves license *Two-sided with refutation*. Code this Tactic only where an identifiable opposing proposition or argument is actually pre-

sented and rebutted; assign any other Tactics only where their own defining message-internal conditions are independently satisfied. The "threat" component discussed in inoculation theory refers to awareness that an existing position may subsequently be challenged and does not by itself license *Threat framing* in this ontology. *Threat framing* requires an independently recoverable representation of danger, risk, hazard, or potential harm to an identifiable valued target.

- **Preemptive inoculation / prebunking.** Forewarn the recipient about an anticipated future persuasive challenge, misleading claim, influence attempt, or recurrent influence technique and provide counterarguments, refutational material, diagnostic cues, or recognition strategies in advance so as to prepare the recipient to resist, discount, scrutinise, or counter that later influence (Banas and Rains, 2010; Compton et al., 2021; Lewandowsky and Linden, 2021; Roozenbeek et al., 2022b). *How to spot:* the message contains an explicitly recoverable preemptive orientation toward anticipated later influence: it warns that the recipient may encounter a claim, argument, persuasion attempt, misinformation strategy, or recognisable influence technique and supplies material intended to prepare the recipient for that encounter. Observable forms may include forewarning about forthcoming claims followed by advance rebuttal; presenting a weakened or illustrative version of an anticipated argument and explaining how to counter it; or teaching a recognisable persuasive or misleading technique and giving cues for identifying or resisting later instances. For example, "You may soon hear claims that the policy increased the total cost. Those claims usually compare only the first-year expenditure and omit later savings; when you encounter them, check what time period is actually being compared" anticipates a later claim and equips the recipient with a counterargument and diagnostic cue. Technique-based prebunking may instead warn that future messages could use a recurrent influence operation and demonstrate how that operation can be recognised without identifying one specific future proposition. Mere rebuttal, correction, fact-checking, warning, media-literacy advice, or presentation of opposing views does not by itself license the Tactic. The available message must make the preemptive resistance relation recoverable: some later persuasive challenge, misleading message, argument, claim, or influence technique is anticipated, and the current communication supplies material for resisting, scrutinising, discounting, or countering it. The anticipated influence attempt need not actually occur within the available annotation sequence. Conversely, annotators do not infer a prebunking operation merely because information supplied now might prove useful against some hypothetical future message. The prospective challenge or future-resistance function

must be explicitly stated or otherwise recoverable from the communication. Annotators do not determine whether the anticipated claim is actually false, whether the warned-against technique is inherently illegitimate, whether the communicator's preferred position is correct, whether the recipient becomes more resistant to persuasion, or whether the intervention is normatively beneficial. Inoculation and prebunking may be used in support of accurate or inaccurate positions; the Tactic records the observable preemptive-resistance operation rather than adjudicating the merits of the position being protected (Banas and Rains, 2010; Compton et al., 2021). *Target constructs: Belief*—the defining operation may protect, strengthen, or stabilise acceptance, rejection, confidence, or epistemic judgement concerning an identifiable proposition or class of propositions by providing advance counterarguments, refutations, or recognition cues for anticipated later challenges (Banas and Rains, 2010; Lewandowsky and Linden, 2021); *Attitude*—inoculation may preemptively protect or stabilise a favourable or unfavourable evaluative position toward an issue, actor, policy, practice, behaviour, or other target against anticipated counter-attitudinal persuasion; resistance of evaluative attitudes is a foundational application of inoculation theory (Banas and Rains, 2010; Compton et al., 2021); *Agency/Choice*—the defining preemptive operation scaffolds epistemic self-direction by equipping the recipient with counterarguments, diagnostic cues, or recognition strategies for monitoring, evaluating, scrutinising, discounting, or resisting anticipated future influence rather than leaving the recipient wholly dependent on the later message's own presentation (Lewandowsky and Linden, 2021; Roozenbeek et al., 2022b). *Boundary:* Code *Preemptive inoculation / prebunking* only where resistance to an anticipated later influence attempt is itself part of the recoverable communicative operation. The fact that a message contains a strong argument, correction, warning, or useful information does not imply that it is inoculating the recipient against future persuasion. Code *Two-sided with refutation* when an identifiable opposing proposition or argument is acknowledged and substantively rebutted in support of an advocated claim or position, without requiring a prospective resistance function. A conventional inoculation message may instantiate both Tactics where it independently raises and refutes an opposing argument and frames that refutation as preparation for anticipated later persuasion. Technique-based prebunking may instantiate *Preemptive inoculation / prebunking* without *Two-sided with refutation* where the message teaches recognition of a recurrent influence technique rather than rebutting a particular opposing proposition. Code *Doubt-seeding* when uncertainty, disagreement, evidential limitations, possible error, or source-reliability concerns are deployed as grounds for reducing confidence in an identifiable epistemic target. Preemptive in-

oculation may warn that future messages will attempt to create doubt, but such warning does not itself constitute *Doubt-seeding* unless uncertainty or qualification is independently deployed to weaken confidence in a focal claim. Code *Source / channel credibility* when the reputation, institutional provenance, or reliability of an information source functions as the warrant for accepting a claim. Warning that a future source may be unreliable may participate in inoculation where it forms part of an explicit preemptive resistance strategy, but source characterisation alone is insufficient. Where a prebunking message teaches another catalogued Tactic—for example, *Scapegoating*, *False equivalence / distorted comparability*, or *Ad hominem / personal-credibility attack*—do not automatically assign that demonstrated Tactic to the prebunking message merely because it is mentioned or illustrated. Apply the ordinary quoted/reported-material rule: code the demonstrated Tactic additionally only where the message itself deploys that operation as its own influence move rather than presenting it solely as an object of instruction or warning.

- **Doubt-seeding.** Invoke, extend, or generalise uncertainty, disagreement, evidential limitations, source-reliability concerns, or procedural limitations as grounds for reducing confidence in, increasing suspension of judgement toward, or otherwise weakening the represented epistemic warrant of an identifiable claim, conclusion, evidence base, source, consensus, or inference (Michaels, 2006; Stocking and Holstein, 2009; Biddle and Leuschner, 2015; Gustafson and Rice, 2020). *How to spot:* the available message or message sequence establishes an identifiable claim, conclusion, evidence base, source, consensus, or inference and deploys uncertainty, disagreement, limitations, possible error, source-reliability concerns, unresolved questions, or similar epistemic qualifications as grounds for reducing confidence in, suspending judgement toward, or treating that epistemic object as less securely established. Observable forms include using residual disagreement as a reason for caution or suspension of judgement; extending a specific evidential limitation into a broader reason to doubt a conclusion; using unresolved questions as grounds for withholding confidence; invoking possible source or methodological fallibility to weaken reliance on a claim; or stressing the incompleteness of available evidence in order to challenge a stronger conclusion. For example, "The assessment itself labels three of its central judgements low confidence, so we should be much more cautious about treating its conclusion as established" deploys the stated uncertainty to reduce confidence in the conclusion. Similarly, "Eighteen studies support the association and two do not, so the remaining disagreement is a reason not to regard the question as fully settled" uses dissent as a basis for greater epistemic caution. Mere mention or

accurate reporting of uncertainty, disagreement, confidence intervals, limitations, unresolved questions, source characteristics, or a need for further research does not by itself license the Tactic. Those features must be communicatively deployed as reasons for reducing confidence in or increasing suspension of judgement toward an identifiable epistemic target. Annotators identify that observable confidence-reducing operation; they do not determine whether the resulting doubt is scientifically warranted, factually correct, excessive, strategically motivated, sincerely held, beneficial, or harmful (Stocking and Holstein, 2009; Gustafson and Rice, 2020). *Target constructs: Belief*—the defining operation is directed at the represented degree of acceptance or epistemic confidence warranted for an identifiable claim, conclusion, evidence base, source, consensus, inference, or proposition, promoting weaker acceptance, greater qualification, or suspension of judgement rather than necessarily promoting a specific contrary proposition (Biddle and Leuschner, 2015; Gustafson and Rice, 2020). *Boundary:* Do not code *Doubt-seeding* merely because a message acknowledges uncertainty, reports disagreement, identifies limitations, questions evidence, requests additional research, or discusses the reliability of a source or method. The uncertainty or epistemic qualification must itself be deployed as grounds for reducing confidence in, qualifying, or suspending judgement toward an identifiable claim, conclusion, evidence base, source, consensus, or inference. Annotation of *Doubt-seeding* does not imply that the doubt is unwarranted, deceptive, malicious, insincere, or epistemically detrimental: a communicator may sincerely and reasonably believe that greater scepticism or caution is appropriate. The Tactic records the observable influence operation rather than adjudicating its truth, justification, motive, or consequences. Code *False balance / distorted evidential parity* when the defining operation represents recoverably unequal positions or bodies of evidence as comparably supported; code *Doubt-seeding* when uncertainty, disagreement, limitations, or reliability concerns are deployed as reasons to reduce confidence or suspend judgement. Code *Strategic vagueness / equivocation* when the communicator's own proposition, stance, responsibility, quantity, or commitment remains materially underdetermined rather than an epistemic target being presented as warranting greater doubt. Code *Ad hominem / personal-credibility attack* when characteristics or circumstances of an identifiable person are used specifically to discount that person's contribution or standing as a source or critic. Code *Gaslighting / epistemic destabilization* when the defining operation attacks the recipient's confidence in their own perception, memory, reasoning, or judgement rather than presenting an external claim, source, evidence base, or inference as warranting greater doubt. These Tactics may co-occur when their respective operations are independently evi-

denced. Code *Two-sided with refutation* when an identifiable opposing proposition or argument is acknowledged and substantively rebutted in support of an advocated claim or position. Code *Doubt-seeding* when uncertainty, disagreement, evidential limitations, possible error, source-reliability concerns, or procedural limitations instead function as grounds for reducing confidence in or suspending judgement toward an identifiable epistemic target; a specific contrary proposition need not be advanced. Code *Source / channel credibility* when the reputation, reliability, institutional provenance, or publication channel of information functions as a positive warrant for accepting the accompanying claim. A source-reliability concern may contribute to *Doubt-seeding* where it is used to reduce confidence, but mere attribution to a more or less credible source does not automatically license both Tactics. Co-code only where the respective warrant structures are independently recoverable.

- **Presuppositional embedding.** Present an identifiable proposition or relation that is not already established in the available message context as presupposed, taken for granted, or already shared rather than asserting it as the focal proposition, so that the recipient is invited to accommodate the embedded content as part of the informational background (Thoma et al., 2023; Macagno, 2024; Lombardi Vallauri, 2021). *How to spot:* the message contains a linguistic construction whose interpretation carries an identifiable proposition as presupposed or backgrounded content, while the available annotation context does not already establish that proposition as shared or given. Observable forms may include factive constructions (e.g., "the minister finally admitted that the payments were concealed"), change-of-state or iterative expressions (e.g., "the government has stopped hiding the figures", "we must make the country prosperous again"), additive expressions, definite or possessive descriptions, cleft-like constructions, or questions whose formulation takes a proposition for granted (e.g., "when will the minister admit the cover-up?"). The presence of a grammatical presupposition trigger does not by itself license the Tactic. Do not code ordinary presuppositions whose content is already established in the available discourse or whose use merely supports routine reference or discourse cohesion. The defining evidence is that a material proposition not already established in the recoverable context is introduced or maintained as taken-for-granted background rather than as the focal assertion. Annotators identify the presuppositional relation expressed by the message; they do not determine whether the presupposed proposition is factually true, whether the recipient actually accommodates it, or whether the communicator deliberately intended to avoid scrutiny. *Target constructs: Belief*—the defining operation

presents a materially relevant proposition or relation as already given, shared, or settled rather than as a new focal assertion, thereby inviting its accommodation into the represented informational background (Thoma et al., 2023; Macagno, 2024). *Boundary:* Code *Presuppositional embedding* when a materially relevant proposition that is not already established in the recoverable context is encoded as presupposed or taken-for-granted background. Mere implicitness, implication, indirectness, or omission does not by itself license the Tactic. Code *Strategic vagueness / equivocation* when materially relevant content is left underdetermined; *Presuppositional embedding* instead requires a recoverable determinate proposition that is communicated as presupposed rather than focal asserted content. Code *Loaded language* when lexical connotation supplies an evaluative orientation without independently embedding a proposition as presupposed. Code *False dilemma* when a restricted alternative set is represented as exhaustive; a question may instantiate both only where it independently presupposes an unestablished proposition and forecloses additional alternatives. A loaded or complex question is treated as one possible interrogative form of *Presuppositional embedding*, not as a separate Tactic. Other Tactics may co-occur where the presupposed proposition independently performs their defining operations.

- **Repetition.** Reiterate an identifiable claim or proposition, verbatim or in close semantic paraphrase, across an available message or recoverable message sequence so that repeated exposure itself functions as a means of reinforcing acceptance or perceived credibility without supplying substantively new evidential support (Hasher et al., 1977; Fazio et al., 2015; Ye et al., 2026). *How to spot:* the same proposition or substantially equivalent propositional content recurs within the available message or recoverable message sequence, and the recurrence functions as reinforcement rather than as a necessary restatement for quotation, clarification, correction, ordinary discourse cohesion, or the introduction of substantively new evidence. The wording need not be identical: repeated slogans, paraphrases, syntactic variants, or condensed restatements may instantiate the Tactic where the same underlying claim is recurrently presented (e.g., "the plan has failed ... this plan is a failure ... make no mistake, the plan has failed"). Lexical repetition, rhetorical parallelism, anaphora, repeated names or keywords, or a conclusion restated once after an argument do not by themselves license the Tactic. Annotators must identify recurrent propositional content and do not infer that repetition actually increased familiarity, processing fluency, credibility, or belief. The repeated claim need not be false, misleading, maliciously motivated, or disputed. *Target constructs: Belief*—the defining operation repeatedly presents the same proposition or sub-

stantially equivalent propositional content as a means of reinforcing its acceptance, familiarity as a claim, or represented credibility without requiring substantively new evidential support (Hasher et al., 1977; Fazio et al., 2015; Ye et al., 2026). *Boundary:* Do not code *Repetition* merely because words, phrases, names, stylistic structures, or rhetorical formulas recur. Require repetition of an identifiable proposition or substantially equivalent propositional content functioning as reinforcement. Code *Information flooding / overload* when repetition contributes to a cumulative volume of competing material that burdens identification, selection, or scrutiny of focal information; repetition alone does not license flooding. Code *Social proof (bandwagon)* only where the beliefs, endorsements, choices, or conduct of multiple others function as social evidence; recurrence of the same claim across speakers does not by itself establish social proof unless that plurality itself performs the warranting operation. Other Tactics may co-occur where the repeated proposition independently satisfies their defining conditions.

- **Statistical evidence.** Use numerical or statistical information as evidence supporting or challenging a claim or inference (Allen and Preiss, 1997; Hornikx, 2005; Hoeken and Hustinx, 2009). *How to spot:* frequencies, percentages, rates, effect estimates, quantitative comparisons, uncertainty intervals, statistical summaries, or data visualisations are used as evidential support for a proposition; the mere presence of numbers or graphs is insufficient unless they function as evidence. *Target constructs: Belief*—the defining operation is to use statistical information as a warrant bearing on acceptance of a proposition, relation, estimate, or inference. *Boundary:* A numerical attribute does not by itself license *Statistical evidence*. Code *Attribute framing* when the number constitutes the positively or negatively framed description of the focal attribute. Code *Statistical evidence* additionally when the numerical information functions as evidence or a warrant for a separate claim or inference.

- **Source / channel credibility.** Invoke the reputation, reliability, institutional provenance, or publication channel of information as a reason to regard the accompanying claim as credible (Hovland and Weiss, 1951; Metzger et al., 2003; Kiousis, 2001; Sanz-Menéndez and Cruz-Castro, 2019). *How to spot:* a recognised news outlet, official institutional channel, scientific publication venue, research institution, or other identifiable information source is used as a reason to trust the information (e.g., "reported by Reuters", "published in *Nature*", "according to the official ministry website"). Mere attribution of a source is insufficient unless its provenance or reputation functions as a credi-

bility warrant. *Target constructs: Belief*—the defining operation is to use the perceived reliability or provenance of the information source or channel as a warrant supporting acceptance of the accompanying proposition. *Boundary:* Code *Appeal to authority* when the warrant is the expertise, status, or authority of an actor who endorses the claim. Code *Source / channel credibility* when the warrant is the provenance, reliability, or institutional standing of the outlet, publication venue, official channel, or source through which the information is communicated.

- **Associative transfer / guilt-or-honor by association.** Use an identifiable association, affiliation, relationship, alignment, co-membership, endorsement relation, symbolic connection, or comparable link between a focal target and another actor, group, institution, movement, symbol, event, or object as a warrant for attributing to the focal target some of the associate's represented commitments, qualities, credibility, credit, discredit, approval, blame, prestige, stigma, or evaluative standing (Mohammed, 2023; Jowett and O'Donnell, 2018; Moran et al., 2023). *How to spot:* the message establishes or invokes (i) an identifiable focal target, (ii) an identifiable associate or associated category carrying a recoverable commitment, property, credibility, or positive or negative standing, and (iii) a relation in which the association itself functions as a reason for transferring that commitment, property, credibility, or evaluative standing to the focal target. Observable forms may include presenting a person's affiliation with a movement as grounds for attributing that movement's commitments to the person; invoking association with a discredited actor as grounds for distrusting or condemning another actor; or connecting a policy, candidate, organisation, or cause with a positively regarded person, institution, group, or symbol so that the associate's prestige, legitimacy, virtue, or approval functions as a reason for favouring the focal target. For example, "She repeatedly works alongside an organisation that the message describes as openly anti-democratic; that tells you what she really stands for" uses the association as a warrant for attributing commitments and negative standing to the focal actor. Conversely, "This movement stands with the nurses and rescue workers who carried the city through the crisis; that tells you what kind of movement it is" may use positively represented associates as a basis for transferring favourable standing to the movement. Mere co-occurrence, contact, membership, endorsement, appearance in the same photograph, shared event attendance, citation, mention, similarity, or association does not by itself license the Tactic. The association must perform inferential or evaluative work: the recipient must be invited to attribute a commitment, quality, credibility, credit,

discredit, approval, blame, prestige, stigma, or comparable standing to the focal target partly because of that association. Annotators identify the association and its communicated warranting function. They do not determine whether the association is factually accurate beyond what is recoverable from the available message, whether the transferred commitment or evaluation is logically justified, whether the parties genuinely share the attributed beliefs or qualities, or whether the communicator intends to deceive or unfairly discredit the target. Arguments by association may be reasonable or unreasonable depending on context; fallaciousness is not an annotation requirement (Mohammed, 2023). *Target constructs: Belief*—the defining operation may use an association, affiliation, relationship, or alignment as a warrant for attributing an identifiable commitment, belief, quality, credibility judgement, responsibility relation, or other proposition to the focal target on the basis of what is represented about the associated actor or category (Mohammed, 2023); *Attitude*—the defining operation may transfer positive or negative evaluative standing from an associated actor, group, institution, movement, symbol, event, or object to the focal target, positioning that target toward greater or lesser approval, trust, esteem, legitimacy, respectability, desirability, blame, stigma, contempt, or other evaluative regard. Both "guilt" and "honor" by association are included as directional realisations of the same association-based warrant (Mohammed, 2023; Moran et al., 2023). *Boundary:* Code *Associative transfer / guilt-or-honor by association* only when the association itself functions as a warrant for transferring a commitment, property, credibility judgement, or evaluative standing to the focal target. Mere association is insufficient. Code *Affective associative cueing* when an affectively, morally, socially, or identity-relevant cue or association functions instead through differential attentional accessibility or salience, making a focal representation disproportionately prominent or difficult to disregard. *Associative transfer / guilt-or-honor by association* does not require such an attentional asymmetry; it requires the association to support an attributed judgement or evaluation. Co-code only where both the associative–attentional operation and the transfer-of-standing or commitment operation are independently recoverable. Code *Appeal to authority* when an actor's expertise, credentials, professional status, or recognised authority functions as a warrant for accepting a claim, judgement, or recommendation. Code *Associative transfer / guilt-or-honor by association* when the focal target instead borrows or acquires represented commitments or standing because of an association with another actor or entity. An endorsement may instantiate both only where expertise or authority independently warrants the claim and the association itself additionally transfers standing to the focal target.

Code *Source / channel credibility* when the provenance, institutional standing, or reliability of the source or channel through which information is communicated functions as the credibility warrant. Association with a prestigious source or institution does not automatically license both Tactics. Code *Social proof (bandwagon)* when information about the beliefs, approval, choices, or conduct of multiple others functions as social evidence for corresponding judgement or conduct. Associative transfer instead depends on who or what the target is associated with and what standing or commitments are transferred through that relation, not on prevalence or normative consensus. Code *Unity (shared identity)* when the recipient's salient ingroup membership or collective "we-ness" itself functions as the influence warrant. Associative transfer may concern relationships among third parties and does not require the recipient to share the relevant identity. Code *Ad hominem / personal-credibility attack* when a person's character, motives, affiliations, circumstances, or conduct are used specifically to discount that person's identifiable claim, argument, criticism, testimony, question, or standing as a source or critic. Associative transfer may instead target a person, group, policy, institution, movement, idea, or other object without any prior contribution being under scrutiny. Where an association both transfers negative standing and independently functions to discount an identifiable contribution, the Tactics may co-occur. Code *Loaded language* when positive or negative lexical connotation itself supplies the evaluative frame. Evaluatively loaded descriptions of the associate may establish the standing that is subsequently transferred, but they do not by themselves establish associative transfer unless that standing is communicatively carried over to the focal target.

- **Reciprocity.** Invoke a favour, benefit, gift, service, or concession provided to the recipient as a basis for expecting or requesting a corresponding return act (Gouldner, 1960; Regan, 1971; Cialdini and Goldstein, 2004). *How to spot:* the message establishes that a favour, benefit, gift, service, concession, or comparable advantage has been provided to the recipient and makes a reciprocal return relation recoverable by presenting that provision as a basis for what the recipient should now do in return. The mere presence of "free", "complimentary", gift, favour, or concession language is insufficient unless the prior provision functions as a reciprocal warrant for the subsequent response. *Target constructs: Behavior*—the defining operation uses a provided benefit, favour, service, gift, or concession as a reciprocal warrant for eliciting a corresponding return act. *Boundary:* Code *Door-in-the-face (DITF)* when an initially larger request is rejected and followed by a smaller target request presented as a retreat or concession. That rejection–retreat sequence does not by itself additionally license *Reci-*

procity; the reciprocal return relation must independently function as an observable warrant in the message. Code both only where the communication not only performs the DITF sequence but also invokes the concession as something for which a corresponding return is expected.

- **Commitment / consistency.** Invoke a person's prior explicit choice, pledge, promise, or commitment as a reason for corresponding follow-through (Cialdini, 2007; Cialdini et al., 1995). *How to spot:* a prior commitment is explicitly established or recalled and the recipient is encouraged to act consistently with it (e.g., "you pledged to support this", "you already committed to participate"). A merely related prior action is insufficient unless consistency with that commitment is itself invoked. *Target constructs: Behavior*—the defining operation is to use a prior commitment or choice to promote subsequent conduct consistent with that commitment. *Boundary:* Do not infer *Commitment / consistency* merely because a Tactic contains a prior act, agreement, or commitment that may psychologically increase later compliance. Code this Tactic when the prior choice, pledge, promise, or commitment is itself invoked as a reason for corresponding follow-through. Code *Foot-in-the-door (FITD)* when prior compliance with a smaller request is followed by a larger target request, whether or not consistency is explicitly invoked. Code *Low-ball* when an already obtained agreement is followed by worsened terms while continued compliance with the same target decision is sought. Code *Induced hypocrisy / self-inconsistency pressure* when an endorsed standard is juxtaposed with the recipient's own contradictory conduct and that discrepancy supplies the pressure. Co-code only where the prior commitment or choice additionally functions as an independent consistency warrant.

- **Low-ball.** Obtain agreement to a request under initially favourable or incomplete cost conditions and, after commitment has been made, increase the cost or make the terms less favourable while seeking continued compliance (Cialdini et al., 1978; Burger and Petty, 1981; Burger and Caputo, 2015). *How to spot:* evidence that the recipient has already agreed or committed to the target action is followed by the removal of an inducement, disclosure of an omitted cost, or other worsening of the terms, while the recipient is asked or expected to maintain the same decision. The prior commitment may have occurred before the annotated message, but it must be explicitly recoverable from message-internal evidence; an initially attractive offer and unfavourable terms presented together before commitment are insufficient. *Target constructs: Behavior*—the defining operation is to use an already obtained commitment to preserve compliance after the target action has become more costly or less attractive. *Boundary:* Code *Low-ball* when

the recipient has already agreed to the focal course of action and the terms of that same decision subsequently become less favourable while continued compliance is sought. Code *Foot-in-the-door (FITD)* when completed compliance with an initial smaller request is followed by a distinct, larger target request. Code *Door-in-the-face (DITF)* when an initially larger request is rejected and followed by a smaller target request. The defining distinction is therefore worsening terms after commitment to the same target decision (*Low-ball*), escalation from completed smaller compliance to a larger request (*FITD*), or retreat from a rejected larger request to a smaller one (*DITF*).

- **Unity (shared identity).** Invoke the recipient's salient ingroup identity, shared social membership, or sense of collective "we-ness" as a basis for making a claim, evaluation, or requested course of action more persuasive (Cialdini, 2016; Spears, 2021). *How to spot:* shared group membership, collective identity, kinship, team membership, nationality, profession, political affiliation, religion, shared mission, or comparable "one of us" language is used as a basis for accepting a claim, adopting an evaluation, or acting accordingly. Mere mention of a social category, demographic characteristic, or collective pronoun is insufficient unless shared identity or belonging performs the influence operation. Patriotic or "flag-waving" appeals are included here where attachment to a nation, polity, movement, or other salient ingroup functions as the identity-based warrant; patriotic symbols, national references, or group language alone are insufficient unless that identity relation performs the influence operation. *Target constructs: Identity*—the defining operation foregrounds or invokes a shared self-category or group identification as part of the influence attempt; *Belief*—shared identity can function as a basis for accepting a claim, judgement, or group-consistent interpretation; *Attitude*—shared identity can function as a basis for adopting a favourable or unfavourable evaluation aligned with the group; *Behavior*—shared identity can function as a basis for eliciting group-consistent, cooperative, supportive, or compliant conduct.

- **Scarcity appeal.** Present an opportunity, product, resource, or course of action as limited in quantity, time, supply, or continuing availability in order to increase its perceived value or motivate action before the opportunity is lost (Lynn, 1991; Worchel et al., 1975; Barton et al., 2022). *How to spot:* message-internal claims of limited quantity, limited time, restricted supply, unusually high demand, or impending unavailability (e.g., "only a few left", "available today only", "last chance") are presented as a reason to value or act on the focal opportunity. Mere mention of rarity or shortage is insufficient unless limited availability func-

tions as an influence warrant. *Target constructs: Attitude*—scarcity is used to increase the perceived value, desirability, or attractiveness of the focal object or opportunity; *Agency/Choice*—scarcity represents the focal option as becoming less available or potentially unavailable, thereby contracting the perceived action space; *Behavior*—scarcity is used to accelerate or elicit action before the opportunity is lost (Lynn, 1991; Barton et al., 2022).

- **Altercasting.** Assign, project, or offer the recipient a social or relational role, category, or identity position carrying an associated behavioural expectation, so that a desired response is presented as congruent with the role or identity position the recipient is invited to assume (Weinstein and Deutschberger, 1963; Pratkanis, 2000; Miller et al., 1975; Turner et al., 2010). *How to spot:* the recipient is characterised as, or invited to become, a particular role-holder, category member, or socially recognisable identity position with an associated expectation for how such a person acts (e.g., "as a responsible parent, you protect your children", "be a voter—make your voice count", "as team captain, you need to set the example") in a way that makes a corresponding response identity-consistent. Mere description of the recipient's demographic characteristics or existing role is insufficient unless that identity assignment performs the influence operation. *Target constructs: Identity*—the defining operation assigns, projects, or offers a self-relevant quality or social role for the recipient to assume; *Behavior*—the projected identity is used to promote conduct consistent with the expectations attached to that role. *Boundary:* Code *Unity (shared identity)* when shared membership in a collective "we" is presented as the basis for influence. Code *Altercasting* when the recipient is assigned, characterised as, or invited to assume a social or relational role, category, or identity position whose associated behavioural expectations provide the influence warrant. Code *Pride / ego appeals* when a positively valued trait, competence, status, or self-image instead functions as leverage for believing, evaluating, choosing, or acting in the promoted way. Co-code only where an independent role-assignment operation and an independent positive-self-image appeal are both recoverable. The two may co-occur when both shared belonging and recipient-directed identity assignment are independently present.

B.5 Framing

- **Loaded language.** Use evaluatively or affectively charged lexical choices to make a target, action, event, or state of affairs appear more favourable, unfavourable, acceptable, threatening, admirable,

contemptible, or otherwise evaluatively marked without requiring a new substantive proposition (Da San Martino et al., 2020; Allan and Burridge, 1991; Capraro and Vanzo, 2019). *How to spot:* words or phrases carry a marked positive or negative connotation that frames how the referent should be regarded, including intensifying, dysphemistic, euphemistic, or otherwise evaluatively loaded wording. The relevant cue is the lexical framing itself rather than merely strong sentiment: a neutral or less evaluatively marked formulation could refer to substantially the same object, action, or event. Mere descriptive precision, ordinary emotional expression, or factual bad news is insufficient unless the lexical choice itself supplies the evaluative frame. *Target constructs: Attitude*—the positive or negative connotations of the lexical choice position the focal target, action, event, or state of affairs toward a more favourable or unfavourable evaluation (Capraro and Vanzo, 2019).*Boundary:* Code *Loaded language* when evaluatively charged wording frames an object, action, event, condition, or actor through lexical connotation. Code *Name-calling / labeling* when a condensed evaluative label is applied to a person or group as a characterisation of who or what they are. The two may co-occur when both operations are independently present. So-called "glittering generalities" or generic positive-value language are coded as *Loaded language* where their influence operation consists in positively evaluative lexical connotation. Where an identifiable moral value, principle, right, obligation, or prohibition instead functions as a substantive criterion for interpreting or evaluating the target, code *Moral framing*; where another specific warrant such as tradition, shared identity, novelty, or authority performs the influence operation, code the corresponding Tactic rather than a generic appeal-to-values label.

- **Name-calling / labeling.** Apply a condensed evaluative label, epithet, or categorical characterisation to a person, group, idea, institution, or other target in order to frame that target as desirable or undesirable, admirable or contemptible, acceptable or threatening (Da San Martino et al., 2019; Jowett and O'Donnell, 2018; Carnaghi et al., 2008). *How to spot:* a person, group, idea, institution, or other target is characterised through a compact evaluative label or epithet (e.g., "traitor", "extremist", "parasite", "hero", "savior") that functions as a characterisation of what or who the target is. Mere mention of a social, political, occupational, or demographic category is insufficient unless the label itself carries an evaluative framing function. *Target constructs: Attitude*—the evaluatively marked label, epithet, or categorical characterisation positions the target toward approval or disapproval, attraction or aversion, praise or condemnation (Da San Martino et al., 2019; Bianchi et al., 2019). *Boundary:* Code *Name-calling / la-*

beling when a condensed evaluative label, epithet, or categorical characterisation is applied to a target as a characterisation of who or what that target is. Code *Loaded language* when evaluatively or affectively charged wording frames a referent through lexical connotation without independently assigning such a condensed characterisation. Code *Altercasting* when a role, quality, or self-description is assigned or offered specifically to the recipient so that a desired response becomes congruent with the identity the recipient is invited to assume. Co-code only where the respective operations are independently recoverable rather than treating every evaluative label as automatically instantiating multiple Tactics.

- **Ridicule / mockery.** Render an identifiable person, group, institution, idea, policy, action, claim, performance, or other target laughable, foolish, absurd, pretentious, incompetent, contemptible, or otherwise worthy of derision through comic incongruity, mocking imitation, parody, caricature, sarcastic re-presentation, ironic quotation, teasing, or another recognizably ridiculing communicative operation (D'Errico and Poggi, 2016; Wang and Silva, 2018; Zappavigna, 2022; Verhulsdonk et al., 2022). *How to spot:* the message represents an identifiable target through a recognizably mocking or ridiculing relation in which comic incongruity, distortion, imitation, irony, parody, caricature, sarcastic praise, ironic quotation, or a comparable device positions the target as ridiculous, foolish, absurd, pompous, inept, unserious, contemptible, or otherwise deserving of laughter or derision. Observable forms may include exaggerated imitation of a target's manner or conduct; ironic repetition of the target's own words so as to expose them to mockery; comic recasting of a policy, argument, or actor as incompetent or absurd; or sarcastic praise whose recoverable evaluative force is derisive rather than favourable. For example, after a sequence of policy reversals, "another masterclass in decisive leadership" may instantiate the Tactic where the ironic praise communicatively renders the target's conduct ridiculous. Likewise, a parody that exaggerates a politician's self-importance so as to make that self-presentation appear absurd may instantiate the Tactic even without a separate insult or explicit negative adjective. Humor, laughter, irony, sarcasm, parody, exaggeration, teasing, imitation, or comic imagery does not by itself license the Tactic. Require a recoverable target-directed operation of ridicule or mockery that evaluatively positions the target as laughable, absurd, foolish, contemptible, or otherwise diminished. Benign joking, affiliative teasing, self-deprecating humour, playful imitation without target devaluation, and humour whose function is not to render an identifiable target ridiculous are insufficient. Annotators identify the observable ridiculing relation and do not determine whether the tar-

get actually deserves ridicule, whether the audience finds the message funny, whether laughter occurs, whether the source consciously intends humiliation, or whether the ridicule successfully damages the target's reputation or persuasive standing (D'Errico and Poggi, 2016; Verhulsdonk et al., 2022). *Target constructs: Attitude*—the defining operation positions an identifiable target toward diminished regard, approval, esteem, seriousness, competence, respectability, or favourable evaluation by communicatively rendering that target laughable, foolish, absurd, pretentious, contemptible, or otherwise worthy of mockery or derision (D'Errico and Poggi, 2016; Zappavigna, 2022). *Boundary:* Code *Ridicule / mockery* when comic, ironic, parodic, imitative, caricatural, or otherwise mocking representation itself performs a target-directed negative evaluative operation. Do not code the Tactic merely because a message is humorous or because some audience could find the target amusing. Code *Name-calling / labeling* when a condensed evaluative label, epithet, or categorical characterisation positions the target. Ridicule does not require a label, and a label does not require ridicule. Co-code where a mocking epithet independently functions both as a categorical label and as part of a recoverable ridiculing operation. Code *Loaded language* when evaluatively or affectively charged lexical choice supplies the framing operation through connotation. *Ridicule / mockery* instead requires a recoverable relation in which the target is communicatively made laughable, absurd, foolish, or worthy of derision; the two may co-occur where both operations are independently evidenced. Code *Humor as deflection* when humour displaces substantive engagement with an identifiable prior question, criticism, accusation, challenge, or point of scrutiny. Ridicule may directly attack or evaluate a target without deflecting from anything. Co-code only where the same communication independently both renders a target ridiculous and uses the humorous turn to displace substantive engagement with recoverable scrutiny. Code *Ad hominem / personal-credibility attack* when a negative characterisation of a person is used as a reason to discount that person's claim, argument, criticism, testimony, question, or standing as a source. Ridicule of a person does not automatically constitute an ad hominem attack, and ridicule may target policies, institutions, ideas, actions, or other non-personal objects. Code *Exaggeration / overstatement* when a scalar, quantitative, evaluative, causal, evidential, or other represented degree is inflated relative to a recoverable baseline. Caricature and parody may use exaggeration as a device, but comic distortion does not automatically license *Exaggeration / overstatement* unless the independent recoverable inflation relation is satisfied. Parody, satire, sarcasm, caricature, and ironic quotation are treated as possible forms through which *Ridicule / mockery* may be realised, not as independent Tactics solely by virtue of their form.

- **False dilemma.** Present two positions, outcomes, or courses of action as an exhaustive choice in a way that forecloses intermediate, combined, neutral, or additional alternatives (Macagno, 2024; Da San Martino et al., 2020). *How to spot:* an either–or construction, "with us or against us", binary question, or equivalent formulation presents two alternatives as exhausting the relevant possibility set. The defining cue is the foreclosure of other possibilities, not merely the presence of two alternatives. Do not code genuinely complementary alternatives (e.g., A versus not-A) or an explicitly binary choice unless the message additionally represents intermediate, combined, neutral, or other possibilities as unavailable. If determining whether additional alternatives exist would require external factual information not recoverable from the message, annotators should not infer them. *Target constructs: Belief*—the defining operation presents an artificially restricted alternative set as an exhaustive representation of the relevant possibilities; *Agency/Choice*—the same foreclosure represents available positions or courses of action as restricted to the presented alternatives, narrowing the perceived option set (Macagno, 2024; Da San Martino et al., 2020). *Boundary:* Code *False dilemma* when discourse represents a restricted set of alternatives as exhausting the relevant possibility set despite message-internal room for additional, intermediate, combined, or neutral possibilities. Code *Scarcity appeal* when the constraint instead concerns the continuing availability, quantity, time, supply, or accessibility of an opportunity itself. Code *Inevitability framing* when one trajectory, development, outcome, or course is represented as necessary, unavoidable, or effectively without viable alternatives rather than two presented alternatives being treated as exhaustive. These Tactics may co-occur only where their respective structures are independently recoverable.

- **Causal oversimplification.** Represent a complex outcome, condition, or problem through a single cause or reason while treating that cause as sufficient, overriding, or exhaustive of the relevant causal explanation (Da San Martino et al., 2020; Dimitrov et al., 2021). *How to spot:* the message presents one factor as the sole, decisive, or comprehensive explanation for an outcome (e.g., "the only reason", "all because of X", "everything happened because of X"). Mere causal attribution ("X contributed to Y" or "X caused Y") is insufficient unless the message itself represents the causal account as exhaustive or suppresses the possibility of contributing causes. Annotators should not infer omitted causes from external factual knowledge. *Target constructs: Belief*—the defining operation is to promote acceptance of a restricted causal model in which one factor is represented as sufficient, overriding, or exhaustive of the relevant explanation.

- **Slippery-slope / consequential escalation.** Represent an initial action, policy, concession, decision, event, or condition through a projected downstream progression in which it is presented as setting in motion, enabling, normalising, or materially increasing the likelihood of one or more further developments that culminate in a substantially more desirable or undesirable outcome, so that the initial step is construed and evaluated through the projected trajectory that follows from it (Walton, 1992; Walton, 2015; Haigh et al., 2016; Piskorski et al., 2023a). *How to spot:* the message establishes an identifiable initial step or condition and represents it as leading beyond its immediate consequences toward a later development, escalation, precedent, cascade, or endpoint. Observable forms may contain linked progression markers such as "then", "next", "soon", "eventually", "once we allow this", "it will not stop there", "this opens the door to", or equivalent formulations indicating that an initial step is projected forward into further consequences (e.g., "If we accept this exception, more exceptions will follow, the rule will lose its force, and eventually there will be no meaningful restriction at all"). A fully explicit multi-step chain is not required where the message itself clearly encodes an escalatory or precedent-based progression from the initial step to a materially more consequential downstream endpoint. Conversely, a simple statement that one action will have one direct consequence does not by itself license the Tactic. The defining evidence is the projected progression beyond an ordinary immediate consequence. The projected endpoint may be undesirable or desirable. Negative slippery-slope arguments commonly use an escalating adverse consequence to frame an initial proposal unfavourably, while structurally analogous positive projections may present an initial action as beginning a cascade toward a highly desirable outcome. These are treated as directional realisations of the same Tactic rather than as separate Tactics (Piskorski et al., 2023a). Annotators identify the consequential progression communicated by the message and do not determine whether every causal or enabling link is factually correct, sufficiently evidenced, probable, exaggerated, logically valid, or ultimately likely to occur. A slippery-slope argument need not be fallacious, deceptive, irrational, or maliciously motivated (Walton, 1992; Walton, 2015). *Target constructs: Belief*—the defining operation represents an initial action, decision, policy, event, or condition as standing in a projected consequential relation to one or more later developments, thereby promoting acceptance of a downstream trajectory, escalation, precedent, cascade, or endpoint as following from or being enabled by the initial step (Walton, 2015; Haigh et al., 2016); *Attitude*—the initial action, decision, policy, event, or condition is positioned toward favourable or unfavourable evaluation through the desirable or undesirable downstream trajectory represented as follow-

ing from it. The evaluative direction may therefore be negative, as in a conventional slippery-slope warning, or positive, as in an analogous escalating-benefit or "stairway to heaven" projection. *Boundary:* Code *Slippery-slope / consequential escalation* when an initial step is construed through a projected downstream progression toward a materially more consequential endpoint. A direct consequence alone is insufficient unless the message additionally encodes escalation, precedent, progressive extension, or another continuing trajectory. Code *Causal oversimplification* when the defining operation instead represents one factor as the sufficient, overriding, or exhaustive explanation for an outcome. *Slippery-slope / consequential escalation* does not require a claim that the initial step is the sole cause of the endpoint; the distinctive operation is the projected progression. Co-code only where both causal exhaustivity and consequential progression are independently recoverable. Code *Inevitability framing* when a development, trajectory, outcome, or course of action is represented as unavoidable, predetermined, necessary, or effectively without viable alternatives. A slippery-slope projection may be uncertain or probabilistic and therefore does not automatically instantiate inevitability. Co-code where the message independently represents both a downstream progression and that progression as effectively unstoppable or unavoidable. Code *Threat framing* when a person, group, event, condition, or development is represented as posing danger or potential harm without the progressive consequential structure required here. Code *Appeal to fear* when a represented danger functions as a reason for an identifiable protective response. A harmful endpoint may participate in all of these relations, but co-code only where each defining operation is independently recoverable. Code *Gain/loss goal framing* when desirable or undesirable consequences of performing or not performing an identifiable focal behaviour are used to promote that same behavioural goal without an independent downstream escalation or cascade. Code *False dilemma* when the operation instead represents a restricted alternative set as exhaustive.

- **Scapegoating.** Concentrate blame or responsibility for a broader negative outcome, condition, or misfortune on an identifiable person or group, constructing that target as the principal culprit or source of the problem (Rothschild et al., 2012; Glick, 2002). *How to spot:* a person or group is assigned concentrated responsibility for a broader negative condition or set of harms (e.g., "they caused all our problems", "everything went wrong because of them"). Attribution of a specific, bounded action or outcome to an identified actor is insufficient unless the message generalises or concentrates broader culpability on that target. Annotation records the blame structure of the message

and does not require determining whether the target is factually innocent or whether the attribution is objectively unjustified. *Target constructs: Belief*—the target is represented as principally responsible for the negative condition or outcome; *Attitude*—the concentration of blame positions the target as culpable or blameworthy and thereby toward unfavourable evaluation (Rothschild et al., 2012). *Boundary:* Code *Causal oversimplification* when the defining operation is an artificially restricted causal explanation, whether or not the proposed cause is a person or group. Code *Scapegoating* when the defining operation is concentrated blame directed at an identifiable person or group. Code both when a person or group is simultaneously presented as the exhaustive cause of a broader problem and as the principal object of blame.

- **Moral framing.** Present an issue, actor, policy, outcome, or course of action through a moral value, principle, virtue, vice, obligation, right, or prohibition so that the target is construed and evaluated through a moral criterion (Feinberg and Willer, 2015; Kodapanakkal et al., 2022; Mobayed and Sanders, 2022). *How to spot:* the message invokes a recognisable moral criterion—for example, care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, purity/degradation, liberty/oppression, justice, rights, duty, virtue, vice, sin, or moral obligation—as a basis for interpreting or evaluating an issue, actor, policy, outcome, or action. Mere positive or negative wording, emotional intensity, or use of an isolated morally connoted word is insufficient unless a moral criterion performs the framing operation. *Target constructs: Belief*—the defining operation construes an issue, actor, action, or outcome through a moral relation or category, shaping what the situation is understood to represent (e.g., harm, injustice, betrayal, subversion, degradation, or oppression); *Attitude*—the same moral criterion positions the target toward moral approval or disapproval, for example as right or wrong, just or unjust, virtuous or vicious, permissible or impermissible, worthy or blameworthy. Existing moral values or moral identities may provide the basis through which the frame is persuasive, but their use as leverage does not by itself yield an *Identity* mapping (Feinberg and Willer, 2015; Kodapanakkal et al., 2022; Mobayed and Sanders, 2022).

- **Metaphor / analogy.** Map relational structure from a source or comparison domain onto a target issue, actor, event, process, or object so that the target is interpreted or reasoned about through that imported structure (Gentner, 1983; Bowdle and Gentner, 2005; Thibodeau and Boroditsky, 2011). *How to spot:* the message explicitly or implicitly represents a target through another domain or compar-

ison structure—for example, "crime is a virus", "the economy is an engine", "this organisation is like a sinking ship", or sustained war, disease, machine, journey, or comparable mappings—and properties or relations of the source domain organise how the target is understood. Explicit comparison markers such as "like", "as", or "analogous to" may signal analogy but are not required. Mere figurative wording, idiom, decorative imagery, or lexical metaphor is insufficient unless the cross-domain mapping performs an interpretive or inferential framing function. *Target constructs: Belief*—the defining operation imports relational structure from a source or comparison domain into the representation of the target, thereby shaping its construal and the relations, causes, consequences, or solutions that are inferred from it (Gentner, 1983; Thibodeau and Boroditsky, 2011).

- **Dehumanizing representation.** Represent an identifiable person, population, or social group as less than fully human, outside the category of ordinary humanity, lacking characteristically human attributes, or more appropriately understood through an animalistic, object-like, mechanistic, contaminating, parasitic, biological, or otherwise nonhuman mode of being (Haslam, 2006; Kteily et al., 2015; Enock and Over, 2023). *How to spot:* the message contains recoverable representational content that denies, diminishes, displaces, or materially qualifies the target's human status or human characteristics, or maps the target onto a nonhuman category in a way that organises how the person or group is to be understood. Observable forms may include explicit denial of humanity (e.g., "they are not human" or "they are barely human"); animalistic characterisation (e.g., people represented as vermin, parasites, livestock, predators, or other animals); mechanistic or objectifying representation that presents people as machines, objects, interchangeable units, or automata; or biological and contamination representations that construe populations as disease, infection, infestation, pollutants, or comparable nonhuman threats (Haslam, 2006; Valtorta et al., 2026). A negative description, insult, comparison, animal term, object term, disease term, or metaphor does not by itself license the Tactic. The nonhuman representation must characterise how an identifiable human target is construed rather than merely provide decorative imagery, describe an external threat, or express hostility. Annotators identify the observable human–nonhuman representational relation and do not determine whether the recipient actually regards the target as biologically nonhuman, attributes fewer uniquely human psychological traits to the target, experiences reduced empathy, endorses exclusion or violence, or otherwise exhibits psychological dehumanisation. Annotation does not imply that the communicator consciously intends to deny the target's humanity or that a figurative

expression is believed literally. Research indicates that dehumanising language may affect evaluations and harm-related responses through mechanisms other than literal denial of human traits; these possible mechanisms and downstream consequences are not part of the Tactic's licensing rule (Enock and Over, 2023). *Target constructs: Belief*—the defining operation alters the represented categorical, ontological, or attribute structure of an identifiable human target by construing that target as less than fully human, as lacking relevant human characteristics, or through a nonhuman animalistic, mechanistic, objectifying, biological, contaminating, or comparable representational category (Haslam, 2006; Kteily et al., 2015); *Attitude*—the same human–nonhuman representation positions the target toward diminished regard, esteem, social worth, moral standing, or favourable evaluation by construing the person or group as degraded, subhuman, object-like, contaminating, parasitic, animal-like, or otherwise outside or below ordinary human regard. The Tactic does not require extreme hatred, prejudice, or conflict, and annotators do not infer that the audience actually adopts a more negative attitude (Haslam, 2006; Kteily et al., 2015). Code *Name-calling / labeling* when a condensed evaluative label, epithet, or categorical characterisation positions a target toward approval or disapproval. A term such as "vermin" or "parasites" may instantiate both Tactics where it independently functions both as an evaluative label and as a nonhuman characterisation; do not infer *Dehumanizing representation* from every insult or *Name-calling / labeling* from every human–nonhuman representation. Code *Metaphor / analogy* when relational structure from a source or comparison domain is mapped onto the target so as to organise interpretation or inference. A dehumanising metaphor may satisfy both Tactics where the cross-domain mapping independently performs an inferential framing function and the target is additionally construed through a materially nonhuman representation. Neither Tactic automatically licenses the other. Code *Loaded language* when affectively or evaluatively charged lexical choice positions a target without independently representing that target as less than fully human or through a materially nonhuman category. Code *Scapegoating* when concentrated responsibility for a broader negative condition or set of harms is assigned to an identifiable person or group. Dehumanization may accompany blame or hostility, but neither operation is required for the other. Code *Threat framing* where a person or group is represented as a danger, hazard, or source of potential harm to an identifiable target. Disease, infestation, predator, or contamination imagery may instantiate both only where the communication independently represents the human target through a nonhuman category and establishes a threat relation. Dehumanizing representation does not require an intergroup or polarising relation. The human tar-

get may be an individual, a set of individuals, a social group, population, institutionally defined category, or other identifiable human target. Do not infer a Polarization annotation merely because the dehumanised target is a group or socially meaningful category. Where the same communication additionally constructs, reinforces, or intensifies an antagonistic relation between social groups, for example by contrasting a dehumanised outgroup with an ingroup, representing the groups as fundamentally opposed, or using the nonhuman characterisation as part of an "us"–"them" division, the applicable Polarization motif is annotated where its additional group-relational conditions are satisfied. The *Dehumanizing representation* Tactic records the human–nonhuman representational operation itself rather than whether that operation is intergroup-polarizing.

- **Collective nostalgic framing.** Invoke a positively valued shared ingroup past, often in contrast with a diminished, disrupted, or changed present, as an interpretive and evaluative frame for current conditions or possible futures (Boym, 2001; Smeekes et al., 2021; Schreurs, 2021; Wildschut et al., 2014). *How to spot:* the message evokes a positively remembered or idealised past associated with a salient collective—for example, "our good old days", "when we stood together", "the country we once were", or a shared "golden age"—and uses that past to frame the present in terms of continuity, loss, decline, or possible restoration. Mere historical reference, commemoration, retro aesthetics, or personal reminiscence is insufficient unless a valued shared group past performs the framing operation. *Target constructs: Belief*—the defining temporal frame construes relations between a valued collective past and the present, for example as continuity, loss, decline, or departure from an earlier condition; *Attitude*—the nostalgic comparison positively evaluates the invoked past and uses it as an evaluative reference point for present actors, conditions, institutions, or possible futures; *Identity*—the invocation of a shared ingroup past foregrounds collective continuity, belonging, or self-definition across time (Smeekes et al., 2021; Wildschut et al., 2014). *Boundary:* Code *Collective nostalgic framing* when a positively valued shared ingroup past functions as a temporal reference frame through which the present or future is construed or evaluated. Code *Unity (shared identity)* when shared collective membership or "we-ness" itself functions as a warrant for accepting a claim, adopting an evaluation, or acting accordingly. The mere presence of a shared group in a nostalgic frame does not automatically license *Unity*; co-code only where shared membership independently performs that persuasive function.

- **Inevitability framing.** Represent a future development, institutional

change, policy trajectory, technological transition, or other outcome as unavoidable, predetermined, necessary, or effectively without viable alternatives (Leonardi, 2008; Séville, 2017). *How to spot:* the message characterises a development or course as "inevitable", "unstoppable", "the only way forward", "there is no alternative", "no going back", or otherwise portrays competing trajectories as futile, obsolete, unrealistic, or unavailable. Mere prediction, confidence, or advocacy is insufficient unless necessity or unavoidability performs the framing operation. *Target constructs: Belief*—the defining operation represents a contingent future, trajectory, or outcome as necessary, predetermined, or unavoidable; *Agency/Choice*—the same necessity frame represents alternative courses, trajectories, or responses as unavailable, non-viable, or futile, thereby narrowing the perceived option set (Leonardi, 2008; Séville, 2017).

- **Threat framing.** Represent a person, group, event, condition, development, or other phenomenon as a danger, risk, hazard, or threat to a valued object, population, interest, institution, condition, or way of life, thereby organising interpretation around potential harm (Eriksson, 2020). *How to spot:* the message characterises a focal phenomenon as threatening, dangerous, hazardous, endangering, destructive, or capable of causing serious loss or harm to an identifiable target (e.g., "threatens our livelihoods", "puts democracy at risk", "could devastate the community", "endangers our way of life"). Mere mention of an unpleasant event, negative outcome, fear-related word, or factual hazard is insufficient unless the danger or threat relation performs the framing operation. Annotators identify the communicated threat relation and do not determine whether the danger is objectively warranted, exaggerated, or likely to occur. *Target constructs: Belief*—the defining operation represents a person, group, event, condition, development, or other phenomenon as standing in a relation of danger, risk, hazard, or potential harm to an identifiable object, population, interest, institution, condition, or way of life (Eriksson, 2020). *Boundary:* Code *Threat framing* when the defining operation is to represent a person, group, event, condition, development, or other phenomenon as standing in a danger, risk, hazard, or potential-harm relation to an identifiable target. Code *Appeal to fear* when represented danger or threatened harm functions as a reason for adopting, maintaining, or performing an identifiable protective response. Where the threat representation functions only as the threat premise required to motivate that protective response, do not automatically add *Threat framing*. Code both only where the threat characterisation independently organises interpretation of the focal phenomenon in addition to supplying the threat–protective-response relation. Code an applicable Coercion Tactic when

a prospective adverse consequence is communicatively deployed as contingent leverage over identifiable target conduct. Co-code only where the respective operations are independently evidenced.

- **Gain/loss goal framing.** Frame a focal recommended behaviour in terms of the desirable consequences associated with performing it or the undesirable consequences associated with failing to perform it, while retaining the same behavioural goal (Levin et al., 1998; Rothman and Salovey, 1997; Meyerowitz and Chaiken, 1987; O'Keefe and Jensen, 2007). *How to spot:* an identifiable focal act is promoted by emphasising either gains, benefits, desirable outcomes, or avoided harms associated with performing the act, or losses, costs, undesirable outcomes, or foregone benefits associated with not performing it (e.g., "if you do X, you can gain or preserve Y"; "if you do not do X, you risk losing or foregoing Y"). The defining cue is the valenced representation of consequences around the same behavioural goal. Positive or negative wording, a favourable or unfavourable attribute, or a gain/loss statement without an identifiable focal behaviour is insufficient. *Target constructs: Behavior*—the defining operation uses positively or negatively framed consequences of compliance or noncompliance to promote performance of the focal behaviour (Levin et al., 1998; Rothman and Salovey, 1997; Meyerowitz and Chaiken, 1987; O'Keefe and Jensen, 2007). *Boundary:* Code *Gain/loss goal framing* when the defining operation is the valenced representation of consequences associated with performing or not performing the same focal behaviour. Code *Appeal to fear* when a represented threat, danger, or harmful consequence functions specifically as a reason for adopting or maintaining an identifiable protective response. The presence of an undesirable consequence in a fear appeal does not by itself additionally license *Gain/loss goal framing*; co-code only where the communication independently organises the behavioural recommendation through a recoverable gain/loss consequence frame beyond the threat–protective-response relation. Code *Attribute framing* when the valenced representation concerns complementary descriptions of an attribute rather than consequences of performing or not performing a focal behaviour.

- **Risky-choice framing.** Present the outcomes of a decision under risk in gain versus loss terms relative to a reference point in a way that structures selection among available choice alternatives (Tversky and Kahneman, 1981; Levin et al., 1998; Kühberger, 1998). *How to spot:* the message presents an identifiable choice among alternatives involving uncertain or probabilistic outcomes and represents the relevant outcomes as gains, preservation, or positive departures from a reference point, or as losses, deterioration, or negative departures from

that reference point. Canonical forms contrast a certain outcome with a probabilistic outcome while expressing their consequences in gain or loss terms. Mere mention of risk, probability, gains, losses, or multiple options is insufficient unless the gain/loss representation structures the risky choice. Annotators identify the frame expressed in the message and do not need to observe an alternative experimental wording or an actual audience preference reversal. *Target constructs: Behavior*—the defining operation uses gain/loss representation of outcomes under risk to structure selection among the available choice alternatives (Tversky and Kahneman, 1981; Levin et al., 1998; Kühberger, 1998). *Boundary:* Code *False dilemma* when two alternatives are presented as exhausting the relevant possibility set despite message-internal room for additional alternatives. Code *Risky-choice framing* when available alternatives involving different levels of certainty or risk are represented through gain/loss outcomes. The presence of two options alone does not license either Tactic.

- **Attribute framing.** Present a focal object, event, outcome, or option through one positively or negatively valenced pole of a semantically or quantitatively complementary attribute description, thereby shaping evaluation while the underlying attribute remains equivalent (Levin and Gaeth, 1988; Levin et al., 1998). *How to spot:* the message characterises the same underlying attribute through a positive or negative complementary formulation—for example, "75% lean" versus "25% fat", "90% survival" versus "10% mortality", or "80% success" versus "20% failure". The defining cue is not merely positive or negative wording, but selection of one valenced pole from complementary descriptions of the same attribute. The complementary relation must be directly recoverable from the message or from the represented quantity itself; annotators should not invent an alternative frame where the categories are not demonstrably exhaustive or complementary. *Target constructs: Perception/Attention*—the defining operation selects one valenced pole from a directly recoverable pair of complementary descriptions, giving the selected representation greater immediate accessibility than its recoverable complement; *Attitude*—the valenced representation of the otherwise equivalent attribute positions the focal object, event, outcome, or option toward a more favourable or unfavourable evaluation (Levin and Gaeth, 1988; Levin et al., 1998).

B.6 Manipulation

- **Guilt trip.** Assign or invoke guilt-relevant responsibility, obligation, indebtedness, inequity, transgression, or failure relative to an interpersonal, moral, or group-based standard in order to pressure the recipient

toward corrective, reparative, compliant, prosocial, inhibitory, or otherwise altered conduct (Baumeister et al., 1994; Tangney et al., 2007; Boster et al., 2016). *How to spot:* the message represents the recipient as responsible, indebted, blameworthy, implicated, or obligated because of their own conduct or omission, another person's sacrifice or suffering, an inequitable benefit, or the actions or history of a group with which the recipient is associated, and uses that guilt-relevant appraisal as pressure around what the recipient should do, stop doing, repair, compensate for, or otherwise change (e.g., "after everything I have done for you", "you'll let everyone down", "how can you just walk away after what happened?"). The recipient need not have personally caused the original harm: responsibility may be represented as interpersonal, vicarious, or collective. An explicit command is not required where the desired correction or change in conduct is recoverable from the message. Mere discussion of guilt, historical wrongdoing, responsibility, inequality, or moral obligation is insufficient unless guilt-relevant responsibility or evaluation is used as pressure on the recipient's conduct. *Target constructs: Belief*—the defining operation promotes acceptance of a responsibility, obligation, indebtedness, implication, or accountability relation connecting the recipient or relevant ingroup to a harm, transgression, inequity, omission, or unmet standard; *Attitude*—the relevant conduct, omission, advantage, or implicated action is positioned as wrong, blameworthy, unfair, disappointing, illegitimate, or otherwise negatively evaluated, providing the evaluative basis for guilt; *Behavior*—the guilt-relevant appraisal and evaluation are used as pressure to promote corrective, reparative, compliant, helping, compensatory, inhibitory, or otherwise altered conduct (Tangney et al., 2007; Boster et al., 2016). *Boundary:* Code *Guilt trip* when responsibility, obligation, indebtedness, implication, transgression, inequity, or failure attributed to the recipient or a relevant ingroup functions as pressure on the recipient's conduct. Code *Appeal to pity / sympathy* when another person's or group's suffering, vulnerability, hardship, or need functions instead as a basis for sympathy, leniency, aid, forgiveness, or another favourable response. The presence of another person's suffering does not by itself distinguish the Tactics: where the recipient is represented as responsible for that suffering and that responsibility supplies the pressure, *Guilt trip* may apply; where the suffering itself supplies the sympathy-based warrant, *Appeal to pity / sympathy* may apply. Co-code only where both operations are independently recoverable.

- **Induced hypocrisy / self-inconsistency pressure.** Foreground a discrepancy between a standard, principle, norm, position, recommendation, or course of conduct endorsed or advocated by the recipient

and the recipient's own represented past or current conduct, using that self-inconsistency as psychological pressure for subsequent conduct aligned with the endorsed standard (Aronson et al., 1991; Stone et al., 1994; Stone et al., 1997; Priolo et al., 2019). *How to spot:* the available message or message sequence establishes (i) a standard, principle, norm, position, recommendation, or behavioural expectation represented as endorsed, advocated, accepted, or promoted by the recipient; (ii) past or current conduct by that same recipient represented as inconsistent with the endorsed standard; and (iii) a communicative operation that foregrounds the discrepancy between the two and uses that self-inconsistency as pressure for corrective, proattitudinal, norm-consistent, or otherwise aligned subsequent conduct. For example, "You have repeatedly argued that every minister should disclose outside income. You still have not disclosed your own consulting payments. Apply the same standard to yourself and publish them" confronts the recipient with a discrepancy between an endorsed transparency standard and their own represented conduct and uses that discrepancy as pressure for behavioural alignment. The endorsed standard or inconsistent conduct may have occurred before the annotated message only where it is explicitly recoverable from the message or available message sequence. A general accusation that the recipient is inconsistent, hypocritical, insincere, or contradictory is insufficient unless both sides of the relevant self-discrepancy are recoverable and the discrepancy itself functions as pressure toward subsequent alignment. The Tactic does not require the words "hypocrisy" or "hypocrite", nor does annotation require determining whether the recipient actually experiences cognitive dissonance, psychological discomfort, guilt, shame, embarrassment, self-threat, or any other internal state. *Cognitive dissonance* is treated as a theorised psychological mechanism through which the observable self-inconsistency operation may exert influence rather than as an independently annotated Tactic or TargetConstruct (Stone et al., 1994; Stone et al., 1997; Priolo et al., 2019). Annotation does not imply deceptive, malicious, exploitative, or otherwise wrongful intent. A communicator may sincerely and reasonably use another person's professed standard and inconsistent conduct to encourage greater consistency. The Tactic is classified as *Manipulation* because the defining influence operation uses the recipient's own psychologically consequential self-discrepancy as pressure for alignment rather than merely supplying substantive evidence or argument concerning the merits of the target conduct. *Target constructs: Perception/Attention*—both the recipient's endorsed standard and their inconsistent conduct are recoverable from the message, and the defining operation places those elements into direct comparison so that their discrepancy becomes the focal relation of the influence attempt rather

than either element merely being mentioned independently (Aronson et al., 1991; Stone et al., 1994); *Behavior*—the salient self-inconsistency is used as pressure to elicit corrective, proattitudinal, norm-consistent, or otherwise aligned subsequent conduct (Stone et al., 1994; Priolo et al., 2019). *Boundary:* Code *Commitment / consistency* when a prior explicit choice, pledge, promise, or commitment is invoked as a reason for corresponding follow-through without additionally confronting the recipient with their own contradictory past or current conduct. Code *Induced hypocrisy / self-inconsistency pressure* when an endorsed standard or position is juxtaposed with the recipient's own inconsistent conduct and that discrepancy itself is used as pressure for behavioural alignment. The two may co-occur where both a prior commitment and contradictory conduct independently perform their respective influence operations. Code *Guilt trip* when guilt-relevant responsibility, obligation, indebtedness, inequity, transgression, or failure supplies the pressure for corrective or compliant conduct without requiring an endorsed-standard–own-conduct discrepancy. A hypocrisy induction may also evoke guilt, but possible guilt is not sufficient to license *Guilt trip*; co-code only where guilt-relevant responsibility or obligation independently performs that Tactic's defining operation. Code *Shaming / derogation* when negative evaluation of the recipient's character, worth, status, identity, or social image supplies the behavioural pressure. Calling the recipient a "hypocrite", "fraud", or comparable derogatory term does not by itself license *Induced hypocrisy / self-inconsistency pressure*; the relevant endorsed-standard–own-conduct discrepancy must be recoverable. Code *Ad hominem / personal-credibility attack* when another actor's inconsistency or hypocrisy is invoked primarily to discount that actor's credibility, standing, claim, argument, or contribution rather than to use the recipient's own self-inconsistency as pressure for corrective alignment. For example, "She preaches climate restraint but flies privately, so do not trust her climate claims" may instantiate a personal-credibility attack, whereas "You ask everyone else to reduce unnecessary flights but continue taking them yourself; follow the standard you advocate" may instantiate *Induced hypocrisy / self-inconsistency pressure*. The Tactics may co-occur where both operations are independently evidenced.

- **Appeal to pity / sympathy.** Present another person's or group's suffering, distress, vulnerability, misfortune, hardship, or need and invoke the pity, sympathy, compassion, or mercy associated with that condition as a basis for accepting a judgement, adopting a more favourable or lenient evaluation, or taking a requested course of action (Walton, 1997; Wood et al., 2014; Condon and DeSteno, 2011). *How to spot:* the message presents another person's or group's suffering,

vulnerability, hardship, misfortune, or need and uses that condition as a reason for the recipient to accept or revise a judgement about responsibility, guilt, culpability, deservingness, or another proposition; to evaluate an actor or conduct more sympathetically, leniently, or mercifully; or to provide aid, leniency, support, an exception, forgiveness, or another favourable response. Mere description of suffering, tragedy, vulnerability, or emotional hardship is insufficient unless that condition functions as an influence warrant. Annotators identify the use of pity- or sympathy-relevant circumstances and do not determine whether the circumstances genuinely excuse the conduct, alter factual responsibility, or justify the requested response. *Target constructs: Belief*—the suffering, hardship, vulnerability, or mitigating circumstances are used as a warrant bearing on acceptance of a proposition or judgement, for example concerning responsibility, guilt, culpability, deservingness, or the interpretation of an actor's conduct; *Attitude*—the same circumstances are used to position an actor, conduct, decision, or outcome toward greater sympathy, leniency, forgiveness, approval, or reduced condemnation; *Behavior*—pity, sympathy, compassion, or mercy elicited by the presented circumstances is used to promote aid, leniency, support, forgiveness, exemption, withholding punishment, or another requested response (Walton, 1997; Wood et al., 2014; Condon and DeSteno, 2011).

- **Love-bombing.** Use unusually intense, accelerated, or repeated affection, attention, praise, gifts, romantic commitment cues, or future-oriented intimacy to establish trust, favourable attachment, or rapid relational commitment that can be leveraged for influence or control (Strutzenberg et al., 2017; Arabi, 2023; Schokkenbroek and Snaphaan, 2025). *How to spot:* the message or recoverable message sequence contains dense or repeated expressions of affection, idealisation, extraordinary praise, intensive attention, gifts, declarations of exceptional closeness, or rapid future-oriented commitment (e.g., "I've never felt this way about anyone", "you're my soulmate", "we were meant to find each other", premature marriage or shared-future claims) and uses that intensity to accelerate trust, attachment, commitment, reciprocation, or subsequent influence. Ordinary affection, a compliment, a gift, enthusiastic romantic expression, or frequent communication is insufficient unless the unusual intensity or acceleration of relational closeness performs the manipulative operation. Annotators do not infer narcissism, deceptive intent, or an abusive relationship merely from affectionate language. *Target constructs: Belief*—the defining operation promotes acceptance of a relational representation in which the source is understood as unusually caring, trustworthy, devoted, intimate, or committed and the relationship as exceptionally close, genuine, or secure; *Atti-*

tude—intense affection, praise, attention, or idealisation is used to cultivate favourable evaluation, attraction, attachment, or positive regard toward the source or relationship; *Behavior*—the accelerated trust and attachment are used to promote relational engagement, reciprocation, commitment, continued interaction, compliance, or other conduct that can subsequently be leveraged for influence or control (Strutzenberg et al., 2017; Schokkenbroek and Snaphaan, 2025). *Boundary:* Code *Love-bombing* when unusually intense, accelerated, dense, or repeated affiliative behaviour functions to establish or accelerate perceived intimacy, trust, attachment, or relational commitment. Code *Ingratiation* when flattery, agreement, similarity, favourable self-presentation, or other rapport-building behaviour functions primarily to increase favourable evaluation or liking of the source without the defining intensity or acceleration of relational closeness required for *Love-bombing*. Gifts or benefits do not by themselves license *Love-bombing*. Code *Dangling contingent rewards* when a prospective benefit is made conditional on identifiable conduct, and code *Reciprocity* when a benefit already provided functions as a warrant for a corresponding return act. Co-code only where the relational-bonding operation and the additional reward or reciprocity structure are independently recoverable.

- **Shaming / derogation.** Use negative social, moral, or status evaluation of the recipient, their character, identity, or anticipated noncompliance as pressure to conform, comply, withdraw, refrain, or otherwise alter conduct (Scheff, 1988; Tangney et al., 2007; Grasmick et al., 1991). *How to spot:* the message links a requested, expected, or disfavoured course of conduct to a derogatory or devaluing judgement about the recipient's character, identity, worth, status, or social image (e.g., "only a coward would back out now", "don't be pathetic—sign it", "everyone will see what kind of person you are if you refuse"). The negative evaluation must function as pressure on the recipient's conduct; an explicit command is not required where the expected conformity, compliance, withdrawal, restraint, or other behavioural change is recoverable from the message. Mere insult, mockery, pejorative labeling, moral condemnation, or discussion of shame is insufficient unless the derogatory evaluation is tied to the recipient's conduct as influence leverage. Public exposure is not required, and annotators do not infer whether the recipient actually experiences shame, accepts the derogatory evaluation, or changes behaviour. *Target constructs: Identity*—the defining operation threatens or negatively positions the recipient's self-concept, character, social image, status, or belonging by linking a course of conduct or noncompliance to a devalued identity or social evaluation; *Behavior*—the threatened negative self- or social evaluation is used as pressure to promote conformity, compliance, withdrawal, re-

straint, or another change in conduct (Scheff, 1988; Tangney et al., 2007; Grasmick et al., 1991). *Boundary:* Code *Shaming / derogation* when a negative evaluation of the recipient's character, identity, worth, status, or social image functions as pressure on conduct. Code *Guilt trip* when responsibility, indebtedness, implication, transgression, or failure of obligation is used as pressure on conduct. The two may co-occur when both guilt-relevant responsibility and derogatory self- or social evaluation independently function as influence warrants. Mere insulting or derogatory language without pressure on the recipient's conduct does not by itself license *Shaming / derogation*.

- **Pride / ego appeals.** Link acceptance of a proposition, evaluation of a target, or a course of conduct to a positively valued self-concept, trait, competence, status, or social image, so that agreeing, approving, or acting is presented as consistent with being intelligent, discerning, brave, successful, sophisticated, principled, elite, or otherwise positively regarded (Walton et al., 2008; Howard et al., 1995; Wheeler et al., 2005; Perkins and Forehand, 2012). *How to spot:* the message presents accepting a claim, evaluating an object or position favourably, or taking a course of action as evidence or expression of a desirable identity, trait, competence, status, or self-image (e.g., "someone as intelligent as you can see that this is true", "people with real taste appreciate this", "successful people act now"). The defining cue is not praise, prestige language, or a positive description alone: the valued self- or social image must function as leverage bearing on what the recipient is encouraged to believe, approve of, choose, or do. Annotators do not determine whether the recipient actually possesses the invoked trait, experiences pride, values the status, or changes their response. *Target constructs: Identity*—the defining operation activates or affirms a positively valued self-concept, trait, competence, status, or social image and makes the targeted response diagnostic of or consistent with that identity; *Belief*—the valued self-image may be used as leverage for accepting, endorsing, or rejecting a proposition, explanation, interpretation, or other claim; *Attitude*—the valued self-image may be used as leverage for adopting a more favourable or unfavourable evaluation of an object, actor, issue, position, behaviour, or outcome; *Behavior*—the valued self-image or status may be made contingent on performing, choosing, joining, supporting, refraining from, or otherwise altering conduct (Howard et al., 1995; Wheeler et al., 2005; Perkins and Forehand, 2012). *Boundary:* Mere praise, compliment, or flattering description does not by itself license *Pride / ego appeals*; the positive identity, competence, status, or self-image must be tied to the response being promoted. Code *Ingratiation* when praise or flattery functions primarily to cultivate favourable regard toward the

source. Code *Shaming / derogation* when a negatively valued character, identity, status, or social image is used as pressure on conduct. Code *Altercasting* when a social role or relational position is assigned or activated and role-consistent expectations provide the influence warrant. These Tactics may co-occur when their respective operations are independently evidenced.

- **Dangling contingent rewards.** Offer or promise a prospective benefit, prize, bonus, privilege, access, discount, or other reward as conditional on the recipient performing a requested action, so that the anticipated reward functions as instrumental leverage for compliance (Kelman, 1958; Mathur et al., 2019; Luguri and Strahilevitz, 2021). *How to spot:* the message makes receipt of, eligibility for, or access to a positively valued reward contingent on an identifiable course of conduct (e.g., "sign up to receive a free month", "complete the survey and earn €10", "refer three friends to unlock your bonus", "donate to enter the prize draw"). The defining cue is the contingency between the requested conduct and the offered reward; mention of a desirable outcome, prize, gift, discount, or benefit alone is insufficient unless it is presented as an incentive conditional on the recipient's action. Urgency, scarcity, uncertain odds, vague eligibility conditions, or obscured redemption terms may co-occur but are not required for this Tactic. Annotators do not determine whether the reward is objectively valuable, sufficient to justify the action, actually motivates the recipient, or is ultimately delivered. *Target constructs: Behavior*—the anticipated reward is made contingent on performing, choosing, joining, signing up, sharing, purchasing, complying, or otherwise altering observable conduct, and functions as instrumental leverage for that action (Kelman, 1958). *Boundary:* Code *Dangling contingent rewards* when a reward offered or controlled by the influencing source is made conditional on the recipient's conduct. Code *Reciprocity* when a benefit, favour, gift, service, or concession has already been provided and that prior provision functions as a warrant for a corresponding return act. The distinction is temporal and relational: *Dangling contingent rewards* uses a prospective conduct–benefit contingency, whereas *Reciprocity* invokes a prior benefit as the basis for return conduct. Code *Gain/loss goal framing* when desirable or undesirable consequences of performing or not performing a focal behaviour are represented as outcomes of that behaviour rather than as a contingent reward offered in exchange for it. Code *Scarcity appeal* when limited availability, quantity, access, or time functions independently as leverage. Where reward terms, eligibility, odds, availability, or redemption conditions are misleading, obscured, or materially equivocal, an applicable Deception Tactic may additionally be coded.

- **Affective associative cueing.** Introduce, juxtapose, insinuate, or foreground an affectively, morally, socially, or identity-relevant cue or association so that its pre-existing emotional or motivational significance makes a focal object, actor, relation, or interpretation disproportionately salient, including where the message contains no explicit emotional evaluation, judgement, or behavioural request (MacKay et al., 2004; Gable and Harmon-Jones, 2010; Brosch and Van Bavel, 2012; Kaiser et al., 2006; Herring et al., 2013). *How to spot:* the message uses an affectively significant cue, association, juxtaposition, taboo, socially meaningful category, identity relation, or identity incongruity to foreground a focal representation or make it difficult to disregard. The affective significance may be explicit or may remain implicit in a recoverable association: the message need not state that the target is frightening, disgusting, shameful, hateful, outrageous, or otherwise negatively evaluated, and need not state a conclusion or request an action. Observable evidence may include strategic juxtaposition of a target with an affectively charged category; insinuation of a socially, morally, or identity-relevant association; repeated or otherwise marked foregrounding of such an association; or construction of an identity incongruity whose affective significance is recoverable from the message. A controversial, sexual, religious, political, identity-related, or otherwise potentially emotional topic is not sufficient by itself. Annotators must identify message-internal evidence that the association is being activated or foregrounded as part of the communicative operation and must not infer a recipient's private prejudices, identity commitments, emotional reaction, or culturally specific associations solely from the topic label. *Target constructs: Perception/Attention*—the defining operation uses the pre-existing affective or motivational significance of a cue or association to capture, hold, redirect, or narrow attentional focus, or to increase the salience of the associated object, actor, relation, or interpretation relative to competing message content (MacKay et al., 2004; Gable and Harmon-Jones, 2010; Brosch and Van Bavel, 2012; Kaiser et al., 2006; Herring et al., 2013; Brady et al., 2020). *Boundary:* Do not code *Affective associative cueing* merely because a topic is controversial or may evoke emotion in some audiences. The message must provide recoverable evidence that an affectively, morally, socially, or identity-relevant association is being activated or foregrounded as part of the influence operation. Code *Loaded language* when explicitly evaluative or emotionally valenced wording positions a target. Code *Appeal to fear* or *Threat framing* when a represented danger or threat relation provides the influence warrant. Code *Shaming / derogation* when negative evaluation of the recipient's character, identity, worth, status, or social image functions as pressure on conduct. Affective associative cueing may co-occur with these or other Tactics when the

associative-attentional operation is independently evidenced. Code *Associative transfer / guilt-or-honor by association* when an association instead functions as a warrant for attributing to a focal target the commitments, qualities, credibility, credit, discredit, prestige, stigma, or evaluative standing represented as belonging to an associated actor, group, institution, symbol, or other entity. Associative transfer does not by itself license *Affective associative cueing*; co-code only where the association independently performs both the transfer-of- judgement operation and a recoverable differential attentional or salience operation.

- **Manipulative choice architecture.** Structure the presentation, availability, sequence, or effort associated with available responses so that one course of action is substantially easier, more immediate, defaulted, or otherwise practically privileged while one or more alternatives are made harder to access, complete, reverse, or select, thereby steering choice through the architecture of the decision environment rather than through the substantive merits of the alternatives (Mathur et al., 2019; Luguri and Strahilevitz, 2021). *How to spot:* the message or recoverable interaction structure presents an identifiable choice in which alternatives are asymmetrically accessible or actionable—for example, one option requires a single immediate response while another requires additional steps, navigation, delay, repeated confirmation, special effort, or an otherwise more burdensome path. The relevant architecture may be verbal, visual, spatial, or interactional; no particular interface modality is required. Defaults, obstruction, difficult reversal or cancellation, and structurally unequal response paths may provide evidence when they function to steer selection. Mere difference in wording, attractiveness, argument quality, or substantive consequences is insufficient unless the structure of access to the alternatives is itself asymmetric. Annotators do not infer the recipient's actual preference, whether the architecture successfully changes choice, or the designer's private intention. *Target constructs: Agency/Choice*—the defining operation alters the practical availability, ease, or viability of alternatives within the represented choice environment, making some responses materially easier or harder to select, complete, or reverse; *Behavior*—the asymmetric choice structure is organised to steer the recipient toward selecting, accepting, continuing, disclosing, purchasing, subscribing, remaining enrolled, or otherwise performing the structurally privileged response (Mathur et al., 2019; Luguri and Strahilevitz, 2021). *Boundary:* Code *Manipulative choice architecture* when the structure of access, effort, default status, sequencing, or reversibility of available responses provides the steering mechanism. Code applicable Framing Tactics when alternatives remain comparably accessible but their rep-

resentation or description is selectively framed. Specific dark-pattern implementations may additionally instantiate Deception, Scarcity, Social proof, Shaming, Loaded language, or other Tactics when those operations are independently evidenced. The label does not require treating every practice described in the dark-pattern literature as an instance of this Tactic.

- **Gaslighting / epistemic destabilization.** Undermine the recipient's confidence in their own perception, memory, reasoning, or judgement by treating their experience or account not merely as mistaken but as evidence that they are unreliable, confused, irrational, unstable, or otherwise unable to judge the relevant matter independently (Abramson, 2014; Spear, 2023; Darke et al., 2025). *How to spot:* the message challenges an identifiable perception, memory, interpretation, or prior account while also discrediting the recipient's capacity to perceive, remember, reason, or judge reliably (e.g., "that never happened—you're imagining things again", "you always distort what people say; you cannot trust your memory", "you're being paranoid, as usual"). Ordinary disagreement, contradiction, correction, denial, or competing recollection is insufficient: the recipient's epistemic competence or self-trust must itself be made part of the influence operation. Where the tactic depends on a repeated pattern, the relevant prior challenges must be contained in or explicitly recoverable from the message or message sequence. Annotators do not determine whose account is factually correct, whether the source is knowingly lying, whether the recipient actually doubts themselves, or whether a broader abusive relationship exists. *Target constructs: Belief*—the defining operation challenges acceptance of the recipient's own perceptions, memories, interpretations, or judgements and promotes reliance on a competing account or on the proposition that the recipient's own cognitive representation is unreliable; *Identity*—the tactic undermines the recipient's self-conception as a competent and independent perceiver, rememberer, reasoner, or judge (Abramson, 2014; Spear, 2023); *Agency/Choice*—the defining operation undermines the recipient's epistemic self-direction by discrediting their capacity to perceive, remember, reason, or judge the relevant matter independently, thereby impairing their capacity for independent evaluation of the information at issue (Spear, 2023; Proust, 2008).*Boundary:* Do not code *Gaslighting / epistemic destabilization* for ordinary disagreement, factual correction, denial, conflicting memories, or criticism of a particular conclusion. The defining operation is an attack on the recipient's epistemic self-trust or standing as an independent judge, not merely an assertion that a proposition is false. Code *Doubt-seeding* when uncertainty, disagreement, evidential limitations, source-reliability concerns, or procedural limitations are used

to reduce confidence in an external claim, conclusion, evidence base, source, consensus, or inference. Code *Gaslighting / epistemic destabilization* when the recipient's own capacity to perceive, remember, reason, interpret, or judge reliably is itself discredited as part of the influence operation. Applicable Deception Tactics may additionally be coded where the message independently satisfies their observable conditions; gaslighting itself is annotated truth-agnostically and does not require establishing that the source's competing account is false.

B.7 Coercion

Within the Coercion family, *Explicit threats* and *Implied threats / coercive intimidation* characterise how the conduct–consequence leverage relation is expressed. The remaining Tactics identify additional observable structure in that coercive relation: *Blackmail / disclosure leverage* identifies informational disclosure as the sanctioning leverage; *Coercive ultimatum* identifies a final, non-negotiable, or bounded compliance condition; and *Collective punishment threats* identify distributed sanction liability extending beyond the immediately responsible actor. These distinctions are not mutually exclusive. A specialised coercive Tactic may therefore be coded together with *Explicit threats* or *Implied threats / coercive intimidation* when the same coercive operation independently satisfies both the relevant expression-form rule and the specialised structural rule. A second or distinct adverse consequence is not required. Do not, however, code both explicit and implied threat forms merely because an explicit utterance also carries ordinary pragmatic implications; both forms require independently recoverable expression structures.

- **Explicit threats.** Explicitly present a prospective adverse consequence, penalty, sanction, deprivation, injury, loss, or other harm and directly make that consequence contingent on the recipient's performance, continuation, cessation, or nonperformance of an identifiable course of conduct, thereby communicatively using the prospective harm as leverage to compel or deter behaviour (Schelling, 1966; Raven, 1993; De Dreu, 1995; Evans and Twyman-Musgrove, 1998; Waxman, 2014; Reichberg and Syse, 2018). *How to spot:* the message establishes (i) an identifiable act, omission, continuation, cessation, or other course of conduct associated with the recipient or target actor; (ii) an explicitly stated prospective adverse consequence, penalty, sanction, deprivation, injury, loss, or other harm; and (iii) an explicitly stated or directly recoverable contingency linking the target's conduct to that adverse consequence, such that the consequence is communicatively deployed as leverage over the target's conduct. The adverse consequence may be represented as imposed directly by the source, by an authority, in-

stitution, associated actor, group, or proxy, or as an outcome whose occurrence the source can facilitate, authorise, permit, withhold protection from, or otherwise communicatively appropriate as leverage over the target's conduct. The source therefore need not represent themselves as the direct physical or institutional agent of the threatened harm. Canonical forms include "if you do X, Y will happen", "do X or Y", and "unless you stop X, you will face Y", where the available message makes the conduct–consequence contingency and the use of Y as coercive leverage explicit. For example, "Vote against the bill and I will remove you from the committee" explicitly uses an organisational sanction to deter an identifiable act. Likewise, "Publish the report and our supporters will make sure you regret it" may instantiate the Tactic where the third-party consequence is explicitly made contingent on publication and deployed as leverage, even though the source does not represent themselves as the direct executor of the harm. The threatened conduct may be compellent—requiring the target to perform or change an action—or deterrent—requiring the target to refrain from or discontinue an action (Schelling, 1966; De Dreu, 1995; Waxman, 2014; Reichberg and Syse, 2018). The source need not use the words "threat", "punishment", "or else", or "sanction". What must be explicit is the relevant conduct–consequence relation and the adverse consequence used as leverage. Annotators do not determine whether the threatened consequence is lawful, morally justified, proportionate, sincerely intended, likely to be carried out, or within the source's actual power to execute. The source need not be represented as the direct agent or controller of the prospective harm; what must be recoverable is that the communication appropriates the adverse consequence as leverage over identifiable target conduct. Nor does annotation imply malicious or hostile intent: the source may sincerely regard the threatened sanction or pressure as justified, protective, necessary, or appropriate (Evans and Twyman-Musgrove, 1998; Beller et al., 2005). *Target constructs: Agency/Choice*—the defining operation makes an otherwise available course of conduct materially more costly or non-viable by attaching a prospective adverse consequence communicatively deployed as leverage to its performance or nonperformance, thereby constraining the practical conditions under which the recipient can exercise choice (Raven, 1993; Schelling, 1966); *Behavior*—the adverse consequence is made contingent on an identifiable act, omission, continuation, or cessation in order to compel, deter, preserve, or otherwise alter that conduct (De Dreu, 1995; Kellermann and Shea, 1996; Waxman, 2014). *Boundary:* Do not code *Explicit threats* merely because a message describes danger, predicts an adverse outcome, reports a penalty rule, states a possible legal or institutional consequence, or contains frightening language. The adverse consequence must be communicatively deployed

as contingent leverage around identifiable target conduct.

Code *Appeal to fear* when a threatening or harmful outcome is used as a reason for adopting a protective response without that prospective harm being communicatively appropriated as coercive leverage over the target's conduct. For example, "If you remain in the evacuation zone, the hurricane may kill you" warns of a dangerous consequence and may motivate evacuation, but does not by itself deploy the hurricane as a coercive sanction. By contrast, "Leave the district or our supporters will make sure your business does not survive the week" may constitute a threat where the prospective third-party harm is explicitly made contingent on the target's conduct and communicatively used as leverage.

Code *Threat framing* when a person, group, event, condition, or development is represented as dangerous or harmful without independently using the prospective harm as leverage over identifiable target conduct.

Code *Implied threats / coercive intimidation* when the same basic conduct–consequence leverage relation is recoverable but the adverse consequence, contingency, source commitment, connection to the harm, or another component is conveyed indirectly through implication, allusion, contextual inference, or communicative action. Code *Coercive ultimatum* when a coercive demand additionally establishes the non-negotiable, final, or time-bounded structure specified by that Tactic. Code *Blackmail / disclosure leverage* where prospective disclosure or non-disclosure of information independently constitutes the sanctioning leverage. Code *Collective punishment threats* where sanction liability extends beyond the immediately responsible actor such that additional members of a group or collective prospectively bear consequences on account of the relevant conduct. These Tactics may co-occur when their respective defining operations are independently evidenced.

- **Implied threats / coercive intimidation.** Communicate an adverse consequence, penalty, sanction, deprivation, injury, exposure, loss, or other harm as leverage over an identifiable course of conduct without stating the full coercive relation directly, such that the relevant conduct–consequence contingency, prospective harm, source commitment, or connection between the source and the harm is recoverable through implication, allusion, contextual inference, or communicative action (Schelling, 1966; Kellermann and Shea, 1996; Haigh et al., 2011; Reichberg and Syse, 2018; Jankovic and Roeben, 2025). *How to spot:* the available message or message sequence supports (i) an identifiable act, omission, continuation, cessation, decision, or other course of conduct associated with the recipient or target actor; (ii) a prospective ad-

verse consequence; and (iii) a recoverable relation in which that consequence is communicatively used as leverage to compel, deter, preserve, or otherwise alter the target conduct, while at least one component of the coercive relation remains indirect rather than fully stated. The relevant consequence need not be represented as harm that the source will personally inflict. It may instead be represented as harm imposed by an associated person or group, an institution or authority, a coercive proxy, or another actor; as harm whose occurrence the source can facilitate, authorise, permit, or decline to prevent; or as a prospective adverse event that the source invokes and communicatively appropriates as leverage over the target's conduct (Reichberg and Syse, 2018; Jankovic and Roeben, 2025). For example, immediately before a politically consequential vote, a party leader's statement that "committee assignments can be surprisingly temporary; I'm sure you'll make the right decision tomorrow" may instantiate the Tactic where the available context makes the vote, possible loss of the assignment, and the use of that consequence as leverage recoverable. Similarly, in response to "I'm still testifying tomorrow", a statement such as "That's your choice. Your neighborhood is rough; things have a habit of disappearing there" may instantiate the Tactic where the message sequence makes the prospective loss function as leverage against testifying. The speaker need not explicitly claim responsibility for the possible loss. Likewise, "You can publish the investigation if you want; some of our supporters are much less restrained than I am" may communicate prospective third-party retaliation as leverage over publication. Ominous wording, aggressive tone, demonstrations of power, statements such as "you'll regret this", references to dangerous environments, or predictions that another actor may cause harm do not by themselves license the Tactic. The available communicative context must make both an identifiable target conduct and the use of a prospective adverse consequence as coercive leverage recoverable. Annotators must not supply an unstated demand, contingency, actor, or threat relation solely from speculation or external knowledge (Haigh et al., 2011; Jankovic and Roeben, 2025). Annotators do not determine whether the source privately intended to threaten, whether the prospective harm is lawful, justified, proportionate, credible, or likely to occur, whether the source actually controls the actor or process through which the harm could occur, or whether the recipient experienced fear or intimidation. Annotation does not imply malicious intent: a source may sincerely regard coercive pressure as justified, defensive, protective, necessary, or otherwise appropriate. The Tactic records the recoverable coercive influence operation expressed by the communication rather than the source's private motive or the normative legitimacy of the threat. *Target constructs: Agency/Choice*—the defining operation makes an otherwise available course of conduct

materially more costly or non-viable by attaching or communicatively invoking a prospective adverse consequence around its performance or nonperformance, thereby constraining the practical conditions under which the recipient can exercise choice (Raven, 1993; Schelling, 1966; Reichberg and Syse, 2018); *Behavior*—the prospective adverse consequence is communicatively used to compel, deter, preserve, terminate, or otherwise alter an identifiable act, omission, decision, or course of conduct (Kellermann and Shea, 1996; Haigh et al., 2011; Reichberg and Syse, 2018). *Boundary:* Do not require the source to represent themselves as the direct physical or institutional agent of the threatened harm. A threatened consequence may be carried out by an associated person or group, an institution, an authority, a proxy, or another actor, or may consist of harm whose occurrence the source represents as able to facilitate, permit, or decline to prevent. What is required is that the prospective adverse consequence be communicatively used as leverage over identifiable target conduct. Conversely, mere reference to environmental danger, independent third-party risk, or a possible adverse event does not constitute an implied threat merely because the information might influence behaviour. For example, "Do not leave your bicycle outside tonight; theft is common in this neighborhood" ordinarily communicates an independent risk or warning. By contrast, following "I will testify tomorrow", "Your bicycle is still outside, isn't it? Things disappear around here" may instantiate an implied threat where the available interaction makes the prospective loss function as coercive leverage concerning the testimony. The distinction is therefore not simply whether the speaker personally causes the harm, but whether the message deploys the prospective harm as leverage over the target's conduct (Reichberg and Syse, 2018; Jankovic and Roeben, 2025). Code *Explicit threats* when the adverse consequence and its contingency on identifiable target conduct are directly stated. Code *Implied threats / coercive intimidation* when the same basic coercive relation is recoverable but the consequence, contingency, source commitment, connection to the harm, or another component is conveyed indirectly through implication, allusion, contextual inference, or communicative action. Do not code both merely because an explicit threat also carries pragmatic implications; co-code them only where distinct explicit and implied threat operations are independently evidenced. Code *Appeal to fear* when an adverse outcome, danger, or risk is used to motivate a protective response without that prospective harm being communicatively appropriated as coercive leverage by the influencing source. Code *Threat framing* when a person, group, condition, event, or development is represented as dangerous without independently using the prospective harm as coercive leverage over target conduct. Code *Coercive ultimatum* when the coercive demand additionally satisfies the final-

ity, non-negotiability, deadline, or other defining structure specified by that Tactic. Code an applicable *Blackmail / disclosure leverage* Tactic when disclosure, exposure, or another specifically extortionate form of leverage independently satisfies its defining conditions. Code *Collective punishment threats* when prospective harm to a family, group, community, organisation, population, or other collective is used as leverage over the target's conduct. These Tactics may co-occur when their respective operations are independently evidenced.

- **Blackmail / disclosure leverage.** Use prospective disclosure, publication, reporting, exposure, or transmission of information as coercive leverage over an identifiable course of conduct by making disclosure or non-disclosure contingent on the recipient's compliance, noncompliance, payment, forbearance, or other response (Shavell, 1993; Gordon, 1993; Lamond, 1996; Elhauge, 2016). *How to spot:* the available message or message sequence establishes (i) information that the source claims, demonstrates, or represents themselves as able to disclose, publish, report, expose, transmit, or otherwise make available to another person, group, institution, authority, or public; (ii) an identifiable act, omission, payment, concession, continuation, cessation, or other course of conduct sought from the recipient or target actor; and (iii) a recoverable contingency in which disclosure or non-disclosure of the information functions as leverage over that conduct. For example, "I have the messages showing the undeclared donation. Withdraw from the race by Friday or I send them to every newspaper" makes prospective disclosure of compromising information contingent on withdrawal and uses that disclosure as coercive leverage. The information need not concern reputation specifically. Prospective disclosure may be represented as consequential because of possible effects on privacy, relationships, employment, legal standing, political position, financial interests, social standing, reputation, or another valued interest. Nor must the source demand money: the demanded response may consist of payment, silence, withdrawal, resignation, support, opposition, voting, non-reporting, non-testimony, transfer of property, provision of information, or any other identifiable conduct. The defining operation is the use of prospective informational disclosure or exposure as contingent leverage (Shavell, 1993; Gordon, 1993). Annotators do not determine whether the information is true, whether the source obtained it lawfully, whether disclosure would actually harm the recipient, whether the threatened disclosure or requested conduct is lawful or morally justified, whether the communication would satisfy the legal elements of blackmail or extortion in a particular jurisdiction, whether the source privately intends to carry out the disclosure, or whether the recipient actually experiences coercion. Annotation records the observable

disclosure-for-conduct leverage structure and does not by itself imply malicious, deceptive, exploitative, or otherwise wrongful intent (Lamond, 1996; Elhauge, 2016). *Target constructs: Agency/Choice*—the defining operation alters the practical conditions under which the recipient can exercise choice by making disclosure or non-disclosure of information contingent on an identifiable response, thereby attaching a prospective informational cost to noncompliance or an informational benefit to compliance and constraining the recipient's practical decision space (Lamond, 1996; Elhauge, 2016); *Behavior*—the threatened disclosure or promised non-disclosure is used as leverage to elicit, deter, preserve, terminate, or otherwise alter an identifiable act, omission, payment, concession, or course of conduct (Shavell, 1993; Gordon, 1993). *Boundary:* Do not code *Blackmail / disclosure leverage* merely because a message possesses, discusses, reveals, threatens to reveal, or announces an intention to publish sensitive, private, embarrassing, compromising, incriminating, or otherwise consequential information. The available communicative context must establish an identifiable target response and a recoverable contingency through which disclosure or non-disclosure functions as leverage over that response. Code *Explicit threats* additionally when the disclosure-based conduct–consequence contingency is directly stated, and code *Implied threats / coercive intimidation* additionally when that same coercive relation is recoverable but one or more components are conveyed indirectly. These labels capture the expression form of the threat, whereas *Blackmail / disclosure leverage* captures the distinctive use of prospective informational disclosure or non-disclosure as sanctioning leverage. A separate adverse consequence is not required for co-coding. Do not require the threatened information to concern the recipient's reputation or identity. Reputational harm is one possible consequence of disclosure but is not constitutive of the Tactic. An independently evidenced Identity, Attitude, Framing, or other Tactic may additionally be coded where the message separately performs that operation. Code *Coercive ultimatum* additionally only where the communication independently satisfies the non-negotiable, final, deadline-based, or other defining structure specified for that Tactic. These Tactics may co-occur when their respective operations are independently evidenced.

- **Coercive ultimatum.** Present a sanction-backed demand as a bounded, final, or non-negotiable compliance condition, such that failure to perform an identifiable act, omission, decision, continuation, or cessation within the represented compliance window carries a prospective adverse consequence (George, 1991; Schelling, 1966; Reichberg and Syse, 2018). *How to spot:* the available message or message sequence establishes (i) an identifiable target conduct;

(ii) a prospective adverse consequence used as leverage over that conduct; and (iii) an explicit deadline, immediate last-chance condition, or otherwise recoverable final or non-negotiable compliance window. For example, "Withdraw the resolution by noon or you will be removed from the committee" combines a sanction-backed demand with a bounded compliance window. A coercive demand does not become an ultimatum merely because it is forceful, urgent, or emphatic. The communication must represent compliance as bounded by an identifiable finality or time condition. Conversely, a final or time-limited offer does not constitute this Tactic unless noncompliance is backed by a prospective adverse consequence used as coercive leverage. Annotators do not determine whether the deadline is reasonable, whether the sanction is credible, lawful, justified, or proportionate, whether the source privately intends to carry it out, or whether the recipient experiences urgency or fear. Annotation records the observable sanction-backed finality structure. *Target constructs: Agency/Choice*—the sanction-backed demand constrains the practical terms of choice and additionally bounds the conditions or interval within which the target may select the non-sanctioned course of action (George, 1991; Reichberg and Syse, 2018); *Behavior*—the final or time-bounded sanction condition is directed at eliciting, preventing, terminating, maintaining, or otherwise altering an identifiable course of conduct. *Boundary:* Code *Coercive ultimatum* only where a prospective adverse consequence is used as leverage over identifiable conduct and the communication additionally represents compliance as subject to an explicit or recoverable final, non-negotiable, last-chance, or bounded condition. Code *Explicit threats* additionally when the sanction-backed conduct–consequence relation is directly stated, and code *Implied threats / coercive intimidation* additionally when that relation is recoverable but one or more components remain indirect. A second sanction is not required: the threat-form Tactic records how the coercive relation is expressed, while *Coercive ultimatum* records its additional finality or bounded-compliance structure.

Code *Scarcity appeal* when the constraint instead consists in the limited continuing availability of an opportunity, offer, resource, or access condition rather than a prospective adverse consequence used as coercive leverage. For example, "Accept this offer by Friday or it expires" may instantiate scarcity, whereas "Sign by Friday or you are fired" instantiates a coercive ultimatum. Co-code scarcity and ultimatum only where disappearing opportunity and coercive sanction are independently present.

- **Collective punishment threats.** Threaten a prospective adverse consequence against a wider group or set of associated persons on

the basis of noncompliance, misconduct, or other conduct attributable to one or some members, using the distributed sanction as leverage to obtain, deter, maintain, or otherwise alter conduct (Heckathorn, 1990; Whitmeyer, 2002). *How to spot:* the available message or message sequence establishes (i) identifiable conduct associated with one or more target actors or group members; (ii) a prospective sanction or adverse consequence extending beyond the immediately responsible actor to additional members of an associated group, organisation, family, community, population, or other collective; and (iii) a recoverable contingency in which that distributed consequence functions as leverage for compliance or behavioural control. For example, "If anyone in this unit leaks another document, the entire unit loses leave" makes the conduct of individual members consequential for the wider group and uses the prospective collective sanction to induce compliance. The defining feature is not merely that a threatened consequence affects a person whom the recipient values. A threat such as "Testify and I will have your brother fired" may instantiate an explicit or implied coercive threat through third-party harm without constituting collective punishment. Code the present Tactic where responsibility and sanction are distributed such that persons or group members are prospectively penalised on account of conduct by one or some members (Heckathorn, 1990; Whitmeyer, 2002). Annotators do not determine whether the group is genuinely responsible for the conduct, whether the proposed sanction is lawful, fair, proportionate, credible, or likely to occur, whether the source privately intends to impose it, or whether recipients experience fear, guilt, solidarity, resentment, or other psychological reactions. Annotation records the observable use of distributed sanction liability as coercive leverage. *Target constructs: Agency/Choice*—the defining operation changes the practical conditions of choice by making the consequences of individual or subgroup conduct extend to additional members, thereby attaching costs to available responses both directly and potentially through group-mediated compliance pressure (Heckathorn, 1990; Whitmeyer, 2002); *Behavior*—the prospective collective sanction is used to elicit, deter, preserve, terminate, or otherwise alter identifiable individual or group conduct (Whitmeyer, 2002). *Boundary:* Do not infer an *Identity* target merely because the threat exploits family, organisational, communal, national, ethnic, religious, or other group relationships. Such relationships may determine the bearer or potency of the sanction without the communication attempting to alter self-concept or group identification. Likewise, fear, guilt, resentment, solidarity, or attentional capture are possible mechanisms or consequences, not fixed Target-Constructs. Code *Explicit threats* additionally when the distributed conduct–sanction relation is directly stated, and code *Implied threats /*

coercive intimidation additionally when that relation is recoverable but one or more components are conveyed indirectly. A second sanction is not required: the threat-form label records how the coercive relation is expressed, while *Collective punishment threats* records the distributed-liability structure. Do not code *Collective punishment threats* for every threat against an associated third party. Require a recoverable structure in which persons beyond the immediately responsible actor prospectively bear sanctions on account of conduct attributable to one or some members of the relevant group or collective.

B.8 Deception (form-level; no intent inference or fact-checking)

- **Exaggeration / overstatement.** Represent a quantity, magnitude, frequency, prevalence, likelihood, certainty, scope, effect, severity, importance, quality, or evidential or causal strength as more extreme than a comparison point, source characterisation, or evidential baseline recoverable from the message, thereby inflating the represented degree or strength of a claim or evaluation (Jowett and O'Donnell, 2018; Da San Martino et al., 2019; Sumner et al., 2014; Haber et al., 2018). *How to spot:* the message contains a recoverable scalar, quantitative, evaluative, modal, causal, or evidential relation in which a target is represented more extremely than an available comparison point or baseline. Observable forms may include upgrading limited or qualified evidence into a stronger or categorical claim, expanding a bounded quantity or frequency into a substantially broader one, or characterising a represented magnitude, effect, quality, severity, or importance in terms that materially exceed the baseline supplied by the message (e.g., "the study reports an association; this proves that X causes Y", "support rose from 48% to 51%—an overwhelming landslide", "several people complained; everyone is furious"). Superlatives, universal quantifiers, certainty terms, intensifiers, large numbers, or extreme adjectives do not by themselves license the Tactic. Annotators do not determine whether a claim is objectively false or exaggerated relative to external facts; the relevant inflation must be recoverable from the message or an explicitly represented source, comparison, or evidential baseline. Conventional hyperbole, humour, idiomatic emphasis, or obviously non-literal extreme expression is insufficient where it does not function to materially inflate the represented claim or evaluation. *Target constructs: Belief*—the defining operation inflates a represented quantity, magnitude, frequency, likelihood, certainty, scope, effect, causal relation, evidential strength, or other proposition beyond the recoverable baseline; *Attitude*—where the exaggerated dimension is evaluative, the defining operation inflates the represented positive or negative quality,

severity, importance, desirability, or undesirability of the focal target (Da San Martino et al., 2019; Sumner et al., 2014; Haber et al., 2018). *Boundary:* Do not code *Exaggeration / overstatement* merely because a message uses an intensifier, superlative, absolute expression, emotional wording, or conventional hyperbole. The excessive degree must be recoverable relative to a comparison point, source characterisation, or evidential baseline represented in the available message context. Code *Loaded language* when evaluatively or affectively charged lexical choice frames a target without a recoverable scalar or evidential inflation. Code *Causal oversimplification* when the defining operation restricts a complex causal account to a single, sufficient, or exhaustive cause. Code *False balance / distorted evidential parity* when the defining operation concerns the relative distribution of support, prevalence, credibility, consensus, or evidential standing across competing positions and represents a recoverable asymmetry as greater parity. Do not additionally code *Exaggeration / overstatement* merely because that parity distortion could be described abstractly as inflating the weaker side's standing; require a distinct recoverable scalar or evidential inflation independently satisfying the present Tactic. The Tactics may co-occur when their respective operations are independently evidenced.

- **Minimization / understatement.** Represent a quantity, magnitude, frequency, prevalence, likelihood, certainty, scope, effect, severity, harm, risk, cost, responsibility, importance, or evidential or causal strength as less extensive, serious, consequential, or significant than a comparison point, source characterisation, or evidential baseline recoverable from the message, thereby downscaling the represented degree or evaluative severity of a claim, event, condition, or outcome (Dimitrov et al., 2021; Kassin and McNall, 1991; Luke and Alceste, 2020). *How to spot:* the message contains a recoverable scalar, quantitative, evaluative, modal, causal, or evidential relation in which a target is represented less extremely or seriously than an available comparison point or baseline. Observable forms may include recasting a represented harm, violation, risk, cost, quantity, frequency, effect, or limitation as minor, negligible, isolated, merely technical, or otherwise less consequential; reducing the represented strength of evidence or causal relations; or using downtoning expressions such as "just", "only", "merely", "minor", or "a little" where the surrounding message establishes the downscaling relation (e.g., "three people were hospitalized, but it was just a minor incident", "the report documents 47 violations; these are only a handful of technical issues", "the insult was just a joke"). Downtoners, diminutives, mitigating adjectives, or low-intensity expressions do not by themselves license the Tactic. Annota-

tors do not determine whether a claim objectively understates external facts; the relevant minimization must be recoverable from the message or an explicitly represented source, comparison, or evidential baseline. *Target constructs: Belief*—the defining operation downscales a represented quantity, magnitude, frequency, likelihood, certainty, scope, effect, severity, harm, risk, responsibility, causal relation, evidential strength, or other propositional relation relative to the recoverable baseline; *Attitude*—where the minimised dimension is evaluative, the defining operation attenuates the represented seriousness, importance, blameworthiness, harmfulness, undesirability, or other negative evaluation of the focal target (Dimitrov et al., 2021; Kassin and McNall, 1991; Luke and Alceste, 2020). *Boundary:* Do not code *Minimization / understatement* merely because a message uses "just", "only", "minor", a diminutive, downtoning expression, euphemism, or other low-intensity wording. The reduced degree must be recoverable relative to a comparison point, source characterisation, or evidential baseline represented in the available message context. Code *Loaded language* when euphemistic or otherwise connotatively attenuating lexical choice frames a target without a recoverable scalar or evidential downscaling. Code *Salience-based concealment* when materially relevant qualifying information remains present in the communication but is made substantially less perceptually prominent or accessible than the focal representation. Code *Minimization / understatement* when the information is instead represented at a recoverably reduced degree, severity, importance, or consequence. The two may co-occur only where information is both perceptually suppressed and independently downscaled. Code *Strategic vagueness / equivocation* when the message leaves the relevant quantity, responsibility, commitment, or interpretation indeterminate without an independently recoverable minimisation relation. These Tactics may co-occur when their respective operations are independently evidenced.

- **Strategic vagueness / equivocation.** Leave a materially relevant proposition, referent, quantity, scope, agent, responsibility relation, condition, stance, or commitment underdetermined by using vague, ambiguous, indirect, evasive, or otherwise noncommittal formulation, so that multiple materially different interpretations or levels of commitment remain available where the message context makes greater determinacy relevant (Eisenberg, 1984; Bavelas et al., 1990; Solt, 2015). *How to spot:* the available message context establishes an informational question, contrast, commitment, responsibility relation, quantity, scope, or other matter for which greater specificity is communicatively relevant, while the focal wording leaves that matter materially unresolved. Observable forms may include an evasive response

that does not commit to the proposition at issue; an unspecified agent where responsibility is explicitly at issue; a vague quantity, population, condition, or scope where the surrounding message establishes the more determinate alternative; or wording that preserves materially different interpretations of what the source asserts or commits to (e.g., "Who authorised the payment?—The payment was authorised through the appropriate procedures"; "Will every customer receive 50% off?—Customers can save up to 50%"). Hedges, modal verbs, vague quantifiers, approximators, passive constructions, or ambiguous expressions do not by themselves license the Tactic. Annotators do not determine whether the statement is false, whether greater precision was objectively available, or whether the source privately intended to mislead; the relevant underspecification or equivocation must be recoverable from the message or message sequence itself. *Target constructs: Belief*—the defining operation prevents or weakens formation of a determinate representation concerning a materially relevant proposition, referent, quantity, scope, agent, responsibility relation, condition, stance, or commitment by leaving multiple interpretations or degrees of commitment open (Bavelas et al., 1990; Eisenberg, 1984). *Boundary:* Do not code *Strategic vagueness / equivocation* merely because a message contains uncertainty language, hedging, approximation, a vague adjective or quantifier, passive voice, or an ambiguous expression. Such forms may accurately communicate uncertainty, efficiently express imprecision, serve politeness or other ordinary communicative functions, and are not inherently deceptive. Code this Tactic only where the available message context makes a more determinate proposition, referent, quantity, scope, responsibility relation, condition, stance, or commitment materially relevant and the focal formulation leaves it unresolved. Code *Minimization / understatement* when a recoverable quantity, severity, risk, responsibility, or evaluative degree is represented as lower rather than merely left indeterminate. Code *Obfuscation (jargon/complexity)* when the defining operation makes message content difficult to process or comprehend rather than semantically or pragmatically underdetermined. Selective omission of otherwise clear information does not by itself license this Tactic. Code *Agenda shifting / evasive refocusing* when an established question, issue, criterion, comparison, actor, time frame, or other focal matter is displaced by a different agenda. Code *Strategic vagueness / equivocation* when the focal matter remains in place but the response leaves the relevant proposition, referent, quantity, agent, responsibility relation, stance, condition, or commitment materially underdetermined. The two may co-occur only where the message both shifts the focal agenda and independently leaves a materially relevant element underdetermined.

- **Selective evidence presentation (cherry-picking).** Select or report a systematically favourable subset of a broader set of evidence, results, cases, observations, examples, or relevant considerations that is recoverable from the available message context, while excluding or discounting countervailing members of that set, so that the presented subset provides a skewed evidential basis for a claim, conclusion, or evaluation (Chiu et al., 2017; Farina et al., 2026). *How to spot:* the message or recoverable message structure establishes a broader evidential or consideration set and selectively uses the members that support a favoured claim, conclusion, or evaluation while relevant countervailing members of the same set are omitted, excluded, or denied comparable evidential use. Observable forms may include reporting only favourable outcomes from a larger represented set of outcomes; citing only supportive studies, cases, polls, periods, subgroups, or observations from an explicitly recoverable pool; or drawing a conclusion from a favourable subset while contrary members of the represented evidence set are left out of the stated evidential basis. For example, a report may identify several favourable and unfavourable outcomes while its summary cites only the favourable outcomes as evidence that an intervention succeeded. A one-sided argument, advocacy message, list of supporting reasons, absence of a rebuttal, or failure to mention unspecified counterevidence does not by itself license the Tactic. Annotators do not search for external counterevidence, determine whether the selected evidence is factually correct, infer what evidence the communicator privately knew, or assume that omitted evidence exists. The broader evidence set and the selective relation must be recoverable from the message, message sequence, or explicitly represented source material available for annotation. *Target constructs: Belief*—the defining operation presents a systematically selected subset of a recoverable evidential set as the basis for accepting a claim, conclusion, relation, estimate, or interpretation, thereby providing a skewed representation of the evidential support available for that proposition; *Attitude*—where the selected subset consists of favourable or unfavourable attributes, outcomes, consequences, or considerations, the defining operation provides a selectively valenced evidential basis for evaluating the focal object, actor, issue, policy, behaviour, or outcome (Chiu et al., 2017; Farina et al., 2026). *Boundary:* Do not code *Selective evidence presentation (cherry-picking)* merely because a message presents arguments for only one side, advocates a position, lists supporting reasons, or fails to mention counterarguments whose existence would have to be supplied from external knowledge. The defining operation requires observable selection from a broader evidential or consideration set that is recoverable within the available annotation context. Code *Minimization / understatement* when countervailing information is represented

but its quantity, severity, importance, or evidential strength is downscaled. Code *Strategic vagueness / equivocation* when materially relevant information remains semantically or pragmatically underdetermined rather than selectively excluded. Code *Salience-based concealment* when materially relevant qualifying information remains present but presentation features make it substantially less perceptually accessible than the focal representation. Code *Selective evidence presentation (cherry-picking)* when the defining operation instead concerns which members of a broader recoverable evidential or consideration set are admitted to, excluded from, or given evidential use in support of the claim or evaluation. Perceptual suppression is not required for cherry-picking, and evidential subset selection is not required for salience-based concealment. Co-code only where both relations are independently evidenced. Code *False balance / distorted evidential parity* when positions or bodies of evidence with unequal represented support are presented as comparably supported. These Tactics may co-occur when their respective operations are independently evidenced.

- **False balance / distorted evidential parity.** Present positions, claims, sources, expert judgements, or bodies of evidence with comparable standing, support, prevalence, credibility, or representativeness despite a substantial asymmetry in their relative support that is recoverable from the available message context, thereby representing a materially unequal evidential or consensus distribution as more evenly divided than the stated baseline warrants (Boykoff and Boykoff, 2004; Koehler, 2016; Dixon and Clarke, 2013). *How to spot:* the message or recoverable message structure establishes that two or more positions differ materially in represented evidential support, expert endorsement, prevalence, or consensus, but subsequently presents those positions as comparably supported, representative, credible, prevalent, or divided. Observable forms may include describing a strongly asymmetric distribution as an even or substantial division; giving positions explicitly represented as majority and fringe views equivalent standing as the two comparable sides of a question; or using symmetric presentation to communicate parity despite a support imbalance established elsewhere in the available message context (e.g., "94 of 100 experts support X, while six oppose it; experts remain evenly divided", "18 studies support X and two do not—the research is split on the issue"). Merely presenting two opposing views, quoting one source from each side, reporting disagreement, or using a balanced format does not by itself license the Tactic. Annotators do not determine from external knowledge which position is correct, mainstream, fringe, or better supported; the relevant asymmetry and the parity representation must be recoverable from the message, message sequence, or explicitly represented

source material available for annotation. Annotation does not imply that the communicator knowingly created the distorted parity or intended to mislead; the Tactic is licensed by the recoverable relation between the represented asymmetry and the subsequent parity representation. *Target constructs: Belief*—the defining operation distorts the represented relative support, prevalence, consensus, credibility, or evidential standing of competing positions, making a materially asymmetric distribution appear more evenly divided or comparably supported than the recoverable baseline (Koehler, 2016; Dixon and Clarke, 2013) *Boundary:* Do not code *False balance / distorted evidential parity* merely because a message reports multiple viewpoints, presents opposing sources, uses a balanced format, or states that disagreement exists. The available annotation context must establish a material asymmetry in relative support, prevalence, consensus, or evidential standing and a subsequent presentation that communicates greater parity than that baseline. Code *Selective evidence presentation (cherry-picking)* when the defining operation selects a systematically favourable subset from a broader recoverable evidential set. Code *Two-sided with refutation* when an opposing argument is acknowledged and substantively rebutted rather than presented with misleading parity. Equating cases, actors, actions, or outcomes that differ in materially relevant respects does not by itself license this Tactic; *False balance / distorted evidential parity* specifically requires a recoverable asymmetry in support, prevalence, consensus, credibility, or evidential standing that is presented as more nearly balanced. Code *Doubt-seeding* when uncertainty, disagreement, evidential limitations, possible error, source-reliability concerns, or procedural limitations function as grounds for reducing confidence in or suspending judgement toward an identifiable epistemic target. Code *False balance / distorted evidential parity* when the defining operation specifically represents a recoverably asymmetric distribution of support, prevalence, consensus, credibility, or evidential standing as more nearly balanced or equivalent.

A false-balance representation does not automatically license *Doubt-seeding* merely because distorted parity may produce uncertainty as a downstream consequence. Co-code only where the represented parity is additionally deployed as an independent reason for reducing confidence or suspending judgement.

- **False equivalence / distorted comparability.** Represent two or more cases, actors, actions, policies, events, outcomes, conditions, or other comparands as equivalent, comparably significant, or sufficiently alike on a relevant dimension despite a material difference on that dimension that is recoverable from the available message context, so that the comparison suppresses or overrides differential weighting war-

ranted by the represented baseline (Piskorski et al., 2025; Walton and Hyra, 2018; Shorten, 2011). *How to spot:* the available message or recoverable message sequence must provide (i) two or more identifiable comparands, (ii) an explicit or recoverable dimension on which equivalence, parity, comparable significance, or sufficiently close similarity is communicated, and (iii) a materially relevant difference on that same dimension that is recoverable from the available annotation context but is subsequently ignored, collapsed, or overridden by the equivalence claim. Observable forms may include statements such as "both sides are equally responsible", "these are really the same thing", "there is no meaningful difference between them", or analogical reasoning in which one shared characteristic is used to transfer a broader judgement despite an available material dissimilarity. For example, "one protest involved a single broken window; the armed attack killed twelve people, but both sides were equally violent" supplies a recoverable difference in represented magnitude and then communicates parity on the comparison dimension. Likewise, "Candidate A misstated one date; Candidate B fabricated several financial records; both candidates are equally dishonest" supplies differential conduct while subsequently equalising the represented evaluative standing of the actors. Mere comparison, similarity, analogy, parallel wording, symmetrical presentation, or assertion that two objects share a characteristic does not by itself license the Tactic. Annotators do not use external factual, historical, legal, scientific, or moral knowledge to decide that a comparison is inappropriate. If the material difference relevant to the asserted equivalence is not recoverable from the available annotation context, do not infer *False equivalence / distorted comparability* merely because the comparison appears implausible, controversial, unfair, or weak. Annotation does not imply that the communicator knowingly distorted the comparison, intended to deceive, or acted in bad faith. The Tactic is licensed by the recoverable relation between the represented material difference and the communicated equivalence or parity. *Target constructs: Belief*—the defining operation alters the represented comparative relation among identifiable cases, actors, actions, policies, events, outcomes, conditions, or other comparands by presenting materially different items as equivalent, comparably significant, or sufficiently alike on the relevant dimension despite a recoverable difference bearing on that comparison (Walton and Hyra, 2018; Piskorski et al., 2025); *Attitude*—where the comparison concerns evaluative standing, the same equivalence operation positions materially differentiated comparands as deserving similar approval, disapproval, blame, praise, legitimacy, seriousness, importance, merit, harmfulness, or other evaluative weight. Political uses of moral equivalence provide a prominent instance in which comparison redistributes or collapses differential judgements of wrongdoing

(Shorten, 2011). *Boundary:* Code *False equivalence / distorted comparability* only where a material difference relevant to the communicated equivalence is recoverable from the available annotation context. Do not adjudicate the quality of a comparison using external knowledge. Code *False balance / distorted evidential parity* when the misrepresented equivalence specifically concerns relative evidential support, expert endorsement, prevalence, consensus, source credibility, or comparable epistemic standing across competing positions or bodies of evidence. Code *False equivalence / distorted comparability* when the parity instead concerns cases, actors, actions, policies, events, outcomes, harms, responsibility, seriousness, merit, legitimacy, or another non-evidential comparison dimension. Co-code only where independent evidential-parity and broader comparability distortions are both recoverable. Code *Metaphor / analogy* when relational structure from a source or comparison domain is mapped onto a target in order to organise interpretation or inference. An analogy does not become a false equivalence merely because the compared cases differ. Code both only where the analogical mapping independently performs its framing operation and the available context additionally establishes a materially relevant difference that is overridden by a communicated equivalence. Code *Exaggeration / overstatement* or *Minimization / understatement* when the defining distortion concerns the represented degree of one target relative to a recoverable scalar or evaluative baseline. *False equivalence / distorted comparability* instead requires a relational parity or equivalence across two or more comparands. The Tactics may co-occur where an independent scalar distortion is used to produce or support the distorted equivalence. Code *Whataboutism* when a criticism, question, or accusation is deflected through a counteraccusation or comparison that shifts attention away from the original issue. A whataboutist comparison does not automatically constitute false equivalence; co-code only where the countercomparison additionally equates materially differentiated cases on a dimension for which the relevant difference is recoverable.

- **Salience-based concealment.** Present materially qualifying, limiting, costly, adverse, or otherwise decision-relevant information within the communication while substantially reducing its perceptual prominence relative to the focal claim, offer, representation, or term that it qualifies, so that the focal representation is substantially easier to notice than information necessary for interpreting it accurately (Gray et al., 2018; Mathur et al., 2019; Luguri and Strahilevitz, 2021). *How to spot:* the message contains both an identifiable focal representation and information that materially qualifies, limits, conditions, or otherwise bears on that representation, while presentation features make

the qualifying information markedly less noticeable or perceptually accessible. Observable forms may include disproportionate differences in size, contrast, placement, duration, volume, rate, layering, formatting, or comparable presentation features; relegating a consequential qualification to fine print or an inconspicuous location; or otherwise presenting relevant information as though it were perceptually secondary or irrelevant (e.g., a highly prominent "PUBLIC SUPPORT: 72%" claim paired with a markedly inconspicuous qualification that the figure refers only to respondents already supporting the policy; or a highly prominent "FREE FOR 30 DAYS" offer paired with a markedly inconspicuous "renews automatically at €49.99/month" disclosure). The relevant cue is the relation between the focal representation, the qualifying information, and their asymmetric perceptual salience. Small text, muted colour, peripheral placement, rapid speech, low volume, or another low-salience feature does not by itself license the Tactic. Annotators do not determine whether either statement is objectively true, whether the qualification legally satisfies a disclosure requirement, whether the recipient actually notices it, or whether the communicator privately intended concealment. *Target constructs: Perception/Attention*—the defining operation makes materially relevant qualifying information substantially less perceptually salient or accessible than the focal representation to which it relates, thereby structuring what is most readily noticed or attended; *Belief*—the suppressed information materially qualifies the proposition, terms, scope, cost, risk, condition, or other representation communicated by the more salient content, so that the attentional asymmetry bears directly on formation of an adequate representation of the focal matter (Gray et al., 2018; Luguri and Strahilevitz, 2021). *Boundary:* Do not code *Salience-based concealment* merely because information is visually or auditorily less prominent than surrounding content. The available message must establish materially relevant information that qualifies or bears on a more salient focal representation and a marked perceptual asymmetry that makes that information disproportionately difficult to notice. Generic omission of information not present in the available communication does not license this Tactic. Code *Manipulative choice architecture* when the defining asymmetry concerns the practical accessibility, effort, default status, sequencing, or reversibility of response alternatives rather than the perceptual accessibility of information. Code *Obfuscation (jargon/complexity)* when information is available to attention but made difficult to understand or process. Code an applicable Distraction Tactic when competing or tangential material redirects attention away from the focal issue rather than the relevant qualification itself being perceptually suppressed. These Tactics may co-occur when their respective operations are independently

evidenced.

B.9 Distraction

- **Whataboutism.** Respond to criticism, accusation, questioning, or scrutiny concerning a focal issue by introducing a comparative or counteraccusatory case—typically another wrongdoing, inconsistency, neglected harm, or allegedly similar case—that both redirects attention from the focal issue and functions as a reason to qualify, suspend, weaken, or reconsider the judgement being advanced about it (Bowell, 2023; Aikin and Casey, 2026; Phi et al., 2024). *How to spot:* the available message or message sequence establishes an identifiable criticism, accusation, question, or issue under scrutiny and a response that introduces another actor, case, wrongdoing, inconsistency, or neglected concern as a comparative counterconsideration. The countercase functions not merely as a new topic but as a reason that the focal criticism or judgement should be regarded as selective, inconsistent, comparatively incomplete, insufficiently contextualised, or otherwise in need of qualification (e.g., "You say Country A should be sanctioned for violating this treaty. What about Country B? You acknowledged that it violated the same provision, yet you opposed sanctions there."). The response need not contain the words "what about", and a "what about" question does not by itself license the Tactic. Annotators identify the comparative and deflective relation communicated by the message; they do not determine whether either allegation is factually correct, whether the cases are objectively equivalent, or whether the criticism is ultimately justified. *Target constructs: Perception/Attention*—the defining operation redirects argumentative and attentional focus from the focal issue under scrutiny toward another case, wrongdoing, inconsistency, actor, or neglected concern; *Belief*—the introduced countercase functions as a comparative or meta-argumentative reason to qualify, suspend, weaken, or reconsider the focal criticism, attribution, judgement, or conclusion by representing it as selectively applied, inconsistent, comparatively incomplete, or insufficiently contextualised (Bowell, 2023; Aikin and Casey, 2026). *Boundary:* Do not code *Whataboutism* merely because a message introduces a comparison, identifies differential treatment, raises another case, or uses the expression "what about". A comparison may be directly relevant to evaluating the general principle, evidential claim, policy criterion, or comparison set explicitly at issue. Code *Whataboutism* when a comparative or counteraccusatory case is used both to redirect scrutiny and to challenge the adequacy, consistency, or unqualified application of the focal judgement. Code *Red herring* when a tangential consideration diverts attention without functioning as this kind of comparative counterconsideration. A

tu quoque form may occur when the countercase specifically alleges inconsistency or hypocrisy by the critic, but whataboutism may instead concern another actor, case, or neglected comparison.

- **Red herring.** Introduce a consideration, issue, controversy, claim, event, or line of argument that diverts attention from an identifiable focal question, claim, criticism, or issue under discussion without substantively resolving or bearing on that focal issue in the role for which it is introduced (Walton, 2004; Walton, 2005). *How to spot:* the available message or message sequence establishes an identifiable issue, question, criticism, claim, or argumentative target and introduces another consideration that displaces engagement with that focal matter while failing to address the issue it replaces. The distracting consideration need not be semantically unrelated to the broader topic; red herrings may concern the same actor, institution, policy area, or controversy while remaining irrelevant to the particular issue requiring resolution (e.g., "Why was the contract awarded without an open tender?"—"This ministry has opened forty new health clinics this year"). The defining cue is the relation between the focal issue and the diversionary consideration, not the presence of an abrupt topic change, emotional language, or an obviously unrelated statement. Annotators assess the communicative relevance relation recoverable from the message context and do not determine whether the distracting claim is factually true or whether the communicator privately intended diversion. *Target constructs: Perception/Attention*—the defining operation redirects argumentative and attentional focus from the issue that the discourse has made focal toward a consideration that displaces engagement with that issue (Walton, 2004; Walton, 2005). The diversion may protect or alter evaluation of an actor, policy, or issue by shifting scrutiny toward more favourable material, but such evaluative influence is instance-dependent rather than constitutive of the Red-herring operation; an independently evidenced evaluative Tactic may additionally be coded. *Boundary:* Do not code *Red herring* merely because a response introduces new information, changes the level of analysis, supplies contextual information, or discusses another aspect of the same broad topic. The introduced consideration must function as a diversion from an identifiable focal issue rather than as evidence, explanation, qualification, rebuttal, or other substantively relevant engagement with it. Code *Whataboutism* when the diversion characteristically takes the form of a comparative or counteraccusatory case that also functions as a reason to qualify, suspend, weaken, or reconsider the focal judgement. Code *Agenda shifting / evasive refocusing* when the response replaces or refocuses the issue, question, dimension, criterion, comparison, actor, time frame, or other agenda to be resolved rather than principally introducing a particular

diversionary consideration as putative engagement with the established focal matter. Do not additionally code *Agenda shifting / evasive refocusing* merely because a *Red herring* succeeds in redirecting discourse; require an independently recoverable substitution or refocusing of the agenda itself. Code *Straw man* when the defining operation instead misrepresents the position or argument under discussion and redirects engagement toward that distorted representation. These Tactics may co-occur when their respective operations are independently evidenced.

- **Straw man.** Represent an identifiable opponent's claim, position, proposal, argument, or commitment as a materially different position and direct criticism, rebuttal, dismissal, or attack toward that substitute representation rather than the position recoverable from the available discourse (Walton, 1996; Aikin and Casey, 2018; Oswald and Lewiński, 2014). *How to spot:* the available message, message sequence, or explicitly represented source material establishes (i) an identifiable position, argument, proposal, or commitment attributable to another actor; (ii) a materially different representation of that position attributed to the same actor; and (iii) criticism, rebuttal, dismissal, or attack directed toward the substitute representation (e.g., "I support requiring background checks for private firearm sales"—"So my opponent wants the government to confiscate everyone's guns"). The defining cue is the recoverable relation between the original position, its materially altered representation, and the subsequent attack. A phrase such as "so you're saying", paraphrase, simplification, unfavourable interpretation, or disagreement does not by itself license the Tactic. Annotators do not determine from external knowledge what an actor truly believes or intended to say; the relevant source position and its distortion must be recoverable within the available annotation context. *Target constructs: Belief*—the defining operation supplies a distorted representation of what another actor claims, proposes, believes, argues, or is committed to, thereby altering the propositional representation of that actor's position; *Perception/Attention*—the defining substitution redirects argumentative scrutiny from the recoverable position toward a materially different surrogate position that becomes the object of criticism or refutation (Walton, 1996; Aikin and Casey, 2018; Oswald and Lewiński, 2014). *Boundary:* Do not code *Straw man* merely because a speaker disagrees with, summarises, paraphrases, simplifies, interprets, criticises, or uncharitably describes another position. The available annotation context must make both the relevant source position and a material distortion of that position recoverable, and the distorted representation must function as the object of criticism, rebuttal, dismissal, or attack. An attributed position whose inaccuracy could be established only through external knowledge does not license this Tac-

tic. Code *Red herring* when engagement is diverted toward another consideration without materially substituting the opponent's position. Code *Whataboutism* when scrutiny is redirected through a comparative or counteraccusatory case used to challenge the adequacy or consistency of the focal judgement. An applicable Deception or Framing Tactic may additionally be coded where the substitute representation independently inflates, minimises, labels, or otherwise transforms the source position.

- **Obfuscation (jargon/complexity).** Present focal information, relations, explanations, or responses through specialised terminology, abstraction, syntactic complexity, nominalisation, technical density, or other linguistically opaque formulation that, in the available communicative context, materially increases the effort required to recover, comprehend, evaluate, or scrutinise the communicated content (Supran and Oreskes, 2017; Brunton and Nissenbaum, 2015; Markowitz and Hancock, 2016; Bullock et al., 2019; Shulman et al., 2020). *How to spot:* the available message context establishes identifiable focal information, a proposition, explanation, relation, answer, or other communicative content, while its linguistic formulation makes that content materially difficult to recover or scrutinise through features such as unexplained specialist terminology, dense technical or bureaucratic language, chains of nominalisation, excessive abstraction, syntactic embedding, or similarly opaque formulation. For example, in response to "Is the government raising the retirement age?", "the reform introduces a recalibration of age-linked eligibility parameters within the contributory pension framework" renders the focal relation less directly accessible than the corresponding plain-language formulation. Jargon, technical terminology, abstraction, formal language, or long sentences do not by themselves license the Tactic: such language may be appropriate or necessary for specialist audiences or intrinsically complex subject matter. The defining evidence is that the linguistic packaging itself materially burdens access to or scrutiny of focal content in the available communicative context. Annotators do not infer that the communicator privately intended confusion, that a simpler formulation was objectively available, or that a particular recipient actually failed to understand the message. *Target constructs: Perception/Attention*—the defining operation increases the processing burden associated with recovering and scrutinising focal message content, consuming limited processing capacity and reducing its cognitive accessibility through linguistically complex or opaque presentation (Markowitz and Hancock, 2016; Bullock et al., 2019; Shulman et al., 2020); *Agency/Choice*—the same opacity impairs epistemic self-direction by making the information materially harder to comprehend, evaluate, and scrutinise

independently, thereby constraining the recipient's capacity for self-directed judgement (Proust, 2008; Bullock et al., 2019; Shulman et al., 2020). *Boundary:* Do not code *Obfuscation (jargon/complexity)* merely because a message contains technical vocabulary, specialist concepts, abbreviations, abstract terminology, formal language, long sentences, or complex subject matter. The linguistic presentation itself must materially increase the difficulty of recovering, comprehending, or scrutinising focal content in the available communicative context. Code *Strategic vagueness / equivocation* when the defining operation leaves materially relevant content semantically or pragmatically underdetermined rather than making determinate content difficult to process. Code *Salience-based concealment* when relevant information is present but made disproportionately difficult to notice rather than difficult to understand. Code *Information flooding / overload* when processing is burdened principally through the volume or accumulation of tangential material rather than through the linguistic opacity of focal content. These Tactics may co-occur when their respective operations are independently evidenced.

- **Agenda shifting / evasive refocusing.** Respond to an identifiable question, criticism, issue, or point of scrutiny by replacing or refocusing the focal issue to be addressed—for example, by shifting its relevant time frame, actor, level of analysis, criterion, comparison, or substantive question—so that discourse proceeds on a different agenda rather than resolving the originally established one (Greatbatch, 1986; Rasiah, 2010; Gabrielsen et al., 2020). *How to spot:* the available message or message sequence establishes an identifiable question, criticism, issue, or dimension requiring engagement, and the response substitutes another issue, dimension, comparison, criterion, or question as the focal matter to be addressed. The shift may remain within the same broad topic: observable forms include changing the relevant time period, actor, level of analysis, comparison point, evaluative criterion, or question (e.g., "Why did unemployment rise during the last quarter?"—"The important thing is that employment is higher today than when we entered government."). The defining cue is not merely introduction of new information or ordinary topic development, but a recoverable displacement of the agenda established by the preceding discourse. Annotators identify the relation between the recoverable focal issue and its substitute; they do not infer that the original issue was objectively sensitive, that the communicator privately intended evasion, or that a particular audience failed to notice the shift. *Target constructs: Perception/Attention*—the defining operation redirects argumentative and attentional focus by replacing or refocusing the question, issue, dimension, comparison, or criterion under scrutiny with another fo-

cal agenda (Greatbatch, 1986; Rasiah, 2010; Gabrielsen et al., 2020). *Boundary:* Do not code *Agenda shifting / evasive refocusing* merely because discourse moves naturally to a related topic, introduces additional context, changes the level of analysis for a substantively relevant reason, or answers only part of a compound question. The available context must establish a focal issue or dimension whose resolution is displaced by a substitute agenda. Code *Red herring* when diversion is accomplished principally by introducing a tangential consideration that displaces engagement with the focal issue without itself redefining the issue to be resolved. Code *Whataboutism* when the substitute focus takes the form of a comparative or counteraccusatory case that also functions as a reason to qualify, suspend, weaken, or reconsider the focal judgement. Code *Strategic vagueness / equivocation* when the focal issue remains in place but the response leaves the relevant proposition, responsibility, quantity, stance, or commitment materially underdetermined. These Tactics may co-occur when their respective operations are independently evidenced.

- **Information flooding / overload.** Saturate an available message or recoverable message stream with a high volume or dense accumulation of secondary, tangential, repetitive, weakly prioritised, or otherwise competing content that materially increases the processing and selection burden required to isolate, prioritise, or scrutinise focal information or an issue under attention (King et al., 2017; Cirone and Hobbs, 2023; Stukal et al., 2022; Persson, 2018). *How to spot:* the available message or message sequence contains an identifiable focal issue, claim, question, event, or other information target together with an unusually dense accumulation or high-volume stream of additional claims, details, examples, statistics, topics, posts, links, or other informational items that collectively compete with the focal material and substantially increase the effort required to identify what is relevant, important, or responsive. The defining cue is the accumulation or volume itself rather than any single distracting item. Individual components may be clear, factually compatible, or broadly related to the surrounding topic; they need not each be independently irrelevant. For example, in response to a question about why a ministry withheld a report, a speaker may produce an extended succession of statistics, achievements, programs, historical comparisons, and ancillary facts that leaves the focal withholding question buried among competing material. Likewise, where a recoverable message stream is available, repeated injection of innocuous, entertainment, or unrelated content may instantiate flooding when its volume visibly displaces focal political material. A long message, detailed explanation, large evidential record, or high posting rate does not by itself license the Tactic: the available annotation context must

show that accumulation of competing material itself performs the distracting or overloading operation. Annotators do not infer that a recipient actually became confused or overloaded, that the communicator privately intended distraction, or that any individual item is factually false (King et al., 2017; Cirone and Hobbs, 2023; Stukal et al., 2022). *Target constructs: Perception/Attention*—the defining accumulation introduces competing informational demands that disperse or capture limited attention and reduce the relative accessibility or salience of focal information or the issue under scrutiny (King et al., 2017; Cirone and Hobbs, 2023; Stukal et al., 2022); *Agency/Choice*—the same accumulation increases the epistemic selection and processing burden required to distinguish, prioritise, compare, and scrutinise information, thereby constraining the recipient's capacity for self-directed evaluation of the material at issue (Persson, 2018; Stukal et al., 2022). *Boundary:* Do not code *Information flooding / overload* merely because a message is long, information-rich, repetitive, highly detailed, or contains numerous pieces of evidence. The volume or accumulation must itself create competing informational demands around identifiable focal content or an issue under attention. Code *Red herring* when diversion is accomplished principally through a particular tangential consideration rather than through the cumulative burden of numerous competing items. Code *Agenda shifting / evasive refocusing* when the defining operation replaces or refocuses the question or issue to be resolved rather than saturating attention with multiple competing informational items. Code *Obfuscation (jargon/complexity)* when focal content is made difficult to comprehend principally through linguistic opacity or complexity rather than informational volume. Code *Doubt-seeding* when the defining operation undermines confidence in evidence, sources, consensus, or epistemic procedures rather than burdening information selection through volume. A high-volume propaganda or "firehose" strategy may instantiate *Information flooding / overload* when the relevant volume and competing-content structure are recoverable within the available annotation context; falsity, inconsistency, or source coordination are not required for this Tactic and are not inferred by annotators. These Tactics may co-occur when their respective operations are independently evidenced.

- **Humor as deflection.** Respond to an identifiable question, criticism, accusation, challenge, or other point of scrutiny through humour, irony, sarcasm, playful absurdity, joking, or another recognizably non-serious mode that both displaces substantive engagement with the focal issue and recasts the evaluative stance in which that issue, criticism, actor, or interaction is treated (Carpenter et al., 2019; Nuolijärvi and Tiittula, 2011; Chernobrov, 2024). *How to spot:* the available message

or message sequence establishes an identifiable question, criticism, accusation, challenge, or issue requiring engagement, and the response shifts into a humorous, ironic, sarcastic, playfully exaggerated, absurd, or otherwise recognizably joking mode that functions in place of, or materially displaces, substantive engagement with that focal matter. For example, in response to "Will you resign over the procurement failure?", "I was hoping you'd start with an easy question" uses a humorous turn rather than directly engaging with the question. Observable indicators of humorous mode may include wordplay, incongruity, comic exaggeration, ironic reversal, sarcasm, playful absurdity, laughter markers, or an explicit joking frame, but none of these features by itself licenses the Tactic. The defining cue is the relation between a recoverable focal issue and humour used as the vehicle of deflection and evaluative repositioning. Annotators do not determine whether the communicator privately intended humour, whether a particular audience actually found the response funny, whether the communicator became more likable, or whether the humorous response successfully reduced scrutiny (Carpenter et al., 2019; Nuolijärvi and Tiittula, 2011). *Target constructs: Perception/Attention*—the defining operation redirects attentional and interactional focus away from substantive engagement with the focal question, criticism, accusation, challenge, or issue by shifting the response into a humorous or non-serious mode (Carpenter et al., 2019; Nuolijärvi and Tiittula, 2011; Chernobrov, 2024); *Attitude*—the same humorous shift changes the evaluative stance in which the focal issue, criticism, actor, or interaction is treated, for example rendering it more playful, ridiculous, dismissible, contemptible, absurd, or otherwise differently valenced than under direct serious engagement (Nuolijärvi and Tiittula, 2011; Chernobrov, 2024). *Boundary:* Do not code *Humor as deflection* merely because a message contains a joke, irony, sarcasm, parody, ridicule, wordplay, comic exaggeration, or other humorous material. Humor may directly communicate an argument, criticism, evaluation, identity claim, or political message without deflecting from anything. The available context must establish a focal question, criticism, accusation, challenge, or issue whose substantive engagement is displaced through the humorous turn. The evaluative shift need not be favourable or reduce perceived seriousness: humour may trivialise, ridicule, mock, soften, intensify contempt, or otherwise reposition the focal issue or actor. Code *Agenda shifting / evasive refocusing* when the defining operation replaces or refocuses the issue to be resolved with another substantive agenda. Code *Red herring* when diversion is accomplished principally through a particular tangential consideration. Code *Whataboutism* when a comparative or counteraccusatory case redirects scrutiny and functions as a reason to qualify, suspend, weaken, or reconsider the focal judgement. Code *Ridicule*

/ *mockery* when humour independently functions to render an identifiable target laughable, foolish, absurd, contemptible, or otherwise worthy of derision without requiring the deflection structure specified here. Code *Ad hominem / personal-credibility attack*, *Name-calling / labeling*, or another applicable evaluative Tactic where their respective operations are also independently satisfied. An explicit "just joking" or equivalent non-serious framing may instantiate *Humor as deflection* when a recoverable prior statement or action is under scrutiny and the joking frame functions to displace substantive engagement with that scrutiny. These Tactics may co-occur when their respective operations are independently evidenced.

- **Ad hominem / personal-credibility attack.** Respond to a claim, argument, criticism, accusation, testimony, question, or other contribution by negatively characterising the person advancing it—for example, their character, motives, trustworthiness, competence, judgement, affiliations, circumstances, conduct, or consistency—and use that person-directed characterisation as a reason to discount the contribution, diminish the person's standing as a source or critic, or displace engagement with the substantive issue (Walton, 1998; Walton, 2008; Aikin, 2025). *How to spot:* the available message or message sequence establishes (i) an identifiable claim, argument, criticism, accusation, testimony, question, or other contribution associated with a person or actor; (ii) a negative characterisation of that person's character, motives, trustworthiness, competence, judgement, affiliations, circumstances, conduct, or consistency; and (iii) a recoverable relation in which that characterisation is used to weaken the person's credibility or standing, discount their contribution, or substitute person-directed scrutiny for engagement with its substantive merits. For example, in response to "Why did your ministry disregard the audit's findings?", "You've spent years campaigning against this government; nobody should take your accusations seriously" redirects scrutiny toward the questioner's alleged partisanship and uses it to undermine the standing of the accusation. Insults, negative descriptions, allegations of bias, references to affiliation, or claims of hypocrisy do not by themselves license the Tactic unless they perform this person-to-contribution undercutting or diversionary function. Annotators identify the represented relation and do not determine whether the personal allegation is factually correct, whether the person is actually biased or untrustworthy, whether the personal consideration is logically relevant, or whether the resulting credibility judgement is justified (Walton, 2008; Aikin, 2025). *Target constructs: Perception/Attention*—the defining operation redirects argumentative or interactional scrutiny from the substantive merits of a claim, criticism, question, testimony, or argument toward charac-

teristics or circumstances of the person advancing it (Walton, 2008; Walton, 1998); *Belief*—the person-directed attack functions as a reason to reduce acceptance, credibility, evidential weight, or argumentative standing of the claim, criticism, testimony, or argument associated with that person (Aikin, 2025; Walton, 2008); *Attitude*—the defining person-directed characterisation positions the target toward negative evaluation, for example as dishonest, biased, incompetent, hypocritical, corrupt, malicious, irrational, or otherwise untrustworthy or discreditable (Walton, 2008; Walton, 1998). *Boundary:* Do not code *Ad hominem / personal-credibility attack* merely because a message insults, criticises, negatively describes, ridicules, or assigns an unfavourable label to a person. The negative personal characterisation must be used to discount an identifiable contribution, undermine the person's standing as a source or critic, or divert substantive engagement toward the person. Code *Name-calling / labeling* when a condensed evaluative label characterises a person without independently performing this person-to-contribution undercutting relation. Code *Red herring* when scrutiny is diverted through another consideration rather than through characteristics or circumstances of the person advancing the focal contribution. Code *Whataboutism* when a comparative or counteraccusatory case is introduced to redirect scrutiny and qualify or reconsider the focal judgement; a *tu quoque* or hypocrisy-based response may instantiate both Tactics when the critic's inconsistency simultaneously functions as the comparative countercase and as grounds for discounting the critic's standing. Code *Humor as deflection* additionally when humour or ridicule independently performs the deflective and evaluative operation. Annotation of this Tactic does not entail that the person-directed argument is logically fallacious: information about competence, conflicts of interest, reliability, or other source characteristics may sometimes be relevant to credibility assessment. Annotators record the communicated undercutting structure rather than adjudicating its factual accuracy or argumentative validity. These Tactics may co-occur when their respective operations are independently evidenced.

C Narrative devices and archetypes

C.1 Narrative devices palette (for reference)

Narrative devices are micro-structural signals that let coders recognise narrative packaging within a single message. They are operationalised as cues and may be used alongside tactic coding (Section B).

Note. Supplementary Appendix C distinguishes independently assigned *NarrativeDevices* from deterministically derived *NarrativeArchetypes*. Coders

identify NarrativeDevices and, where applicable, populate their closed-vocabulary attributes from message-internal evidence. NarrativeArchetypes are not independently selected; they are derived when assigned NarrativeDevice instances, their applicable attributes, and their configuration memberships satisfy a minimal signature specified below.

Cross-dimension boundary with Method. The Narrative and Method dimensions are analytically independent. NarrativeDevice assignment records how actors, events, relations, temporal structure, causal structure, moral structure, symbols, or stakes contribute to the message's narrative organisation. Tactic assignment in the Method dimension records a communicative influence operation satisfying the corresponding definition and spotting rule in Supplementary Appendix B. Shared lexical or representational material does not by itself license annotation in both dimensions.

Accordingly, role-casting does not imply *Altercasting* unless a role or identity position is assigned or activated for the recipient and its behavioural expectation functions as an influence warrant; causal emplotment does not imply *Causal oversimplification* unless a causal account is restricted in the manner specified by that Tactic; moral valuation does not imply *Moral framing* unless a moral criterion functions as an interpretive or evaluative frame; metaphorical mapping does not imply *Metaphor / analogy* unless the cross-domain relation independently performs the interpretive or inferential influence operation specified there; and stakes or urgency do not imply *Appeal to fear* unless represented harm functions as a reason for an identifiable protective response. Conversely, a Method Tactic does not automatically license a NarrativeDevice merely because it contains narrative-compatible material. Co-code across dimensions only where the respective operations are independently recoverable.

C.2 Narrative device attributes (coding fields)

Narrative devices may additionally carry a small set of attributes that capture who is doing what to whom and, where applicable, the relevant temporal, moral, metaphorical, causal, or stakes configuration. Together with NarrativeDevice instances and their configuration memberships, these attributes provide the inputs from which qualifying NarrativeArchetypes are derived deterministically.

Supplementary Table S4: Narrative devices palette.

Device	**Definition & anchors**
Role-casting (actants)	Assigns actors to narrative roles (e.g. hero/protector, villain/opponent, victim, traitor, savior/strong leader), enabling plot expectations (Propp, 2010; Greimas, 1966; Roselle et al., 2014).
Conflict–resolution arc	Minimal story grammar: situation → problem → struggle → (promised) resolution/call-to-action (Bruner, 1990; Fisher, 1984; Green and Brock, 2000).
Temporal sequencing & contrast	Before/after, decline/restoration, crisis/solution orderings that scaffold coherence (Ricœur, 1984; Linde, 1993).
Causal emplotment	Links events into cause–effect chains (who did what to whom, why it matters) (White, 1973; Ricœur, 1984; Entman, 1993).
Metaphorical mapping	Uses cross-domain metaphorical structure to organise narrative actors, events, causal relations, problems, or stakes, including war, invasion, contagion, machine, or comparable mappings (Lakoff, 2004; Thibodeau and Boroditsky, 2011).
Exemplars & vignettes	Concrete cases that stand in for general claims, heightening vividness (Iyengar, 1991; Zillmann, 1999).
Moral valuation	Organizes actors, actions, events, or outcomes within the narrative through morally valenced plot relations such as virtue/vice, purity/degradation, loyalty/betrayal, or fairness/cheating (Entman, 1993; Haidt, 2012).
Symbolic condensation	Icons, slogans, emblems that compress complex stories (Edelman, 1988; Hariman and Lucaites, 2007).
Stakes & urgency cues	Represents what is at risk in the narrative as severe, escalating, time-critical, catastrophic, or existential, including through dread, emergency, tipping-point, or last-chance language (Tannenbaum et al., 2015; Moffitt, 2016; Brady et al., 2017).

Supplementary Table S5: Narrative device attributes and allowed values.

Attribute group	**Field**	**Allowed values**
Role-casting (actants)	`ActorGroup`	`elite_in_group`, `people_in_group`, `minority_in_group`, `external_outgroup`, `abstract_cabal`, `institution`
	`ActorLocation`	`internal`, `external`
	`Function`	`hero/protector`, `villain/opponent`, `victim`, `traitor`, `savior/strong_leader`
	`ReferentObject`	`territory/state`, `people/lives`, `economy`, `culture/identity`, `institutions/rule_of_law`
Temporal sequencing & contrast	`Direction`	`decline`, `restoration`, `tipping_point`, `steady_progress`, `cyclical_crisis`
	`ReferencePoint`	`golden_past`, `normal_baseline`, `future_utopia/dystopia`
Moral valuation	`Axis`	`care/harm`, `fairness/cheating`, `loyalty/betrayal`, `authority/subversion`, `purity/degradation`, `liberty/oppression`
Metaphorical mapping	`Type`	`war/invasion/siege`, `flood/swarm/liquid`, `disease/parasite/contamination`, `rot/decay`, `machine/puppet`, `none`

Continued on next page

Supplementary Table S5: Narrative device attributes and allowed values (continued)

Attribute group	Field	Allowed values
Stakes & urgency cues	`Level`	`existential/systemic`, `high_nonexistential`, `low/moderate`
	`Domain`	`physical_security`, `economic`, `cultural/moral`, `institutional/order`
Causal emplotment	`Pattern`	`conspiracy/hidden_hand`, `corruption/self_enrichment`, `incompetence/mismanagement`, `structural_crisis`, `open_hostility/attack`, `subversion/collusion`

Attribute fields are populated only for the corresponding NarrativeDevice and where the relevant value is recoverable from message-internal evidence; otherwise they remain empty. Qualifying device–attribute configurations provide the basis for deterministic NarrativeArchetype derivation in the next subsection.

Attribute glossary (coder-facing)

For transparency and UI alignment, we provide brief coder-facing definitions of the narrative device attributes introduced above.

Supplementary Table S6: Narrative device attribute glossary (coder-facing).

Value	Description
ActorGroup	
`elite_in_group`	National or local leaders, politicians, high officials, or wealthy/powerful figures presented as part of "us".
`people_in_group`	Ordinary members of the in-group ("ordinary people", citizens, voters, everyday supporters).
`minority_in_group`	Minority groups inside the in-group (ethnic, religious, regional, or other subgroups still treated as "us").
`external_outgroup`	Actors framed as outsiders ("they"): foreign states, migrants as outsiders, supranational bodies, enemies.
`abstract_cabal`	Vague or hidden power networks with no clear membership (e.g. "globalists", "deep state", "shadowy elites").
`institution`	Impersonal bodies such as courts, government agencies, media outlets, political parties, firms, NGOs, or international organisations.
ActorLocation	
`internal`	Framed as operating from within the focal polity, community, or institutional setting; this does not by itself imply in-group identification or loyalty.
`external`	Framed as operating from outside the focal polity, community, or institutional setting; this does not by itself imply hostility or out-group identification.

Continued on next page

Table S6 (continued)

Value	Description
Function	
hero/protector	Defends the threatened group or object; keeps harm away.
villain/opponent	Causes harm, blocks solutions, or is the main opponent.
victim	Is harmed or endangered; other actors act on or against them.
traitor	Was expected to be loyal but is portrayed as betraying "us" from the inside.
savior/strong_leader	Individual leader presented as exceptionally or uniquely capable of rescuing the in-group, restoring a valued order, or resolving a major collective problem or crisis.
ReferentObject	
territory/state	Land, borders, sovereignty, or survival of the state or homeland.
people/lives	Safety, health, or survival of people, families, or populations.
economy	Jobs, prices, prosperity, economic stability, or cost of living.
culture/identity	Values, way of life, traditions, language, religion, group identity.
institutions/rule_of_law	Courts, constitution, democracy, media, and basic rules and procedures.
Direction	
decline	Things are getting worse over time.
restoration	Things have declined but can be brought back to a better earlier state.
tipping_point	A decisive moment after which things will change sharply.
steady_progress	Things are gradually improving or moving forward.
cyclical_crisis	Crises recur in cycles; this is one more turn in a repeated up–down pattern.
ReferencePoint	
golden_past	An idealised earlier period when things were much better.
normal_baseline	A recent or ordinary state of affairs treated as "normal".
future_utopia/dystopia	A projected very good or very bad future that anchors the comparison.
Axis	
care/harm	Focus on suffering, protection, harm, cruelty, helping or hurting others.
fairness/cheating	Focus on justice, equality, cheating, corruption, double standards.
loyalty/betrayal	Focus on loyalty to group, nation, or cause vs. betrayal or treason.
authority/subversion	Focus on respect for leaders, rules, hierarchy vs. disobedience or subversion.
purity/degradation	Focus on purity, contamination, disgust, sacred vs. profane.
liberty/oppression	Focus on freedom, control, censorship, or domination by powerful actors.
Type (metaphor)	
war/invasion/siege	War language (battle, enemy, attack, invasion, defend, siege).
flood/swarm/liquid	Water or swarm imagery (floods, waves, tides, flows, swarms).
disease/parasite/contamination	Illness or infestation imagery (virus, cancer, plague, toxicity, infection).
rot/decay	Decay imagery (rot, decay, erosion, crumbling, rust).
machine/puppet	Control or mechanism imagery (puppet, strings, machine, programming, remote control).
none	No clear recurring metaphor pattern; message is largely literal.
Level	
existential/systemic	Threat to survival of people, state, system, or a basic way of life.
high_nonexistential	Serious harm or large losses, but not described as total collapse.
low/moderate	Limited or local harm; inconvenience, cost, or moderate risk.
Domain	
physical_security	Safety of bodies and lives (violence, war, crime, serious health threats).
economic	Money, jobs, prices, markets, and material living standards.
cultural/moral	Values, norms, traditions, religion, and moral codes.
institutional/order	Functioning of institutions, democracy, law, and public order.
Pattern	
conspiracy/hidden_hand	Events or circumstances are attributed to covert coordination or a secret plan by multiple actors seeking to produce, manipulate, or control outcomes. Mere suspicion or concealment without such coordinated causal attribution is insufficient; coding describes the form of the allegation and does not determine its factual truth or falsity.
corruption/self_enrichment	Elites or officials misuse entrusted power for personal or insider benefit.
incompetence/mismanagement	Problems blamed on stupidity, negligence, or poor management.
structural_crisis	Problems attributed to deep structural forces (systems, long-term trends) rather than individuals.
open_hostility/attack	Harm caused by direct, visible attacks or openly hostile acts.
subversion/collusion	Harm or collective disadvantage is attributed to actors undermining the in-group, its institutions, or its interests from within, including collaboration with hostile outside actors; a secret coordinated plot need not be claimed.

C.3 Narrative binding and archetype composition

Appendix B already provides that where defining evidence is relational, the relevant cue spans the elements required to establish that relation. Narrative binding extends the same principle across NarrativeDevice instances: devices may contribute jointly to an archetype signature only where message-internal evidence represents them as belonging to the same narrative configuration.

This follows accounts of emplotment as the configuration of heterogeneous actors and events into a meaningful whole, and work on narrative situation models in which events are integrated through relations including protagonist continuity, temporality, causality, and intentionality (Ricœur, 1984; Zwaan et al., 1995; Zwaan and Radvansky, 1998).

Configuration assignment. Configuration membership is recorded as part of the assigned Narrative representation. Each NarrativeDevice instance receives exactly one configuration identifier, numbered locally within the message (e.g. `C1`, `C2`). The configuration identifier is a structural binding field rather than a NarrativeDevice attribute and therefore is not part of the closed-vocabulary attribute set in Table S5. Annotators begin with `C1` and open a further configuration where device instances belong to a distinct narrative thread or substructure and therefore should not be eligible to combine with the existing instances for archetype derivation. A single overall story may consequently contain more than one narrative configuration.

Devices need not be textually adjacent: a narrative configuration may recur across non-contiguous portions of a message. Devices are placed in the same configuration where the message represents them as participating in the same narrative thread through recoverable relations such as shared actors or referents, a common focal event or problem, causal or temporal linkage, common goals or stakes, or discourse continuity. Co-occurrence within the same artefact, discussion of the same broad topic, or textual proximity does not by itself establish common configuration membership. Where the same message evidence independently supports distinct NarrativeDevice instances belonging to different configurations, those instances are annotated separately rather than assigning one device instance to multiple configurations.

Signature evaluation. NarrativeArchetype signatures are evaluated separately within each narrative configuration, not over the undifferentiated set of NarrativeDevices occurring in the message. Configuration membership supplies the interpretive binding required for composition; once NarrativeDevice instances, their applicable attributes, and their configuration memberships are fixed, signature evaluation introduces no further narrative-coherence judgement.

Where a signature condition requires multiple attributes of one NarrativeDevice, those attributes must be jointly satisfied by the same device instance. Where a configuration contains multiple instances of the required device type, the condition is satisfied if at least one instance in that configuration jointly satisfies the complete device condition. Where a signature requires several device types, the required conditions may be satisfied by qualifying instances of those types within the same configuration.

A single narrative configuration may satisfy zero, one, or multiple Narra-

tiveArchetype signatures. Each satisfied signature independently yields its corresponding NarrativeArchetype. A configuration that satisfies only part of a signature retains its assigned NarrativeDevices and attributes but does not yield a partial NarrativeArchetype.

Signature notation. All signature conditions are evaluated within one narrative configuration. A condition of the form `Device.Attribute = value` means that at least one instance of that NarrativeDevice in the current configuration satisfies the condition. Where multiple conditions in the same signature refer to attributes of the same NarrativeDevice type, those conditions must be jointly satisfied by one instance unless the signature explicitly states otherwise.

Supplementary Table S7: Coder-facing rules for narrative binding and archetype composition.

Rule	**Operational guidance**
What to record	Assign each NarrativeDevice instance exactly one configuration identifier, local to the message. Begin with `C1`; open `C2`, `C3`, and so forth where device instances belong to a distinct narrative thread or substructure and therefore should not be eligible to combine with the existing instances for archetype derivation. A single overall story may contain more than one configuration.
How to bind	Place devices in the same configuration when the message represents them as parts of the same story or situation, for example through shared actors or referents, a common event or problem, causal or temporal linkage, common goals or stakes, or discourse continuity. Devices may be non-contiguous.
Boundary	Do not bind devices merely because they occur in the same artefact, concern the same broad topic, occur near one another, or contain attributes that could jointly complete an archetype signature. If a common narrative relation is not recoverable from the message, keep the instances in separate configurations.
Multiple instances	A configuration may contain multiple instances of the same NarrativeDevice type. Evaluate each required device condition against individual instances; attributes required jointly by that condition must occur on the same instance.
Multiple archetypes	A single configuration may satisfy several archetype signatures. Evaluate each signature independently; every completely satisfied signature yields its corresponding archetype.
Incomplete signatures	If the required device–attribute configuration is not completely satisfied, no NarrativeArchetype is derived. Retain the underlying NarrativeDevices and attributes; do not assign a partial archetype.

C.4 Narrative archetypes and deterministic signature derivation

NarrativeArchetypes are macro-story packages derived within narrative configurations from diagnostic combinations of assigned NarrativeDevice instances and their attributes. They are not Methods and are not independently annotated. Table S8 summarises characteristic devices and spotting cues for interpretive guidance; these descriptive indicators are not themselves activation rules. Each archetype is specified by a minimal signature over coded device attributes in Table S9. Those signatures define the deterministic mapping from qualifying device–attribute configurations to NarrativeArchetypes.

Supplementary Table S8: Narrative archetypes: definitions, devices, and spotting cues.

Archetype	Definition	Narrative devices	Cues
Threat / Enemy at the gates	Frames an external actor or out-group as an existential threat to a valued collective referent, requiring protection (Buzan et al., 1998; Keen, 1986; Miskimmon et al., 2013).	Role-casting (external villain/opponent; optionally protector), stakes/urgency, causal emplotment, conflict–resolution arc; metaphorical mapping (especially invasion/siege or flood/swarm) may reinforce the archetype.	"They're coming", "defend/protect", emergency verbs; border/crime visuals; maps/arrows; swarm/crowd shots; red alert motifs.
Enemy within (fifth column)	Frames actors positioned within the focal polity, community, or its institutions as internal threats who undermine collective interests from within, including through alleged collaboration with hostile outside actors (Wodak, 2015; Chernobrov, 2019; Radnitz and Mylonas, 2022).	Role-casting (internal villain/opponent or traitor), causal emplotment (subversion/collusion); moral valuation (especially loyalty/betrayal), conspiracy/hidden-hand emplotment, metaphors of infiltration or parasitism, and exemplars of alleged insider activity may reinforce the archetype.	"Fifth column", "enemy within", "working for them", "agents from within"; accusations of internal subversion, infiltration, divided loyalty, or collaboration with hostile interests.
Crisis & collapse	Frames a valued social system or domain as undergoing or approaching severe breakdown, such that ordinary conditions are portrayed as no longer sustainable and urgent action as necessary (Moffitt, 2016; Boin, 2005).	Temporal sequencing (decline/tipping point), stakes/urgency; conflict–resolution arcs (e.g. restoring order) and symbolic condensation (e.g. burning-city or empty-shelf imagery) may reinforce the archetype.	"On the brink", countdowns, catastrophe charts; riot/fire/disorder B-roll; "last chance" copy; failure-of-institutions claims.
Corruption / rot at the top	Frames political elites or institutions as abusing entrusted power for personal or insider benefit, thereby casting their exercise of authority as corrupt (Edelman, 1988; Mudde and Kaltwasser, 2017).	Role-casting (elite/institution as villain or opponent), causal emplotment (corruption/self-enrichment), exemplars (e.g. scandals or self-dealing); moral valuation (especially fairness/cheating, and sometimes purity/degradation) may reinforce the archetype.	"Corrupt", "self-dealing", "lining their pockets", "cronies", "kickbacks"; secret-deal or money-transfer imagery; leaked-document screenshots; procurement or patronage claims.
Betrayal	Frames an actor expected to be loyal to an in-group, constituency, or shared cause as violating that loyalty or obligation (Haidt, 2012; Chernobrov, 2019).	Role-casting (internal traitor); moral valuation (especially loyalty/betrayal) and temporal sequencing or contrast (e.g. promise → breach, commitment → reversal) may reinforce the archetype.	"Stabbed in the back", "sold us out", "betrayed us"; oath or pledge imagery; juxtaposed statements or promises then/now; accusations of abandoning the group or cause.
Moral decline / decadence	Frames the community's cultural or moral order as deteriorating through degradation, impurity, or decadence, threatening valued norms or ways of life (Cohen, 2011; Haidt, 2012).	Temporal sequencing (decline), moral valuation (purity/degradation); metaphorical mapping (especially contamination or rot/decay) and exemplars of perceived moral transgression may reinforce the archetype.	"Moral decay", "degeneracy", "our values are disappearing", "protect the children"; disgust or impurity language; shock exemplars; contamination, biohazard, or rot/decay imagery.

Continued on next page

Table S8 (continued)

Archetype	Definition	Narrative devices	Cues
Golden age / restoration	Idealises an earlier period as a superior social or political order and frames its recovery or restoration as a desirable future (Boym, 2001; Roselle et al., 2014).	Temporal sequencing (restoration with a golden-past reference point); symbolic condensation (e.g. heritage symbols), conflict–resolution arcs (restoration), and role-casting (e.g. guardians or restorers) may reinforce the archetype.	"Return to", "restore", "back to basics", "make X great again"; idealised then/now contrasts; heritage imagery, flags, monuments, pastoral or retro visual motifs.
Savior / strong leader	Frames a particular leader as possessing exceptional personal capacity or authority to rescue, restore, or decisively resolve a collective problem or crisis (Weber, 1947; Eatwell, 2006).	Role-casting (savior/strong leader); conflict–resolution arcs and symbolic condensation (e.g. leader-centred or heroic imagery) may reinforce the archetype.	"Only X can fix it", "we need X to save/restore us"; claims of exceptional strength or capacity; leader-centred imagery, heroic or quasi-salvific presentation, adoring-crowd or rescue motifs.
Collective victimhood	Frames an in-group as having been subjected to significant harm, persecution, or unjust treatment because of its group membership (Noor et al., 2017).	Role-casting (in-group as victim); exemplars of collective harm, causal emplotment identifying perpetrators or sources of harm, and moral valuation may reinforce the archetype.	"They have persecuted us", "our people have suffered", "we were targeted"; accounts or lists of harms and casualties; memorial imagery; testimony, remembrance, or historical grievance claims.
Contamination / disease	Frames an actor, group, idea, or social condition through disease, parasite, or contamination metaphors, construing it as infecting, spreading through, or damaging a wider social body or order (Sontag, 1978; Musolff, 2015; Thibodeau and Boroditsky, 2011).	Metaphorical mapping (disease/parasite/contamination); role-casting (e.g. harmful actor as villain/opponent), moral valuation (especially purity/degradation), and stakes/urgency may reinforce the archetype.	"Infestation", "plague", "virus", "cancer", "parasite", "infecting", "toxic", "contaminating"; biohazard or disease imagery; cleansing, sterilising, or containment language.
Conspiracy / hidden hand	Attributes significant events or circumstances to covert coordination, secret plots, or hidden plans by actors seeking to produce or control outcomes (Hofstadter, 1965; Sunstein and Vermeule, 2009; Douglas et al., 2019).	Causal emplotment (conspiracy/hidden hand); role-casting of alleged plotters and exemplars presented as evidence or anomalies may reinforce the archetype, as may symbolic condensation of hidden networks or control.	"They don't want you to know", "behind the scenes", "this was planned", "follow the money"; hidden-network or red-string diagrams; alleged leaked documents; selectively highlighted connections or anomalies; revelation and cover-up language.
Order vs. chaos (law-and-order)	Frames disorder, crime, or norm-breaking as threatening social or institutional order and positions authority, control, or the restoration of order as the appropriate counterpoint (Hall et al., 1978; Wodak, 2015).	Moral valuation (authority/subversion), stakes/urgency (institutional/order); conflict–resolution arcs (restoring order), role-casting of disorderly actors as villains/opponents or authorities as protectors, and symbolic condensation (e.g. uniforms, badges, flags) may reinforce the archetype.	"Restore order", "lawlessness", "anarchy", "take back control"; riot/disorder imagery; curfews, police lines, sirens, badges, uniforms, or other symbols of public authority and control.

Archetype overlap. Derived NarrativeArchetype outputs are non-exclusive: a narrative configuration may instantiate multiple archetypes whenever their respective signatures are independently satisfied, and a message may contain multiple such configurations. Co-instantiation does not imply a parent–child relation between archetypes. For example, *Enemy within (fifth column)* may co-occur with *Betrayal* when an internally positioned actor is cast as a traitor, and with *Conspiracy / hidden hand* when covert coordinated causation is also attributed. Neither relation is required by the Enemy-within signature.

Supplementary Table S9: Narrative archetype signatures from device attributes.

Archetype	Signature (minimal conditions over device attributes)
Threat / Enemy at the gates	∃ `RoleCasting` with `ActorLocation` = `external`, `Function` = `villain/opponent`, and a populated `ReferentObject`; and `Stakes.Level` = `existential/systemic`.
Enemy within (fifth column)	∃ `RoleCasting` with `ActorLocation` = `internal` and `Function` ∈ {`villain/opponent`,`traitor`}; and `CausalEmplotment.Pattern` = `subversion/collusion`.
Crisis & collapse	`TemporalSequencing.Direction` ∈ {`decline`, `tipping_point`}; `Stakes.Level` = `existential/systemic`; and `Stakes.Domain` is populated.
Corruption / rot at the top	∃ `RoleCasting` with `ActorGroup` ∈ {`elite_in_group`, `institution`} and `Function` = `villain/opponent`; and `CausalEmplotment.Pattern` = `corruption/self_enrichment`.
Betrayal	∃ `RoleCasting` with `ActorLocation` = `internal` and `Function` = `traitor`.
Moral decline / decadence	`TemporalSequencing.Direction` = `decline`; `Stakes.Domain` = `cultural/moral`; and `MoralValuation.Axis` = `purity/degradation`.
Golden age / restoration	`TemporalSequencing.Direction` = `restoration`; and `TemporalSequencing.ReferencePoint` = `golden_past`.
Savior / strong leader	∃ `RoleCasting` with `Function` = `savior/strong_leader`.
Collective victimhood	∃ `RoleCasting` with `Function` = `victim` and `ActorGroup` ∈ {`people_in_group`, `minority_in_group`}.
Contamination / disease	`Metaphor.Type` = `disease/parasite/contamination`.
Conspiracy / hidden hand	`CausalEmplotment.Pattern` = `conspiracy/hidden_hand`.
Order vs. chaos (law-and-order)	`Stakes.Domain` = `institutional/order`; and `MoralValuation.Axis` = `authority/subversion`.

D Polarization

D.1 Polarization scope and axes

The polarization dimension records message-internal group-positioning moves: cues that construct, harden, reinforce, soften, cross, or supersede relations and boundaries between social groups. It does not infer population-level polarization from a single message. Polarization motifs and their derived directions are non-exclusive: a message may contain multiple DIVIDE, COHERE, and BRIDGE moves when independently supported by observable cues (Tajfel and Turner, 1979; Iyengar et al., 2012; Mason, 2015;

McCoy et al., 2018).

Each coded motif maps deterministically to exactly one polarization direction and a fixed nonempty set of polarization axes. The axes synthesise distinct strands of the polarization literature for message-level annotation; their labels should not be read as claims that each slash-pair constitutes a single standardised construct in the source literature. The derived axis set records the dimensions of group relations instantiated by the motif, not every downstream outcome with which that motif may be empirically associated.

Cross-dimension boundary with Method. Polarization and Method are analytically distinct dimensions. No general cross-dimensional inheritance rule is defined between them. A Method Tactic does not by itself license a Polarization motif, which additionally requires the relevant operation to position a salient social or political group relation. Conversely, a Polarization motif does not automatically determine a Method Tactic; where the same evidence also satisfies a Method definition, the Tactic is annotated on its own criteria. The same cue span may therefore support annotations in both dimensions without making either dimension reducible to the other.

For example, a dehumanizing representation licenses a Polarization *Dehumanization* motif only where the message additionally positions the represented target within a salient intergroup relation; concentrated blame licenses the Polarization *Scapegoating / concentrated blame* motif only where that blame performs a group-relational positioning operation; and representation of danger licenses *Existential threat* only where a salient outgroup is positioned as threatening the continued physical or symbolic existence of an ingroup or collective referent. Conversely, ingroup continuity, consensus, nostalgia, pride, or shared identity in the Polarization dimension does not automatically license *Appeal to tradition*, *Social proof*, *Collective nostalgic framing*, *Pride / ego appeals*, or *Unity (shared identity)* unless the independent Method-level influence operation is also recoverable.

D.2 Polarization axes (coder-facing glossary)

Axis	Operational definition and boundary
Affective/Identity	Group-relational positioning through social identification, group attachment, or affective and evaluative orientation toward groups, including collective belonging, pride, favouritism, dislike, distrust, contempt, hostility, or social distance (Iyengar et al., 2012; Iyengar and Westwood, 2015; Mason, 2015). Excl.: substantive policy or ideological disagreement without salient group-based identification, evaluation, or social distance.
Issue/Ideological	Group-relational positioning through convergence, divergence, or opposition in substantive political attitudes, policy preferences, issue positions, group interests, values, or ideological commitments (Abramowitz and Saunders, 2008). Excl.: group dislike, identity opposition, or perceived extremity without substantive issue or ideological positioning.
Sorting/Alignment	Group-relational positioning that aligns social or political identities with ideological, issue, value, or normative positions into common camps, aligns multiple identities or cleavages with one another, or conversely makes cross-cutting identities or positions salient across an existing division (Levendusky, 2009; Mason, 2015; Mason, 2016). Excl.: solidarity, hostility, or issue disagreement involving only one already-defined group boundary without alignment or cross-cutting of additional identities, cleavages, or positions.
Perception/Distance	Message-level representation of the apparent extremity, homogeneity, consensus, composition, commonality, or distance of political or social groups insofar as those properties position groups as more or less internally uniform or intergroup-distant (Westfall et al., 2015; Levendusky and Malhotra, 2016; Ahler and Sood, 2018). The axis records how the degree or structure of group polarization is represented in the message and does not determine whether that representation is factually accurate.
Network/Segregation	Group-relational positioning concerning social contact, association, exposure, or separation across political or social group boundaries, including calls to isolate, avoid, connect, cooperate, or interact across groups (Boutyline and Willer, 2017; Mepham et al., 2025). At message level the axis records an orientation toward relational segregation or integration; it does not infer the actual structure of a social network. Excl.: group hostility or solidarity without relational positioning concerning contact, association, exposure, or separation.
Legitimacy/Tolerance	Group-relational positioning concerning legitimate standing, rights, procedural inclusion, acceptable treatment, or tolerated participation of other groups or their members. The axis includes frames that expand or contract who is regarded as entitled to ordinary moral, civic, or procedural protections (Opotow, 1990; Svolik, 2019; Tilley et al., 2026). Excl.: mere hostility, dislike, or disagreement without such normative positioning.

D.3 Relational interpretation of polarization directions

Groups and group boundaries are coded as they are represented or constructed within the message; they need not pre-exist the communication as

fixed social categories. A message may construct a new collective identity, redefine the boundaries of an existing one, divide a previously inclusive category, or recast previously differentiated groups as sharing a broader identity.

DIVIDE marks message-internal moves that construct, sharpen, or reinforce a socially or politically salient boundary between groups, including by differentiating an outgroup from an ingroup or by partitioning a previously represented collective into opposing camps.

COHERE marks message-internal moves that construct, extend, or reinforce a shared collective "we", including through common identity, belonging, solidarity, valued characteristics, commitments, continuity, or collective agency, where the constitutive move is not itself the crossing, softening, or superseding of a salient boundary between differentiated groups.

BRIDGE marks message-internal moves that cross, soften, or supersede a salient boundary between differentiated groups, including by emphasising cross-cutting commonality or incorporating previously differentiated groups within a broader shared identity.

The three directions are non-exclusive. A single message may, for example, construct a broader shared "we" (COHERE) while simultaneously reducing a previously salient subgroup boundary (BRIDGE), or may strengthen solidarity within one collective (COHERE) while sharpening its distinction from another (DIVIDE). Coding records these relational operations as represented in the message and does not imply that audiences actually adopt the identities, boundaries, or relations being advanced.

Ingroup / outgroup. These terms denote message-relative relational positions rather than fixed or externally verified social memberships. An *ingroup* is a collective represented, constructed, or offered by the message as a relevant "we"; an *outgroup* is a collective represented or constructed in contrast to such a "we". Addressees need not already belong to either category, and coding does not require that they accept the proposed categorisation.

Supplementary Table S10: Polarization motifs (Divide): operational definitions and coding boundaries.

Motif	Operational definition and coding boundary
Dehumanization	Represents a social group or its members as less than fully human, including through animalistic, objectifying/mechanistic, or biological representations that recast people themselves as animals, objects, machines, vermin, disease, or contaminating agents (Haslam, 2006; Kteily et al., 2015; Esses et al., 2013). *Incl.:* "vermin", "animals", "infestation", people described as "things", "cogs", a "plague", or a contaminant. *Excl.:* generic insult, dislike, moral condemnation, or claims that a group carries or spreads disease without representing the people themselves as less than human. A dehumanizing representation does not by itself license this Polarization motif unless a salient intergroup positioning relation is additionally recoverable; otherwise the applicable representation may remain solely a Method-level Tactic under Supplementary Appendix B.
Existential threat	Attributes to an outgroup a threat to the continued physical existence or defining identity and continuity of an in-group or collective referent (Hirschberger et al., 2016; Stephan et al., 2009). *Incl.:* annihilation, invasion, extinction, replacement, or disappearance frames; claims that "our people", collective identity, culture, or "way of life" will be destroyed, erased, or cease to exist. *Excl.:* serious threats to safety, resources, interests, or values that are not framed as endangering the group's continued physical or symbolic existence; high-risk claims without salient us–them positioning.
Zero-sum	Represents the interests, gains, resources, status, rights, opportunities, or power of salient social or political groups as necessarily opposed, such that gains for one group entail corresponding losses for another (Esses et al., 1998; Różycka-Tran et al., 2015; Roberts and Davidai, 2022). *Incl.:* "their gain is our loss", "every advantage they get is taken from us", or "take back what's ours" where the message frames one group's gain as the other's loss. *Excl.:* disagreement, rivalry, scarcity, or competition without an express or clearly implied inverse gain–loss relation between groups.
Scapegoating / concentrated blame	Concentrates responsibility for a negative collective condition or outcome on a salient social or political group, representing that group as the principal, overriding, or sole cause of the problem (Rothschild et al., 2012). *Incl.:* "they caused all our problems", "everything went wrong because of them", or repeated attribution of multiple societal harms to the same group. *Excl.:* attribution of a specific, bounded action or outcome to identified actors without generalised group blame; multi-causal accounts in which the group is presented as one contributing factor without concentrated responsibility. Concentrated blame directed at a person or category does not by itself license this Polarization motif unless the blame also positions a salient social or political group within a recoverable group-relational divide.

Continued on next page

Table S10 (continued): Divide motifs.

Motif	Operational definition and coding boundary
Political/procedural delegitimization	Denies a salient social or political group, or its representatives, legitimate standing, authority, or entitlement to participate as a valid actor within the relevant political or civic order (Rivlin-Angert and Mor-Lan, 2025; Daur, 2025). *Incl.:* claims that a group or party has "no legitimate place" in politics, is "not a legitimate opposition", has no right to govern or participate, or falls outside the bounds of acceptable political membership. *Excl.:* criticism of a group's policies, competence, conduct, morality, or specific decisions that does not deny its legitimate standing or participation; terms such as "traitor", "corrupt", "rigged", or "illegitimate" do not qualify by themselves unless they perform such a legitimacy denial.
Shunning / segregation	Advocates, endorses, or normalises reducing social contact, association, interaction, shared space, or information exposure across a salient social or political group boundary (Carlson and Settle, 2022; Peterson and Kagalwala, 2021). *Incl.:* "don't listen to them", "stay in our own spaces", calls to avoid, unfollow, block, exclude, or cease interacting with people or sources because they belong to or are aligned with the outgroup. *Excl.:* avoiding, muting, or blocking a particular person or source because of specific conduct, harassment, relevance, or tone rather than group membership or alignment; selective media choice without salient intergroup positioning.
Extremity portrayal	Represents extreme or radical political or social positions as prevalent, typical, or defining of a salient outgroup (Westfall et al., 2015; Levendusky and Malhotra, 2016; Ahler and Sood, 2018). *Incl.:* "most of them are radicals", "their party has been taken over by extremists", "moderates barely exist on their side", or equivalent portrayals of outgroup extremity as widespread or characteristic. *Excl.:* claims about particular extreme individuals or subgroups that do not generalise extremity to the wider outgroup; neutral descriptions of group composition or issue positions without portraying the group as predominantly extreme.
Moral condemnation	Represents a salient outgroup or its members as characteristically immoral, morally corrupt, depraved, or in violation of fundamental standards of right and wrong (Leach et al., 2007; Brambilla et al., 2013). *Incl.:* characterisations such as "immoral", "evil", "morally bankrupt", "degenerate", "depraved", or "corrupt", including impurity or contamination rhetoric when it conveys moral degradation of the group. *Excl.:* policy disagreement, incompetence, dislike, threat attribution, or allegations of specific misconduct that do not characterise a salient outgroup in moral terms; literal disease or contamination claims without moral evaluation.

Continued on next page

Table S10 (continued): Divide motifs.

Motif	Operational definition and coding boundary
Collective punishment	Advocates or endorses punitive or otherwise adverse treatment of a salient social or political group beyond individually responsible perpetrators, extending sanction to wider group members on the basis of shared group membership or collective responsibility (Lickel et al., 2006; Pereira et al., 2015). *Incl.:* "they should all pay for what their people did", or punishment, expulsion, deprivation, or categorical sanctions extended to wider group members because of wrongdoing attributed to particular members or factions. *Excl.:* sanctions directed at persons for their own specified conduct; or categorical exclusion or adverse policy toward a group without a punitive or retributive link to wrongdoing attributed to other group members.
Homogenization	Represents a salient social or political outgroup as a monolithic opposing bloc by portraying its members as uniformly or near-uniformly sharing beliefs, traits, motives, identities, or recurrent behaviours relevant to the intergroup division, thereby minimising or denying meaningful within-group variation (Rubin and Badea, 2007; Rubin and Badea, 2012). *Incl.:* "they are all the same", "they all think alike", "none of them care about X", or categorical attribution of a non-definitional characteristic that portrays the outgroup as internally undifferentiated. *Excl.:* bounded prevalence claims that preserve meaningful within-group variation; descriptions of a specific unanimous event; category-defining attributes; or neutral descriptions of shared characteristics that do not position the outgroup as a monolithic opposing bloc.
Criminalization / public-order threat	Represents a salient social or political outgroup as characteristically criminal, lawless, violent, or dangerous to public safety or social order (Brouwer et al., 2017; Gruber, 2023). *Incl.:* categorical portrayals such as "criminals", "thugs", "terrorists", "violent mob", or claims that the group characteristically brings crime, violence, lawlessness, or disorder. *Excl.:* descriptions of specific unlawful or violent acts by identified individuals that do not generalise criminality or dangerousness to the outgroup; non-criminal threats to resources, values, identity, or collective survival; labels such as "enemy" that do not themselves imply criminality, violence, or public-order danger.
Conspiratorial attribution	Attributes a consequential social or political event, condition, or harm to the covert, coordinated, and intentional action of a salient outgroup or coalition, thereby positioning that group as acting adversarially against an ingroup or collective referent (Prooijen and Douglas, 2018; Douglas et al., 2019). *Incl.:* claims that a hidden cabal, hostile network, or outgroup is secretly "pulling the strings", engineering events, coordinating a cover-up, or working together against "us" or a valued collective. *Excl.:* ordinary or openly acknowledged coordination; secrecy without a coordinated causal plot; conspiratorial explanations that do not themselves position a salient outgroup adversarially against an ingroup or collective referent; or claims about an individual's hidden motives without group or coalition agency.

Continued on next page

Table S10 (continued): Divide motifs.

Motif	Operational definition and coding boundary
Identity-based ridicule / disparagement	Mocks, belittles, or derisively devalues a salient social or political outgroup by making the group identity itself, or characteristics presented as emblematic of that identity, an object of contempt or ridicule (Ford and Ferguson, 2004; Abrams and Bippus, 2011). *Incl.:* group-directed mockery, derisive caricature, contemptuous jokes, or claims that members are laughable, pathetic, absurd, or beneath respect because of their group identity. *Excl.:* satire or ridicule directed only at an argument, policy, specific action, or individual performance without disparaging a salient social or political group or its identity.
Membership denial	Denies, revokes, or challenges the authentic membership or belonging of a person, subgroup, or constituency in a salient collective "we", positioning the target as not genuinely, no longer, or never properly part of that collective (Cheryan and Monin, 2005; Kteily et al., 2014; Howard et al., 2023). *Incl.:* "they are not real Americans", "RINOs are not real Republicans", "you are no longer one of us", "you ain't Black", or claims that "fake", "false", or insufficiently authentic members do not genuinely belong to the collective. *Excl.:* criticism, deviance, disloyalty, betrayal, or moral condemnation without denial of membership; describing someone as atypical or unrepresentative without denying that they belong; factual or administrative membership-status statements with no social-identity boundary function; or denial of political or procedural standing without denial of membership in the relevant collective. Narrative-dimension betrayal or internal-traitor framing may independently co-occur; terms such as "traitor" or "sellout" qualify here only when they also deny or revoke authentic belonging to the collective.

Supplementary Table S11: Polarization motifs (Cohere): operational definitions and coding boundaries.

Motif	Operational definition and coding boundary
Symbolic / ritual affirmation	Affirmatively invokes, displays, celebrates, or calls for participation in shared group symbols or collective ritual practices as expressions of common ingroup membership, belonging, or unity (Schatz and Lavine, 2007; Whitehouse and Lanman, 2014; Callahan and Ledgerwood, 2016). *Incl.:* affirming use of ingroup flags, emblems, anthems, uniforms, chants, salutes, or collective ceremonies to mark "us", shared membership, or standing together as a group. *Excl.:* merely depicting, mentioning, describing, criticising, or rejecting a group symbol or ritual without using it to affirm common membership or unity; collective activity lacking a symbolic or ritual ingroup-affirmation function.

Continued on next page

Table S11 (continued): Cohere motifs.

Motif	Operational definition and coding boundary
Collective nostalgia	Positively and nostalgically evokes a shared ingroup past in a way that affirms common group identity, belonging, or continuity across time (Wildschut et al., 2014; Smeekes et al., 2021). *Incl.:* sentimental or affirmative references to "our good old days", a valued period "when we stood together", or a positively remembered shared ingroup past presented as expressive of who "we" are. *Excl.:* neutral or merely factual historical references; personal nostalgia without a salient group referent; or positive reference to a specific past achievement without nostalgic or continuity-oriented ingroup framing.
Achievement / collective pride	Affirmatively presents an accomplishment, success, contribution, or valued performance of a collective represented as, or offered as, a shared "we" as grounds for positive collective regard, celebration, or pride (Smith and Mackie, 2016; Mackie and Smith, 2018). *Incl.:* celebrating what "we" achieved, built, won, contributed, or overcame; messages presenting collective accomplishments as grounds for shared positive regard or pride. *Excl.:* neutral or merely factual reports of achievements without affirmative collective evaluation; accomplishments of particular individuals that are not represented as achievements of the collective.
Service / sacrifice	Affirmatively presents the contribution of personal effort, time, resources, comfort, interests, or safety for a shared collective good as an expression or marker of service, commitment, or sacrifice to a collective "we", including where such contribution is presented as a basis for belonging to that collective (Swann et al., 2010; Levine et al., 2005). *Incl.:* calls to serve, volunteer, contribute, endure hardship, give resources, or accept personal costs "for us", "for our people", or for a shared collective good; praise of people for such service or sacrifice; or claims that willingness to serve or sacrifice marks someone as "one of us". *Excl.:* generic duty, loyalty, or obligation language without a specified collective contribution or personal cost; merely compelled compliance where service or sacrifice is not independently affirmed; helping without a salient shared-group relation.
Mutual support / solidarity	Affirmatively represents or offers a shared collective "we" through relations of support, care, or assistance among those represented or invited to belong to it, including commitments or expectations that "we" support one another (Drury et al., 2009; Drury et al., 2016). *Incl.:* "we take care of our own", "when one of us needs help, the rest step in", commitments to mutual material, practical, informational, or emotional support; praise of people for supporting fellow members; or offers of mutual support that invite others into the shared "we". *Excl.:* generic expressions of unity, affection, or belonging without a support relation; individual service or sacrifice for a collective goal without support to fellow members; helping where no salient shared-group relation is represented or offered.

Continued on next page

Table S11 (continued): Cohere motifs.

Motif	Operational definition and coding boundary
Collective efficacy	Represents a salient ingroup as capable, through joint or coordinated action, of achieving a shared goal or producing a desired collective effect (Bandura, 2000; Zomeren et al., 2010). *Incl.:* "together we can", "if we organize, we have the power to succeed", or claims that the group's combined effort, capacity, numbers, skills, or coordination can achieve a shared objective. *Excl.:* generic optimism or confidence without collective capability; calls for action without representing the ingroup as capable of producing the desired outcome; individual efficacy; or external actors presented as the source of success.
Character valorization	Affirmatively attributes traits, qualities, or dispositions to a salient ingroup and presents those characteristics as valued, identity-affirming, or worthy of embrace within the message's own evaluative frame. (Tajfel and Turner, 1979; Luhtanen and Crocker, 1992) *Incl.:* claims that "we" are brave, resilient, disciplined, honest, ruthless, unconventional, defiant, or other characteristics presented as sources of collective worth or identity; affirmative reclamation of otherwise derogatory traits or labels, such as "we're weird and proud of it." *Excl.:* descriptions of ingroup characteristics that are merely factual, critical, ashamed, or externally attributed without affirmative uptake by the ingroup voice; praise of a specific accomplishment without characterising the group more generally.
Tradition / cultural continuity	Affirmatively represents inherited customs, practices, norms, or ceremonies as linking the present ingroup to earlier generations or periods of the same group, thereby portraying the ingroup as culturally continuous through time (Sani et al., 2007; Smeekes and Verkuyten, 2013). *Incl.:* preserving, continuing, passing on, or participating in traditions described as inherited from ancestors, predecessors, or earlier generations; claims that a custom or practice "has always been part of who we are." *Excl.:* contemporary rituals or collective activities with no heritage- or continuity-oriented framing; neutral historical descriptions that do not connect past and present ingroup identity; or personal/family traditions without a salient social or political ingroup referent.
Consensus / normative alignment	Uses represented agreement around a belief, value, preference, norm, goal, or position to construct, extend, or affirm a shared collective identity, including by portraying such agreement as characteristic of "us" or as a basis for belonging to the group (Turner et al., 1989; Hogg and Reid, 2006; Reicher and Hopkins, 1996). *Incl.:* "people like us agree", "this is what we believe", "if you believe in X, you are one of us", or equivalent claims that make shared agreement a marker of common group membership or internal alignment. *Excl.:* generic popularity or polling claims with no shared-identity function; agreement between particular individuals without construction or affirmation of a collective "we"; or issue agreement that is not presented as relevant to group belonging or alignment.

Continued on next page

Table S11 (continued): Cohere motifs.

Motif	Operational definition and coding boundary
Norms / commitment pledges	Defines, affirms, or enacts norms, rules, principles, oaths, or pledges as standards or commitments constitutive of a shared collective "we", including where adherence or commitment is presented as a basis for belonging to the group (Hogg and Reid, 2006; Feldman, 1984; Ellemers et al., 1999). *Incl.:* "this is what we stand for", "members live by this code", "if you accept these principles, you are one of us", oaths of allegiance, or pledges that explicitly enact commitment to the group's identity, values, or membership standards. *Excl.:* administrative or behavioural rules with no shared-identity function; generic statements of duty without a salient collective "we"; promises or pledges unrelated to group membership, allegiance, or normative commitment; or mere enforcement/compliance demands that do not independently present the norm as constitutive of the collective.
Collective mourning / remembrance	Uses shared grief or commemorative remembrance of deaths, victims, losses, or collectively suffered events to construct, extend, or affirm a collective "we", including by presenting mourning or remembrance as something shared by those represented or invited to belong to the collective (Wagoner and Brescó de Luna, 2021; Gur et al., 2024). *Incl.:* "we mourn together", "we remember our fallen", shared remembrance presented as binding "us", invitations to join a collective in mourning or remembrance, or commemorative appeals that make a shared loss part of common group identity. *Excl.:* merely reporting a death, funeral, memorial event, or expression of individual grief without a shared-identity function; generic commemoration of achievements, founding events, or traditions without loss-oriented collective remembrance.
Self-directed humor	Uses playful self-reference, self-mockery, teasing, or insider humour about a collective "we", its members, or shared identity-relevant characteristics or experiences to construct, extend, or affirm common identity, familiarity, or inclusion (Flesher Fominaya, 2007; Moody, 2019). *Incl.:* affectionate or affiliative jokes about "us", shared foibles or stereotyped habits treated playfully as recognizably "ours", insider jokes that mark common membership, or humorous invitations such as "if you get this, you're one of us". *Excl.:* humour unrelated to a shared group identity; serious criticism or disparagement of the group without an affiliative shared-identity function; or ridicule that constructs its target as a differentiated outgroup rather than as part of the collective "we".

Supplementary Table S12: Polarization motifs (Bridge): operational definitions and coding boundaries.

Motif	Operational definition and coding boundary
Humanization	Represents a salient outgroup or its members as fully human persons by foregrounding their personhood, human mental or emotional capacities, agency, individuality, or other features that represent them as persons rather than as merely faceless, object-like, animal-like, machine-like, or psychologically undifferentiated category members (Haslam, 2006; Prati et al., 2016; Capozza et al., 2017). *Incl.:* explicit affirmations that outgroup members are human beings or "people too"; representations of their hopes, fears, grief, attachments, choices, agency, or complex individual lives when these function to affirm their humanity; or individualised portrayals that clearly counter a depersonalised or dehumanizing representation of the outgroup. *Excl.:* merely naming or showing an outgroup member; generic civility, sympathy, liking, or positive evaluation without a humanizing representation; or mentioning shared roles, suffering, or personal details without foregrounding outgroup personhood or humanity.
Superordinate / common identity	Represents members of two or more salient, differentiated social or political groups as belonging to a broader shared category or collective "we", thereby recasting the intergroup relation through a more inclusive common identity; subgroup identities may remain simultaneously salient (Gaertner et al., 1993; Gaertner et al., 1996; Dovidio et al., 2000). *Incl.:* "Democrats and Republicans, we are all Americans", "whatever our faction, we are all members of the same community", or dual-identity formulations that acknowledge subgroup differences while affirming membership in a shared encompassing collective. *Excl.:* an ingroup-only "we" with no salient differentiated groups being brought under a common identity; generic similarity or shared characteristics that are not used to represent the groups as members of a broader collective; or an additional identity that merely cross-cuts the original group boundary without functioning as an encompassing common category.

Continued on next page

Table S12 (continued): Bridge motifs.

Motif	Operational definition and coding boundary
Contact invitation	Invites, encourages, or affirmatively endorses constructive interpersonal interaction across a salient social or political group boundary, presenting members of differentiated groups as people who should meet, speak, listen, engage, or otherwise enter into direct social contact (Pettigrew and Tropp, 2006; Esses and Dovidio, 2002; Meleady, 2021). *Incl.:* "go meet people from the other side", "sit down and listen to one another", invitations to dialogue or conversation between members of opposed communities, or encouragement to seek direct interaction with outgroup members rather than avoid them. *Excl.:* mere co-presence or reported contact without an invitation or affirmative endorsement; interaction with no salient intergroup boundary; hostile confrontation or contact advocated primarily to threaten, denounce, or antagonise the other group; or calls for joint action whose distinctive bridging operation is a shared goal rather than interpersonal contact.
Perspective-taking	Invites or encourages consideration of a situation, experience, action, or disagreement from the subjective standpoint of a member or members of a salient differentiated outgroup, in a way oriented toward understanding their perspective across the group boundary (Galinsky and Moskowitz, 2000; Vescio et al., 2003; Todd et al., 2011). *Incl.:* "put yourself in their shoes", "try to see this from their side", "imagine what this experience is like for them", or invitations to consider how outgroup members' circumstances, experiences, concerns, or viewpoints make the situation appear from their perspective. *Excl.:* merely reporting or describing an outgroup's beliefs, emotions, or position without inviting perspective-taking; imagining an opponent's perspective solely to predict, manipulate, exploit, or defeat them; generic empathy or imagination with no salient intergroup relation; or caricaturing a purported outgroup perspective rather than inviting consideration of that perspective as their subjective standpoint.
Cross-cutting identity	Invokes an identity or category dimension that intersects a salient social or political group boundary, representing members of differentiated groups as sharing, or being distributed across, category memberships that do not simply reproduce the focal division, thereby complicating the message's categorical "us–them" structure (Brown and Turner, 1979; Crisp and Hewstone, 1999; Crisp et al., 2001). *Incl.:* "we may vote differently, but many of us are parents", "veterans are found on both sides", or appeals emphasising overlapping or mixed memberships such as occupation, religion, region, class, generation, or other identities that cut across the salient group boundary. *Excl.:* merely mentioning additional identities without using them to cross or complicate the focal group division; identities that simply align with and reproduce the same "us–them" boundary; or a broader category that directly subsumes the differentiated groups as a single collective, which is coded as superordinate / common ingroup identity.

Continued on next page

Table S12 (continued): Bridge motifs.

Motif	Operational definition and coding boundary
Superordinate goal	Represents two or more salient differentiated social or political groups as sharing a valued objective whose attainment requires their joint or coordinated contribution, thereby positioning the groups as functionally interdependent in relation to that goal (Sherif et al., 1961; Gaertner et al., 1999). *Incl.:* "neither side can solve this alone", "both communities must work together to keep the hospital open", or appeals in which differentiated groups are represented as needing one another's cooperation to achieve a shared desired outcome. *Excl.:* merely stating that groups have similar preferences, interests, or aspirations without representing joint action or interdependence as necessary for attaining the goal; generic calls for unity or cooperation without a specified shared objective; or goals pursued independently by each group rather than through cross-group cooperation.
Procedural fairness	Affirms impartial, reciprocal, or consistently applied procedural standing for salient differentiated social or political groups, including by representing those groups as entitled to legitimate participation, voice, consideration, or access to the same fair decision-making procedures (Tyler, 1988; Tyler, 1994; Tyler, 2000). *Incl.:* "both sides deserve to be heard", "the same rules should apply whether the proposal comes from us or them", equal or impartial opportunities for participation, representation, appeal, or consideration across a salient group boundary, or affirmations that opposing groups retain legitimate procedural standing despite disagreement. *Excl.:* generic fairness, due-process, or administrative rules with no salient intergroup relation; calls for compromise or mutual concessions without a procedural-fairness relation; formally identical treatment invoked without affirming fair procedural standing across the differentiated groups; or procedures that explicitly deny one group legitimate participation or standing.
Epistemic common ground	Represents salient differentiated social or political groups as sharing or accepting a factual claim, body of evidence, premise, observation, or other epistemic reference point despite their broader disagreement, thereby establishing common epistemic ground across the group boundary (Echterhoff et al., 2009; Stern and Ondish, 2018; Tessler et al., 2024). *Incl.:* "we disagree about what should be done, but both sides agree that X occurred", "whatever our politics, we can start from the same evidence", or appeals identifying facts, observations, premises, or evidence as mutually accepted or jointly usable across differentiated groups. *Excl.:* factual or evidentiary claims with no salient cross-group commonality; claims about consensus only within one group; generic acknowledgement of uncertainty or personal fallibility without establishing an epistemic relation across differentiated groups; or merely presenting competing claims without identifying shared epistemic ground.

Continued on next page

Table S12 (continued): Bridge motifs.

Motif	Operational definition and coding boundary
Joint success exemplars	Depicts or recounts a successful instance of contact, friendship, cooperation, problem solving, or joint action between members of salient differentiated social or political groups, where their cross-group engagement is represented as having achieved, resolved, produced, or successfully sustained a valued relational or collective outcome (Wright et al., 1997; Mazziotta et al., 2011; Vezzali et al., 2019). *Incl.:* examples of differentiated groups successfully working together, solving a problem, reaching an agreement, achieving a shared outcome, forming or sustaining a positive cross-group friendship or relationship, or otherwise demonstrating that cross-group engagement succeeded in some meaningful respect. *Excl.:* mere co-presence, contact, dialogue, or attempted cooperation without a represented successful outcome; positive portrayal of an outgroup member without a successful cross-group relation; or success attributed to only one group without salient cross-group engagement.

Supplementary Table S13: Deterministic motif → direction and axis-set mapping (PolarizationEffect lookup).

Motif	Direction	`onAxis`
Dehumanization	Divide	{Affective/Identity}
Existential threat	Divide	{Affective/Identity}
Zero-sum	Divide	{Issue/Ideological}
Scapegoating / concentrated blame	Divide	{Affective/Identity}
Political/procedural delegitimization	Divide	{Legitimacy/Tolerance}
Shunning / segregation	Divide	{Network/Segregation}
Extremity portrayal	Divide	{Perception/Distance}
Moral condemnation	Divide	{Affective/Identity}
Collective punishment	Divide	{Legitimacy/Tolerance}
Homogenization	Divide	{Perception/Distance}
Criminalization / public-order threat	Divide	{Affective/Identity}
Conspiratorial attribution	Divide	{Affective/Identity}
Identity-based ridicule / disparagement	Divide	{Affective/Identity}
Membership denial	Divide	{Affective/Identity}
Symbolic / ritual affirmation	Cohere	{Affective/Identity}
Collective nostalgia	Cohere	{Affective/Identity}
Achievement / collective pride	Cohere	{Affective/Identity}
Service / sacrifice	Cohere	{Affective/Identity}
Mutual support / solidarity	Cohere	{Affective/Identity}
Collective efficacy	Cohere	{Affective/Identity}
Character valorization	Cohere	{Affective/Identity}
Tradition / cultural continuity	Cohere	{Affective/Identity}

Continued on next page

Table S13 (continued): Polarization axis mapping.

Motif	Direction	onAxis
Consensus / normative alignment	Cohere	{Affective/Identity, Sorting/Alignment, Perception/Distance}
Norms / commitment pledges	Cohere	{Affective/Identity, Sorting/Alignment}
Collective mourning / remembrance	Cohere	{Affective/Identity}
Self-directed humor	Cohere	{Affective/Identity}
Humanization	Bridge	{Affective/Identity}
Superordinate / common identity	Bridge	{Affective/Identity}
Contact invitation	Bridge	{Network/Segregation}
Perspective-taking	Bridge	{Affective/Identity}
Cross-cutting identity	Bridge	{Sorting/Alignment}
Superordinate goal	Bridge	{Issue/Ideological, Network/Segregation}
Procedural fairness	Bridge	{Legitimacy/Tolerance}
Epistemic common ground	Bridge	{Perception/Distance}
Joint success exemplars	Bridge	{Network/Segregation}

E TopicDomain groupings

`TopicDomain` indexes the substantive issue area addressed by a message. It is independent of the message's Method, TargetConstruct, Narrative, Polarization, and AboutObject annotations. A topic category does not imply that communicative influence is present, does not license any particular Tactic or group-positioning motif, and does not determine the broader civic or commercial outcome toward which influence is oriented. The same TopicDomain may therefore occur with many different configurations of the other ontology dimensions.

The top-level categories listed above are adapted from three established policy and issue taxonomies: EuroVoc's 21 domains (Publications Office of the European Union, n.d.), the Comparative Agendas Project's 21 major topics (Baumgartner and Jones, 2009), and the 17 top-level IPTC Media Topics (Kuzman and Ljubešić, 2025). The list is not a direct merge; categories salient in contemporary political-communication and influence research (abortion/reproductive rights, gender and sexuality, race/ethnicity as distinct from civil rights, public-health mandates as distinct from general health) are broken out as separate domains where the source taxonomies fold them into larger categories.

E.1 Cultural/Jurisdictional adaptation

Issue salience is culturally bound and time-varying. Agenda-setting and policy-agenda research suggests that issue categories cluster stably across countries, but which domains become politically divisive differs across countries and eras (McCombs and Shaw, 1972; Baumgartner and Jones, 2009).

Supplementary Table S14: TopicDomain groupings with illustrative subtopics.

TopicDomain	Illustrative subtopics
Elections & Governance	voting rules, turnout, courts, parties, corruption
Immigration	border policy, asylum, refugees, deportation
Race/Ethnicity	discrimination, policing, reparations
Religion	religious rights, blasphemy, interfaith conflict
Abortion/Reproductive Rights	abortion access, contraception, IVF
Guns/Crime/Policing	gun rights/control, crime stats, policing policy
Environment & Climate	emissions, energy transition, conservation, impacts/adaptation
Public-Health Mandates	vaccination, masking, lockdowns
Gender & Sexuality	gender identity, LGBTQ rights, sex education, bathroom policies
National Security/War	conflicts, terrorism, military actions
Family & Children	parental rights, curriculum, child protection/abuse, adoption, fertility/IVF
Culture & Media	arts/culture policy, media bias, public broadcasting, content moderation
Health (general)	healthcare quality, insurance, wellness
Economy & Taxation	inflation, jobs, benefits, trade
Education	curriculum, funding, higher ed
Energy & Utilities	fuel prices, grids, nuclear (non-climate)
Housing & Urban Affairs	affordability, zoning, homelessness
Labor & Employment	unions, wages, workplace safety
Science & Technology	AI, data privacy, platforms
Foreign Affairs (non-war)	diplomacy, treaties, sanctions
Consumer/Marketplace	pricing, subscriptions, product claims
Transportation	public transit, EVs, aviation
Disasters & Emergencies	natural disasters, relief, resilience

F AboutObject: Construct and controlled vocabularies

AboutObject as logical outcome. `AboutObject` records a broader civic or commercial outcome that is logically advanced, opposed, strengthened, weakened, maintained, or otherwise oriented by the message as recoverable from message-internal content. It does not represent an empirically observed audience effect and does not require inference of the communicator's private intention.

AboutObject is analytically independent of the Method branch and of the Tactic-derived TargetConstruct profile. The Method branch records how influence is performed through an assigned Tactic and its derived Method family; TargetConstruct records the fixed audience-state profile inherited from that Tactic; AboutObject identifies the broader logical outcome to which the communication bears. Semantic overlap between these analytical components does not constitute redundancy and is not a reason to suppress an otherwise independently satisfied AboutObject label.

Multiple AboutObject labels may be assigned where the message independently advances multiple logical outcomes, including where one represented outcome functions as a premise, intermediate condition, or reason for another. However, two labels should not be maintained as distinct AboutObjects where they merely provide alternative descriptions of the same logical endpoint.

Independence from TopicDomain. `TopicDomain` and `AboutObject` answer different questions and are coded independently. `TopicDomain` records the substantive issue area or areas addressed by the message. `AboutObject` records the broader civic or commercial outcome toward which the message is observably oriented.

No correspondence between the two is required. A message may use material from one TopicDomain as a reason, example, threat, comparison, or other communicative resource in support of an AboutObject located in a different outcome domain. Conversely, the same AboutObject may be pursued through messages concerning many different TopicDomains. TopicDomain therefore does not determine, constrain, or supply an AboutObject label, and an AboutObject label does not determine the message's TopicDomain.

For example, a message about immigration may be oriented toward electoral choice, policy support, institutional delegitimization, intergroup hostility, civic disengagement, or suppression of public expression. These distinctions are retained even where the topical material used to support them is similar. Where independently evidenced, multiple TopicDomains and multiple AboutObject labels may co-occur.

Outcome identity and causal-chain rule. Distinct rhetorical mechanisms, intermediate audience states, or plausible downstream consequences do not constitute distinct AboutObject labels merely because they occur at different points in an influence pathway. AboutObject records the broader outcome toward which the message is observably oriented.

Where two candidate labels describe substantially the same broader outcome, they are represented by a single AboutObject even if the message reaches that outcome through different Methods, TargetConstructs, Narrative structures, or intermediate states. Conversely, multiple AboutObject labels are assigned only where the message independently bears on more than one broader outcome.

Annotators do not infer one AboutObject from another merely because the two commonly stand in a causal relation. For example, reduced trust in journalism does not automatically imply information avoidance, and information avoidance does not automatically imply reduced trust in journalism. Both are coded only where both orientations are independently recoverable from the message.

Directional neutrality. The direction of an AboutObject label carries no normative evaluation of the message, communicator, institution, or promoted outcome. Labels indicating increased trust, cohesion, participation, compliance, support, engagement, or other apparently affirmative outcomes are not treated as inherently benign, prosocial, democratic, truthful, or non-propagandistic. Likewise, labels indicating decreases in those outcomes are

not inherently malign.

AboutObject records the direction of the broader outcome toward which the communication is observably oriented. The Method dimension independently records how influence is performed. The same AboutObject may therefore be pursued through Persuasion, Framing, Manipulation, Coercion, Deception, Distraction, or combinations thereof, subject to the ordinary Tactic coding rules. No Method family or moral characterisation is inferred from the direction of an AboutObject.

Substantive-object rule. An AboutObject label must identify a substantive civic or commercial object of influence. Such an object may be a behaviour, participation pattern, policy or political stance, social relation, trust relation, legitimacy relation, epistemic condition, institutional condition, resilience capacity, or other durable civic or commercial state whose strengthening, weakening, adoption, rejection, maintenance, or disruption constitutes a meaningful logical outcome of communication.

A candidate is not excluded merely because it concerns an epistemic, psychological, relational, or capacity-based condition. Trust in journalism, trust in science, institutional trust or legitimacy, generalised epistemic trust, social cohesion, and resilience to information or influence threats may themselves constitute AboutObjects where alteration of that condition is the broader logical outcome toward which the message bears.

By contrast, generic processing states or intermediate conditions such as confusion, overload, indecision, or paralysis do not constitute AboutObjects merely because they may occur during influence or contribute causally to another outcome. Where such a condition bears on a substantive object, annotate the corresponding substantive AboutObject or AboutObjects instead.

Resilience as an influence object. Resilience concerns the capacity of persons, groups, institutions, or societies to understand, recognise, resist, and recover from manipulative, misleading, or otherwise challenging influence in the information environment. It is distinct from whether particular information or sources are trusted. Communication may therefore logically bear toward strengthening or weakening resilience itself by developing or degrading transferable capacities, practices, norms, or resources for recognising and responding to influence.

Do not infer a resilience outcome merely because a message itself uses, resists, refutes, or exposes an influence tactic. The message must independently bear toward changing broader capacity to withstand or respond to subsequent influence. Resilience direction carries no normative implication: propaganda or other influence communication may itself seek to strengthen resistance to competing messages, sources, or influence attempts.

F.1 L1: Sphere tags (`aboutObjectFamily`)

Non-exclusive tags (closed set). `aboutObjectFamily` is represented at message level by up to two deterministically derived tags:

- `CivicOutcome`
- `CommercialOutcome`

Tags are not mutually exclusive: a message may have one or both.

Deterministic rule. `CivicOutcome` is present if at least one instantiated L3 label has L2 $\in \{A, \ldots, K\}$. `CommercialOutcome` is present if at least one instantiated L3 label has L2 = L. Both sphere tags are present when independently instantiated L3 labels occur in both sets. If no specific L3 label is instantiated, or if only `Unclear` is instantiated, derive no sphere tag.

F.2 L2: Domain taxonomy (A–L; closed list)

A. **Politics** (political-actor support; elections and political-system legitimacy; policy stance and compliance; civic participation) (Kim et al., 2018; Brennen et al., 2020; Kavanagh and Rich, 2018; Anderson and Just, 2013; Dowling and Legrand, 2023).

B. **Identity, Intergroup Relations & Social Cohesion** (collective identity, polarization, outgroup hostility, societal cohesion, group inclusion/exclusion, and humanization/dehumanization) (Kavanagh and Rich, 2018; EUvsDisinfo, n.d.; United Nations, 2019).

C. **Security & Violence** (political-violence support and participation; terrorism support; violent-extremism support and engagement/ disengagement) (Fortify Rights, 2018; United Nations, 2019; McCauley and Moskalenko, 2017).

D. **Epistemic Trust, Legitimacy & Attribution** (trust and legitimacy positioning concerning media, science, institutions, and the wider information environment; generalised epistemic confidence; responsibility attribution and reassignment) (Ognyanova et al., 2020; Lazer et al., 2018; Kavanagh and Rich, 2018).

E. **Ideological Worldview & Community Engagement** (conspiratorial worldviews, anti-pluralist worldviews, and engagement with worldview-reinforcing ideological communities) (EUvsDisinfo, n.d.; Malcolm, 2023; Medzihorsky and Lindberg, 2024).

F. **Public Health & Civil Protection** (health-guidance adherence, health-care engagement, substitute-remedy adoption, crisis protective action, and emergency-service operational capacity) (Nascimento et al., 2022; Hilberts et al., 2025).

G. **Environment, Climate Action & Resilience** (environmental-protection action; climate-change acceptance and policy stance; climate mitigation and adaptation action; climate resilience) (Lewandowsky, 2021; Chatterjee et al., 2024).

H. **Economic, Financial & Infrastructure Stability** (economic and financial-system confidence, resource stockpiling and deposit-run behaviour, and critical-infrastructure confidence and resilience) (Ali et al., 2023; Sandri et al., 2023; OECD, 2019).

I. **Information & Influence Resilience** (capacity to understand, recognise, resist, and recover from information and influence threats; active information seeking and information avoidance) (Roozenbeek et al., 2022a; Hicks et al., 2025).

J. **Geopolitics & Conflict Alignment** (support for international military force and military-service participation; foreign assistance and sanctions; alignment with conflict parties; alliance support and confidence in alliance reliability) (Krishnarajan and Tolstrup, 2023; Geissler et al., 2023; Tomz and Weeks, 2021; Pretorius, 2019).

K. **Personal Security, Privacy & Public Expression** (personal-privacy, personal security, and freedom or suppression of public expression) (Chen et al., 2019; Eckert and Metzger-Riftkin, 2020; Björkenfeldt, 2026).

L. **Commercial Acquisition, Relationship & Advocacy** (commercial inquiry and conversion; subscription, installation, and trial enrollment; customer retention and reactivation; brand loyalty, brand trust, recommendation, review generation, and advocacy) (Lahtinen et al., 2020; Gordon, 2012; Buttle, 1998; Friedman, 1991; Chaudhuri and Holbrook, 2001).

F.3 L3 label catalogue (L1–L2–L3 mapping)

Definition and closure. `aboutObjectLabel` (L3) is a closed vocabulary of cue-based influence-object labels. Each label has a single L2 domain (A–L) and inherits its L1 sphere deterministically from L2. Definitions are cue-based (message-internal) and truth/intent agnostic. If a broader civic or commercial outcome orientation is recoverable but the available message-internal evidence is insufficient to select a specific L3 label, assign `Unclear`. If no AboutObject is recoverable, assign no AboutObject label.

Cross-dimension boundary. AboutObject labels are not inherited from Method Tactics, TargetConstructs, Narrative labels, Polarization motifs, or

TopicDomains. Those annotations may contribute evidence, but an AboutObject label requires message-internal evidence that the communication is oriented toward the corresponding broader civic or commercial outcome. The same Tactic may therefore occur with different AboutObject labels, and the same AboutObject may be pursued through different Tactics. A Tactic or TargetConstruct alone is insufficient to instantiate an AboutObject label.

Specificity and same-endpoint rule. Where two candidate AboutObject labels describe the same logical endpoint at different levels of specificity, instantiate only the most specific label whose definition is satisfied. A generic label is not additionally instantiated merely because the specific outcome is an instance of the broader category.

The broader label may co-occur only where the message independently advances a broader or additional outcome beyond the specific endpoint. This rule applies both within and across L2 domains. TopicDomain does not determine AboutObject specificity; selection is based on the substantive outcome toward which the message bears.

Accordingly, a climate-policy outcome represented by `ClimatePolicy-Support` or `ClimatePolicy-Opposition` is not additionally duplicated as generic `Policy-Support` or `Policy-Opposition`; foreign-aid and sanctions outcomes are likewise represented by their Domain J labels rather than duplicated under generic policy labels. The same principle applies wherever a domain-specific label and a broader label would otherwise describe the identical logical endpoint.

Political-outcome boundaries. Domain A distinguishes support for political actors, legitimacy of elections and political systems, policy stance, policy compliance, and forms of civic participation. These outcomes may co-occur but are not inherited from one another.

`PoliticalActorSupport-*` concerns support for or rejection of an identifiable domestic political leader, candidate, party, governing coalition, or comparable actor. `PoliticalSystemLegitimacy-*` instead concerns the rightful standing or legitimacy of the broader political regime, constitutional order, or form of government. Support for an incumbent does not automatically establish legitimacy of the political system, and system legitimacy does not require support for the incumbent.

`ElectionLegitimacy-*` concerns the validity, fairness, authority, or legitimacy of a particular electoral process or outcome. `InstitutionalLegitimacy-*` in Domain D concerns the rightful authority of a particular public or civic institution. These labels are not substituted for one another merely because elections, institutions, and political systems stand in causal or constitutional relations.

`Policy-Support` and `Policy-Opposition` concern stance toward a pol-

icy or governmental measure. `PolicyCompliance-*` concerns behavioural conformity with or departure from the represented rule, requirement, or implementation measure. Policy stance does not mechanically determine policy compliance.

Where a domain-specific AboutObject represents the identical political or behavioural endpoint more precisely, apply the global specificity rule rather than duplicating the generic Domain A label.

Boundary with the Polarization dimension and adjacent intergroup outcomes. The Polarization dimension and identity/intergroup AboutObjects are coded independently. The Polarization dimension records message-internal relational operations: how the communication constructs, reinforces, crosses, softens, or otherwise positions salient group boundaries through DIVIDE, COHERE, and BRIDGE motifs. These annotations describe the group-relational move enacted within the message and do not by themselves establish a broader societal outcome.

`CollectiveIdentity-Strengthening` and `CollectiveIdentity-Weakening` instead concern the broader logical outcome of increasing or decreasing identification, attachment, belonging, or investment in an identifiable collective. A COHERE move does not automatically instantiate either label: shared identity may be used only as a communicative resource for another outcome. Collective identity is coded as an AboutObject only where alteration of identification or belonging is itself independently recoverable as an outcome toward which the message bears.

`CollectiveIdentity-Reconfiguration` concerns alteration of what a collective identity means, who belongs to it, or what membership normatively entails. Message-internal construction or invocation of an ingroup is insufficient unless such alteration of identity content or boundaries is itself a broader logical outcome.

`Polarization-Increase` and `Polarization-Decrease` concern the broader condition of division between salient social or political groups. Accordingly, a DIVIDE or BRIDGE move does not automatically instantiate the corresponding AboutObject. The communication must independently bear toward increasing or decreasing the wider intergroup division itself.

`SocietalCohesion-Increase` and `SocietalCohesion-Decrease` concern the condition of the broader social fabric within the relevant society or civic community, including social trust, shared belonging, reciprocal commitment, cooperation, and capacity for constructive relations across its constituent groups. Cohesion within a single ingroup is insufficient. A message may strengthen an ingroup while simultaneously weakening societal cohesion by excluding, antagonising, or separating other constituent groups.

Likewise, hostility, inclusion, and humanization AboutObjects require the corresponding broader intergroup condition to be independently recov-

erable. Hostile, exclusionary, humanizing, dehumanizing, COHERE, DIVIDE, or BRIDGE representations may supply evidence but do not mechanically determine an AboutObject.

Security-outcome boundaries. Security and violence AboutObjects distinguish attitudes toward violence from participation in violence, and attitudinal support for violent extremism from behavioural engagement with violent-extremist actors or activities. These outcomes may co-occur but are not inherited from one another.

A message that praises, glorifies, justifies, condemns, or delegitimizes violence does not automatically establish a participation outcome. Likewise, a call to join or leave an extremist organisation does not automatically establish a change in ideological or attitudinal support. Multiple labels are assigned only where each broader outcome is independently recoverable.

AboutObject labels in this domain are also independent of Method, Polarization, and Narrative annotations. Incitement, threat, glorification, enemy construction, dehumanisation, or other message-internal operations may provide evidence for a security outcome but do not mechanically determine one.

`PoliticalViolenceSupport-*` is the generic support object for physical political or intergroup violence. It is not additionally instantiated where the identical support outcome is already represented by the more specific `TerrorismSupport-*` or `ViolentExtremismSupport-*` label. Use the generic political-violence label additionally only where support for political violence more broadly is independently advanced.

`TerrorismSupport-*` concerns support for or rejection of represented terrorist acts, terrorist violence, or actors specifically in relation to their terrorist conduct. `ViolentExtremismSupport-*` concerns broader support for or rejection of a violent-extremist movement, group, orientation, or its violent ideological project. Support for a terrorist act does not automatically establish support for the wider movement or ideology, and support for a violent-extremist movement does not automatically establish support for every represented terrorist act.

Epistemic-trust, legitimacy, and attribution boundaries. Domain D distinguishes trust in specific epistemic source classes, institutional trust, institutional legitimacy, generalised epistemic trust, and responsibility attribution. These outcomes may co-occur but are not inherited from one another.

`MediaTrust-*` and `ScienceTrust-*` concern trust in their respective epistemic source classes. They are not additionally duplicated as `InstitutionalTrust-*` merely because a particular media or scientific organisation is institutionally organised. Use `InstitutionalTrust-*` for

trust in the reliability, competence, honesty, or procedural dependability of other public or civic institutions, or where a distinct institutional-trust outcome is independently recoverable.

`GeneralizedEpistemicTrust-*` requires a broader outcome concerning whether reliable knowledge, evidence, authentic information, or dependable means of distinguishing warranted from unwarranted claims remain available in the relevant information environment. Distrust of one source, institution, scientific claim, or item of evidence does not automatically instantiate a generalised epistemic-trust outcome.

Institutional trust and institutional legitimacy are coded separately. Reliability, competence, honesty, or procedural dependability concern trust; rightful authority, standing, or entitlement to exercise an institutional role concern legitimacy. Both are assigned only where both relations are independently recoverable.

`ResponsibilityAttribution-Shift` concerns a change in who or what is represented as causally, morally, or practically responsible for an event, condition, or outcome. Blame assignment, counter-attribution, or false-flag representations may provide evidence, but responsibility reassignment does not automatically instantiate distrust, hostility, delegitimization, or another AboutObject.

Ideological-worldview boundaries. Domain E records broader ideological or conspiratorial orientations and engagement with communities organised around such orientations. It does not instantiate an AboutObject merely because a message contains a conspiracy narrative, ideological claim, anti-pluralist representation, or invitation to an ordinary political or civic group.

A specific conspiracy claim or hidden-plot narrative does not automatically instantiate `ConspiratorialWorldview-Increase`. The broader outcome requires an independently recoverable orientation toward strengthening a generalised conspiratorial interpretation of consequential social or political events. Conversely, a generalised conspiratorial worldview does not require reduced generalised epistemic trust: a message may express strong certainty that hidden actors coordinate events.

Anti-pluralist worldview outcomes concern broader acceptance or rejection of pluralism as a legitimate organising principle, including the legitimacy of competing groups, viewpoints, opposition, and societal diversity. Exclusion or delegitimization of one identifiable group does not automatically instantiate an anti-pluralist worldview unless the broader rejection of pluralism is independently recoverable.

Ideological-community engagement is coded separately from worldview adoption, collective identity, ordinary civic participation, and violent-extremist engagement. Multiple labels may be assigned where changes in worldview, community engagement, identity, or violent engagement are each

independently recoverable.

Public-health and civil-protection boundaries. Domain F records substantive health and civil-protection outcomes recoverable from the message rather than the medical truth, safety, efficacy, or appropriateness of the communication. Annotators do not determine whether a represented treatment, remedy, recommendation, protective action, or warning is medically correct, evidence-based, safe, or effective.

Health-guidance adherence concerns adherence to or departure from an identifiable health recommendation, intervention, or guidance relation. Health-care engagement instead concerns seeking, maintaining, avoiding, delaying, refusing, or withdrawing from engagement with health-care services or professional care. These outcomes may co-occur but are not inherited from one another.

Substitute-remedy adoption concerns a remedy or treatment represented within the message as an alternative or substitute for another health intervention, recommendation, treatment, or professional care. The label records the represented substitution and adoption outcome; it does not classify the substitute as medically safe, dangerous, effective, ineffective, proven, or unproven.

Crisis protective action concerns behaviour directed at responding to an emergency or hazard, such as evacuation, sheltering, preparedness, or other represented protective response. Emergency-service capacity instead concerns the operational availability or burden of emergency-response services themselves. Neither outcome is inferred automatically from the other.

Trust in health or emergency institutions is coded under the applicable Domain D trust object. Transferable capacity to recognise and resist health misinformation is coded under Domain I information resilience. Multiple AboutObjects may be assigned where these outcomes are independently recoverable.

Where one represented behaviour simultaneously qualifies as health guidance and crisis protective action, do not duplicate an identical behavioural endpoint across both labels. Use `CrisisProtectiveAction-*` where the substantive object is immediate or emergency crisis-response behaviour, and use `HealthGuidanceAdherence-*` where the substantive object is adherence to an identifiable health recommendation or health-directed course of action. Code both only where distinct adherence and crisis-response outcomes are independently recoverable.

Environmental and climate-outcome boundaries. Domain G distinguishes environmental-protection behaviour, acceptance of climate-change propositions, support or opposition toward climate policy, climate-related mitigation and adaptation action, and the longer-term resilience capacity of

relevant social, institutional, economic, or ecological systems. These outcomes may co-occur but are not inherited from one another.

`ClimateChangeAcceptance-*` concerns acceptance or rejection of substantive represented claims about climate change, including its occurrence, anthropogenic contribution, or comparable core climate-change propositions. It is distinct from trust in science or scientists. Claims positioning scientists, scientific institutions, or scientific expertise as trustworthy or untrustworthy are coded under the applicable Domain D `ScienceTrust-*` label. A message may independently bear on both.

Climate-policy stance is distinct from climate-related action. Supporting or opposing a governmental measure does not automatically establish a change in individual or collective mitigation or adaptation behaviour, and behavioural mobilisation does not automatically establish policy support or opposition.

Mitigation action concerns behaviour oriented toward reducing or limiting represented contributions to climate change. Adaptation action concerns adjustment to represented actual or expected climate effects. Neither is identical to climate resilience, which concerns the broader capacity of a social, institutional, economic, or ecological system to cope, continue functioning, reorganise, adapt, learn, or transform under climate-related disturbance.

A mitigation or adaptation action does not automatically instantiate `ClimateResilience-Increase`. Code resilience additionally only where strengthening or maintaining broader capacity to withstand or respond to climate-related disturbance is itself independently recoverable as a logical outcome. Multiple AboutObjects may be assigned where the respective outcomes are independently supported.

Where the identical resilience endpoint concerns a critical infrastructure or essential-service system, use the more object-specific `CriticalInfrastructureResilience-*` label in Domain H rather than duplicating that endpoint as `ClimateResilience-*`. Climate change or a climate-related hazard may remain the relevant TopicDomain or hazard context. Use `ClimateResilience-*` for broader climate-resilience outcomes concerning communities, social systems, ecological systems, or other climate-adaptive capacity not already represented by the more specific critical-infrastructure object.

`EnvironmentalProtectionAction-*` covers environmental-protection or conservation behaviour not already represented by a more specific climate behavioural object. Climate mitigation and adaptation labels take precedence where the identical endpoint is explicitly climate-related. Environmental TopicDomain alone does not instantiate environmental-protection action.

Economic, financial, and infrastructure boundaries. Domain H distinguishes confidence in economic or financial systems from behaviour that

may follow from such confidence, and distinguishes perceived infrastructure dependability from infrastructure resilience capacity. These outcomes may co-occur but are not inherited from one another.

`EconomicConfidence-*` concerns broader confidence in the condition, stability, prospects, or functioning of the relevant economy. `FinancialSystemConfidence-*` concerns confidence in the safety, stability, reliability, or continued functioning of banks, deposits, or the represented financial system. Claims about public or civic institutions are coded under the applicable Domain D trust or legitimacy object where that institutional relation is independently targeted.

Loss of economic or financial confidence does not automatically instantiate stockpiling or deposit-run behaviour. Behavioural labels are added only where increased or decreased stockpiling or run participation is itself an independently recoverable logical outcome.

`EssentialResourceStockpiling-*` records increased or decreased accumulation of essential goods or resources in a represented shortage, crisis, or run context. The ontology does not infer that such behaviour is irrational, emotionally driven, or properly characterised as "panic."

`CriticalInfrastructureConfidence-*` concerns confidence in the dependability or continuity of essential infrastructure or services. `Critical InfrastructureResilience-*` instead concerns the broader capacity of such systems to withstand, absorb, adapt to, recover from, or continue functioning during disruption. Confidence and resilience are coded together only where both outcomes are independently recoverable.

Multiple AboutObjects may be assigned where confidence, behaviour, and resilience outcomes are each independently supported.

Information-behaviour boundaries. Domain I distinguishes transferable resilience capacity from particular information practices. `Informatio nSeeking-Increase` concerns active acquisition, consultation, comparison, monitoring, or verification of information. `InformationAvoidance-Increase` concerns practices that reduce, restrict, or exclude engagement with information.

Neither behaviour is inferred mechanically from the other, and they are not treated as exact directional opposites. A message may encourage seeking from one source while encouraging avoidance of another.

A one-off information-seeking or verification instruction does not automatically instantiate `InformationResilience-Increase`. Resilience requires an independently recoverable orientation toward strengthening transferable capacity to recognise, evaluate, resist, or recover from subsequent influence. Source-specific or generalised trust outcomes are coded under Domain D where independently recoverable.

Separate Decrease labels are intentionally omitted for `InformationS`

eeking and InformationAvoidance. These are represented as directional practices rather than as endpoints of a single bipolar continuum: increased information acquisition is coded as InformationSeeking-Increase, while increased disengagement or non-exposure is coded as InformationAvoidance-Increase. Neither label is inferred solely from the absence or reduction of the other.

Geopolitical and conflict-alignment boundaries. Domain J records broader foreign-policy and geopolitical outcomes concerning international military force, military-service participation, foreign assistance, sanctions, alignment with conflict parties, and alliance relationships. These outcomes may co-occur but are not inherited from one another.

Support for a foreign actor or conflict party does not automatically imply support for military intervention, foreign assistance, sanctions, or alliance commitments on that actor's behalf. Likewise, support for a particular foreign-policy instrument does not automatically imply broader support for the actor benefiting from that instrument. Multiple labels are assigned where each outcome is independently recoverable.

AllianceSupport-* concerns support for maintaining, strengthening, weakening, leaving, or otherwise changing an alliance or comparable security alignment itself. AllianceReliabilityConfidence-* instead concerns confidence that represented alliance partners or the alliance relationship will reliably perform, honour, or sustain relevant commitments. Confidence in alliance reliability does not automatically establish support for the alliance, and alliance support does not automatically establish confidence in its reliability.

Domain J military-force outcomes concern interstate or foreign-policy uses of organised military force, including intervention, defence of allies, or continuation or withdrawal of military operations. Political or intergroup violence, terrorism, and violent-extremist participation are coded under Domain C where those corresponding outcomes are independently recoverable.

Foreign-aid and sanctions labels are domain-specific policy AboutObjects. They are not additionally duplicated under generic Policy-Support or Policy-Opposition for the identical policy endpoint. Other policy AboutObjects may nevertheless co-occur where genuinely distinct policy outcomes are independently advanced.

MilitaryServiceParticipation-* concerns behavioural entry into, continuation of, or withdrawal from regular military service. MilitaryForceSupport-* instead concerns public or political support for the use of organised military force. Support for war or intervention does not imply willingness to serve, and willingness to serve does not imply support for every represented use of military force.

Where an identical military-service endpoint could also be characterised

as generic policy compliance, the more specific Domain J label applies under the same-endpoint rule.

Personal-security and expression boundaries. Domain K records substantive outcomes concerning personal privacy, personal security, and public expression. It does not instantiate an AboutObject merely because a message contains a threat, blackmail, disclosure leverage, harassment, intimidation, reputational pressure, or another coercive operation.

Coercion is represented in the Method dimension. The substantive outcome sought through coercion is coded as the corresponding AboutObject wherever that outcome belongs in the ontology. Thus coercion used to suppress speech may support `PublicExpression-Decrease`; coercion used to alter voting, policy stance, commercial behaviour, organisational engagement, or another substantive outcome is coded under the corresponding AboutObject rather than as a generic coercive-compliance outcome.

`PersonalPrivacy-*` concerns the broader condition of privacy, including protection, confidentiality, anonymity, control over disclosure, and unwanted accessibility or dissemination of identifiable personal information. Doxxing, disclosure, identification, tracing, or dissemination may provide evidence for privacy decrease but are not themselves AboutObjects.

A threat to disclose or expose personal information does not automatically instantiate `PersonalPrivacy-Decrease` where the threatened privacy loss functions only as a conditional consequence used to obtain another outcome.

`PersonalSecurity-*` concerns the broader security of identifiable persons or populations against targeting, intimidation, retaliation, harassment, or comparable personal harm. A threat or hostile communication does not automatically instantiate security decrease unless weakening personal security itself is independently recoverable as an outcome toward which the communication bears.

`PublicExpression-*` concerns increased or decreased participation in public expression, including speaking, publishing, reporting, communicating, or otherwise making views or information publicly available. Threat, harassment, surveillance, or reputational pressure may provide evidence where silencing or sustained expression is independently recoverable, but those means do not mechanically determine the AboutObject.

Civic participation outside public expression is coded under the applicable Domain A participation object. Multiple AboutObjects may be assigned where personal privacy, personal security, public expression, or another substantive outcome are each independently recoverable.

Commercial-outcome boundaries. Domain L distinguishes commercial acquisition and conversion behaviours, continuation or restoration of cus-

tomer relationships, longer-term brand loyalty, and customer communication or advocacy behaviours. These outcomes may co-occur but are not inherited from one another.

Commercial inquiry, purchase, subscription enrollment, app installation, and trial enrollment are distinct observable market behaviours. A message may advance one as a pathway toward another, but a later conversion outcome is not inferred merely because an earlier funnel stage is advanced.

`CustomerRetention-*` concerns continuation or termination of an existing customer relationship. `CustomerReactivation-*` concerns restoration or continued non-restoration of a relationship that is represented as already lapsed. `BrandLoyalty-*` instead concerns the broader preferential commitment or durable attachment of customers toward a brand, provider, product, or service. Retention, reactivation, and loyalty may influence one another but are coded separately.

`BrandRecommendation-*` concerns recommending or referring the brand, product, or service to others. `ReviewGeneration-*` concerns production or withholding of public consumer reviews or ratings. `BrandAdvocacy-*` concerns broader active public endorsement, defence, promotion, or ambassador-like representation extending beyond an isolated recommendation or review.

Commercial boycott, switching, cancellation, avoidance, or comparable market withdrawal does not receive a separate AboutObject merely because a particular mobilising form is used. Code the substantive market relationship affected, such as `PurchaseBehavior-Decrease`, `CustomerRetention-Decrease`, or another independently recoverable commercial outcome.

TopicDomain is independent of commercial outcome. A commercially instantiated AboutObject may therefore be pursued through political, environmental, identity, health, geopolitical, or other topical material. Multiple AboutObjects are assigned where distinct commercial outcomes are independently recoverable.

`BrandTrust-*` concerns confidence in the reliability, honesty, dependability, competence, or promise-fulfilment of a commercial brand or provider. It is distinct from `BrandLoyalty-*`, which concerns durable preferential commitment, and from `CustomerRetention-*`, which concerns continuation of an existing commercial relationship. Trust, loyalty, and retention may influence one another but are not inherited from one another.

Where the relevant outcome concerns confidence in the stability, solvency, or continued functioning of the banking or financial system rather than trust in a commercial provider as an exchange partner, use the applicable Domain H `FinancialSystemConfidence-*` label.

Supplementary Table S15: L3 label catalogue with deterministic L3-to-L2 domain and L2-to-L1 sphere inheritance.

L3 label	L2	Cue-based definition
VoterRegistrationIncrease	A	Cues increased voter registration or registration checking through mobilisation, procedural guidance, encouragement, or other message-internal orientation toward becoming or remaining registered to vote.
VoterRegistrationDecrease	A	Cues reduced voter registration, registration completion, or maintenance of registration through discouragement, avoidance, boycott, procedural deterrence, or other message-internal orientation away from being registered to vote. *Boundary:* Do not infer `VoteTurnout-Decrease` merely because registration is discouraged or impeded. Registration and ballot participation are distinct AboutObjects. Code turnout decrease additionally only where the message independently bears toward not casting a ballot or reduced voting participation.
VoteTurnoutIncrease	A	Cues increased participation in voting through mobilising calls, logistics, efficacy claims, encouragement, or other orientation toward casting a ballot in an election or referendum.
VoteTurnoutDecrease	A	Cues reduced participation in voting through abstention appeals, electoral boycott, demobilisation, fatalism, discouragement, or other orientation toward not casting a ballot. *Boundary:* A message directed only at voter registration does not by itself instantiate `VoteTurnout-Decrease`, even though failure to register may causally prevent later voting. Code both only where reduced registration and reduced ballot participation are independently recoverable logical outcomes.
PoliticalActorSupportIncrease	A	Cues a broader logical outcome of increased approval, preference, backing, or support for an identifiable domestic political leader, candidate, party, governing coalition, or comparable political actor. Support may be advanced through endorsement, favourable comparison, voting recommendation, approval of leadership or performance, or another message-internal orientation toward supporting that actor. *Boundary:* Support for a policy associated with the actor does not automatically instantiate this label unless support for the political actor itself is independently recoverable. Where the relevant actor is represented as a party to an international conflict, the more specific Domain J `ConflictPartySupport-*` label applies to the identical geopolitical alignment endpoint.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
PoliticalActorSupportDecrease	A	Cues a broader logical outcome of reduced approval, preference, backing, or support for an identifiable domestic political leader, candidate, party, governing coalition, or comparable political actor, including electoral rejection or broader withdrawal of political support. *Boundary:* Criticism of one act, statement, or policy is insufficient unless reduced support for the political actor itself is independently recoverable.
PolicySupport	A	Cues increased support for a specific policy, regulation, law, governmental measure, or comparable policy course through stance, endorsement, advocacy, or mobilisation toward its adoption, maintenance, or implementation. *Boundary:* Mere attribution of a policy position to a candidate or party is insufficient. Where a policy is mentioned only as a reason to support or oppose an electoral option, code the applicable `PoliticalActorSupport-*` label unless a stance toward the policy itself is also independently recoverable.
PolicyOpposition	A	Cues increased opposition to a specific policy, regulation, law, governmental measure, or comparable policy course through rejection, resistance, advocacy, or mobilisation against its adoption, maintenance, or implementation. *Boundary:* Mere attribution of a policy position to a candidate or party is insufficient. Where a policy is mentioned only as a reason to support or oppose an electoral option, code the applicable `PoliticalActorSupport-*` label unless a stance toward the policy itself is also independently recoverable.
PolicyComplianceIncrease	A	Cues a broader logical outcome of increased behavioural compliance with an identifiable law, regulation, governmental rule, requirement, obligation, or public-policy implementation measure represented in the message. *Boundary:* Support for or approval of the policy does not automatically instantiate compliance. The message must independently bear toward behaviour conforming to the represented rule or requirement. Where the identical behavioural endpoint is represented by a more specific domain label, such as health-guidance adherence, crisis protective action, climate action, or military-service participation, use that more specific label rather than duplicating the endpoint here.
PolicyComplianceDecrease	A	Cues a broader logical outcome of reduced, refused, evaded, avoided, or discontinued behavioural compliance with an identifiable law, regulation, governmental rule, requirement, obligation, or public-policy implementation measure. *Boundary:* Opposition to the policy itself is insufficient unless non-compliance with the represented rule or requirement is independently advanced.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
ProtestParticipationIncrease	A	Cues increased participation in demonstrations, rallies, strikes, marches, or comparable protest action through mobilisation, logistics, encouragement, or other orientation toward joining or supporting the action.
ProtestParticipationDecrease	A	Cues reduced participation in demonstrations, rallies, strikes, marches, or comparable protest action through discouragement, demobilisation, withdrawal, avoidance, or other orientation away from joining or supporting the action.
CivicEngagementIncrease	A	Cues increased engagement with civic or democratic participation beyond a single specifically coded voting or protest act, including contacting public officials, petitioning, volunteering, participating in civic or political organisations or meetings, contributing to campaigns or causes, or otherwise engaging through institutional or collective civic channels. *Specificity boundary:* Do not instantiate `CivicEngagement-Increase` or `CivicEngagement-Decrease` merely because a message bears on one specifically represented participation form already captured by a narrower AboutObject such as voter registration, voting, or protest participation. Use the broader civic-engagement label where participation or withdrawal across civic or democratic channels is itself the recoverable object, or where the message addresses civic engagement in a generalised form extending beyond one specifically coded act.
CivicEngagementDecrease	A	Cues reduced engagement with civic or democratic participation broadly, including withdrawal from civic institutions, organisations, advocacy, consultation, campaigning, or other participatory channels through disengagement, fatalism, demobilisation, or rejection of participation. Mere discouragement of one specific voting or protest act is insufficient unless broader civic withdrawal is independently advanced. *Specificity boundary:* Do not instantiate `CivicEngagement-Increase` or `CivicEngagement-Decrease` merely because a message bears on one specifically represented participation form already captured by a narrower AboutObject such as voter registration, voting, or protest participation. Use the broader civic-engagement label where participation or withdrawal across civic or democratic channels is itself the recoverable object, or where the message addresses civic engagement in a generalised form extending beyond one specifically coded act.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
ElectionLegitimacyIncrease	A	Cues strengthened acceptance, confidence, or recognition of an electoral process or outcome as valid, fair, properly constituted, authoritative, or legitimate, without requiring factual adjudication of the underlying claims. *Boundary:* This label concerns the legitimacy or validity of an electoral process or outcome, not merely trust in the institution administering it. Claims about the competence, honesty, reliability, or rightful authority of an electoral commission or other institution license the corresponding Domain D trust or legitimacy label only where that institutional outcome is independently recoverable.
ElectionLegitimacyDecrease	A	Cues weakened acceptance, confidence, or recognition of an electoral process or outcome as valid, fair, properly constituted, authoritative, or legitimate, including representations of elections as rigged, stolen, fraudulent, invalid, or otherwise unworthy of recognition, without requiring factual adjudication. *Boundary:* Criticism or distrust of an electoral institution does not by itself instantiate `ElectionLegitimacy-Decrease`. The electoral process, result, or its validity must itself be placed in question. Conversely, an election may be represented as illegitimate without independently attacking the trustworthiness or rightful authority of a particular institution.
PoliticalSystemLegitimacyIncrease	A	Cues a broader logical outcome of strengthened acceptance of the relevant political system, regime, constitutional order, or form of government as rightful, appropriate, entitled to govern, or worthy of continued political recognition and support. *Boundary:* This label concerns legitimacy of the broader political order rather than approval of a particular leader, validity of a particular election, trust in a specific institution, or acceptance of one policy. Those outcomes are coded separately where independently recoverable.
PoliticalSystemLegitimacyDecrease	A	Cues a broader logical outcome of weakened acceptance of the relevant political system, regime, constitutional order, or form of government as rightful, appropriate, entitled to govern, or worthy of continued political recognition and support. *Boundary:* Criticism of an incumbent, institution, election, or policy does not automatically instantiate this label. Delegitimation of the broader political order itself must be independently recoverable.
CollectiveIdentityStrengthening	B	Cues a broader logical outcome of strengthened identification, attachment, belonging, solidarity, or psychological investment in an identifiable social, political, national, religious, civic, or other collective identity. Mere use of an existing shared identity as an argument or rhetorical resource is insufficient unless strengthening that collective identification is itself an independently recoverable outcome.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
CollectiveIdentityWeakening	B	Cues a broader logical outcome of weakened identification, attachment, belonging, solidarity, or psychological investment in an identifiable collective identity, including orientations toward detachment, disidentification, fragmentation, or erosion of collective belonging.
CollectiveIdentityReconfiguration	B	Cues a broader logical outcome of altering the represented content, boundaries, norms, obligations, or meaning of an identifiable collective identity, including changing who properly belongs to the group or what membership in that group is represented as requiring or entailing. Mere identity reference is insufficient.
PolarizationIncrease	B	Cues a broader logical outcome of increased polarization between salient social or political groups through harder intergroup division, increased zero-sum opposition, greater perceived incompatibility, segregation, or reduced scope for coexistence, cooperation, dialogue, or compromise. *Boundary:* `Polarization-Increase` is not inherited automatically from a Polarization-dimension DIVIDE annotation. A DIVIDE move records the message-internal construction, sharpening, or reinforcement of a salient group boundary. The AboutObject is additionally instantiated only where a more divided, antagonistic, incompatible, segregated, or less cooperative intergroup condition is itself recoverable as part of the broader logical outcome toward which the communication bears. A group boundary used primarily as a premise or mobilising structure for another outcome—for example electoral choice, policy support, protest participation, or collective action—does not by itself license `Polarization-Increase`. Multiple AboutObjects may nevertheless be assigned where polarization increase and those other outcomes are each independently recoverable.
PolarizationDecrease	B	Cues a broader logical outcome of reduced polarization between salient social or political groups through softened intergroup division, reduced zero-sum opposition, decreased perceived incompatibility, or increased scope for coexistence, cooperation, dialogue, or compromise. *Boundary:* A BRIDGE or other depolarising relation in the Polarization dimension does not automatically instantiate this AboutObject. Reduced polarization of the broader intergroup relation must itself be independently recoverable.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
OutgroupHostilityIncrease	B	Cues a broader logical outcome of increased fear, hatred, contempt, aversion, antipathy, or comparable hostile orientation toward an identifiable social or political outgroup. Evidence may include generalised threat, derogation, enemy construction, contempt, or other representations where the communication bears toward cultivating, intensifying, or normalising hostility toward the group. *Boundary:* Mere criticism, disagreement, negative description, insult, or hostile wording directed at a group is insufficient unless increased hostility toward that group is itself a recoverable broader outcome.
OutgroupHostilityDecrease	B	Cues a broader logical outcome of reduced fear, hatred, contempt, aversion, antipathy, or comparable hostile orientation toward an identifiable social or political outgroup. Evidence may include communication bearing toward reduced animosity, increased empathy, greater interpersonal warmth, or rejection of hostility toward the group. *Boundary:* Polite language, reduced derogation, or the absence of hostility is insufficient. Reduced hostility toward the group must itself be an independently recoverable logical outcome.
SocietalCohesionIncrease	B	Cues a broader logical outcome of strengthened cohesion within the relevant society or broad civic community through increased solidarity, mutual trust, shared belonging, reciprocal commitment, cooperation, or capacity for constructive relations across its constituent groups. *Boundary:* Stronger solidarity within one narrowly defined ingroup does not by itself instantiate societal cohesion where other constituent groups are excluded, antagonised, or positioned outside the relevant social whole.
SocietalCohesionDecrease	B	Cues a broader logical outcome of weakened cohesion within the relevant society or broad civic community through fragmentation, generalised social distrust, erosion of shared belonging or solidarity, breakdown of cooperation, or deterioration of the broader social fabric. *Boundary:* Hostility toward one group or polarization of one group relation does not automatically instantiate this label unless deterioration of the wider societal cohesion is independently recoverable.
GroupInclusionIncrease	B	Cues a broader logical outcome of increased acceptance of an identifiable social group's belonging, membership, participation, standing, rights, protection, or legitimate place within the relevant civic, social, or political community. Positive affect toward the group is not required.
GroupInclusionDecrease	B	Cues a broader logical outcome of decreased acceptance of an identifiable social group's belonging, membership, participation, standing, rights, protection, or legitimate place within the relevant civic, social, or political community. Increased hostility or dehumanization is not required.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
GroupHumanizationIncrease	B	Cues a broader logical outcome of strengthened recognition of an identifiable social group's full humanity, human dignity, mental complexity, moral standing, or capacity for ordinary human experience. The outcome may be advanced even where disagreement or conflict with the group remains.
GroupHumanizationDecrease	B	Cues a broader logical outcome of weakened recognition of an identifiable social group's full humanity, human dignity, mental complexity, moral standing, or capacity for ordinary human experience, including communication bearing toward treating the group as subhuman, non-human, animal-like, object-like, contaminating, or otherwise less than fully human. *Boundary:* The Method-level Tactic *Dehumanizing representation* does not automatically instantiate this AboutObject. The broader outcome of reducing recognition of the group's humanity or human standing must itself be independently recoverable.
PoliticalViolenceSupportIncrease	C	Cues a broader logical outcome of increased acceptance, endorsement, justification, normalisation, or perceived legitimacy of physical violence as a means of pursuing, defending, or expressing political, ideological, civic, or intergroup objectives. Mere description of violence, violent imagery, or prediction that violence will occur is insufficient unless support for or acceptability of political violence is independently advanced.
PoliticalViolenceSupportDecrease	C	Cues a broader logical outcome of reduced acceptance, endorsement, justification, normalisation, or perceived legitimacy of physical violence as a means of pursuing, defending, or expressing political, ideological, civic, or intergroup objectives. Condemnation of a particular perpetrator is insufficient unless reduced support for the relevant use of political violence is independently advanced.
PoliticalViolenceParticipationIncrease	C	Cues a broader logical outcome of increased participation in, commission of, or material contribution to physical political or intergroup violence, including calls, encouragement, instructions, mobilisation, or other communication bearing toward violent action by recipients or represented participants. *Boundary:* Support for, praise of, or justification of violence does not automatically instantiate this label. Participation or contribution to violent conduct must itself be independently recoverable as an outcome.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
PoliticalViolenceParticipationDecrease	C	Cues a broader logical outcome of reduced participation in, commission of, or material contribution to physical political or intergroup violence, including calls to refrain from violence, cease violent activity, disarm, withdraw from planned attacks, or substitute non-violent conduct. *Boundary:* Condemnation or reduced approval of violence does not automatically instantiate this label unless reduced violent participation or conduct is itself independently advanced.
TerrorismSupportIncrease	C	Cues a broader logical outcome of increased approval, endorsement, justification, normalisation, admiration, or support for terrorist violence, terrorist acts, terrorist causes, or actors represented in the available message context as engaged in terrorism. Glorification, martyr or hero representations, justification of attacks, or celebratory depictions may provide evidence where increased support is independently recoverable. *Boundary:* Do not infer or adjudicate an actor's legal status as a terrorist from outside the available annotation context. Mere reporting, quotation, historical description, or depiction of terrorist violence is insufficient.
TerrorismSupportDecrease	C	Cues a broader logical outcome of reduced approval, endorsement, justification, normalisation, admiration, or support for terrorist violence, terrorist acts, terrorist causes, or actors represented in the available message context as engaged in terrorism. Mere factual reporting of harms or condemnation without a recoverable support-reduction orientation is insufficient.
ViolentExtremismSupportIncrease	C	Cues a broader logical outcome of increased approval, endorsement, normalisation, justification, or support for a violent extremist movement, group, orientation, or its use of violence. The defining outcome is support for violent extremism rather than adoption of every component of its ideology or participation in its activities.
ViolentExtremismSupportDecrease	C	Cues a broader logical outcome of reduced approval, endorsement, normalisation, justification, or support for a violent extremist movement, group, orientation, or its use of violence. Reduced organisational engagement or abandonment of extremist doctrine is not inferred unless independently recoverable.
ViolentExtremistEngagementIncrease	C	Cues a broader logical outcome of increased behavioural involvement with a violent extremist or terrorist group, movement, network, or activity, including joining, recruitment uptake, continued participation, material contribution, training, onboarding, or deepened organisational involvement. The presence of recruitment language is evidence only where increased engagement by the relevant recipient or represented population is itself the recoverable outcome.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
ViolentExtremistEngagementDecrease	C	Cues a broader logical outcome of reduced behavioural involvement with a violent extremist or terrorist group, movement, network, or activity, including leaving, disengaging, withdrawing participation, ceasing material contribution, severing organisational ties, or remaining disengaged. *Boundary:* This label concerns behavioural disengagement and does not require abandonment of extremist beliefs. Conversely, reduced ideological support does not automatically instantiate behavioural disengagement.
MediaTrustIncrease	D	Cues increased trust or confidence in journalism, news media, or a represented class of news-producing institutions as reliable, credible, honest, or dependable sources of information.
MediaTrustDecrease	D	Cues reduced trust or confidence in journalism, news media, or a represented class of news-producing institutions as reliable, credible, honest, or dependable sources of information.
ScienceTrustIncrease	D	Cues increased trust or confidence in science, scientific expertise, scientists, or scientific institutions as reliable, credible, or dependable sources of knowledge.
ScienceTrustDecrease	D	Cues reduced trust or confidence in science, scientific expertise, scientists, or scientific institutions as reliable, credible, or dependable sources of knowledge.
InstitutionalTrustIncrease	D	Cues increased trust or confidence in a public or civic institution as reliable, competent, honest, procedurally dependable, or capable of performing its represented institutional function.
InstitutionalTrustDecrease	D	Cues reduced trust or confidence in a public or civic institution as reliable, competent, honest, procedurally dependable, or capable of performing its represented institutional function.
InstitutionalLegitimacyIncrease	D	Cues strengthened acceptance of a public or civic institution's authority, standing, rightfulness, or entitlement to exercise its represented institutional role or make binding or authoritative decisions.
InstitutionalLegitimacyDecrease	D	Cues weakened acceptance of a public or civic institution's authority, standing, rightfulness, or entitlement to exercise its represented institutional role or make binding or authoritative decisions. Mere claims of incompetence, unreliability, or factual error are insufficient unless the institution's legitimate standing or authority is also placed in question.
GeneralizedEpistemicTrustIncrease	D	Cues strengthened confidence that reliable knowledge, evidence, authentic information, or dependable means of distinguishing warranted from unwarranted claims remain available within the relevant information environment.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
GeneralizedEpistemicTrustDecrease	D	Cues erosion of confidence that reliable knowledge, evidence, authentic information, or dependable means of distinguishing warranted from unwarranted claims remain available within the relevant information environment, including broad representations that truth is unknowable, evidence is generally unreliable, or authentic and fabricated information cannot meaningfully be distinguished.
ResponsibilityAttributionShift	D	Cues movement or reassignment of causal, moral, or practical responsibility for an event, condition, or outcome away from one represented actor and toward another actor, where the reassignment of responsibility itself constitutes a recoverable logical influence outcome. Claims that an event was staged, conducted under false attribution, or deliberately made to appear to have been caused by another actor are included where they perform this attributional reassignment.
ConspiratorialWorldviewIncrease	E	Cues a broader logical outcome of strengthened generalised orientation toward explaining consequential social, political, institutional, or historical events through covert coordination, hidden powerful actors, secret plots, or systematic concealed control. A specific allegation of conspiracy or one hidden-actor causal account is insufficient unless strengthening a broader conspiratorial interpretation of social reality is independently recoverable.
ConspiratorialWorldviewDecrease	E	Cues a broader logical outcome of weakened generalised orientation toward explaining consequential social, political, institutional, or historical events through covert coordination, hidden powerful actors, secret plots, or systematic concealed control. Mere rejection or correction of one specific conspiracy claim is insufficient unless reduced reliance on the broader conspiratorial worldview is independently advanced.
AntiPluralistWorldviewIncrease	E	Cues a broader logical outcome of strengthened rejection of political or societal pluralism as a legitimate organising principle, including representations that competing viewpoints, opposition, diversity, or coexistence among legitimately different groups should be replaced by ideological, political, or social homogeneity, exclusive correctness, or monistic authority. *Boundary:* Strong ideological commitment, criticism of an opponent, or exclusion of one group is insufficient unless broader rejection of legitimate pluralism is itself independently recoverable.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
AntiPluralistWorldviewDecrease	E	Cues a broader logical outcome of weakened rejection of political or societal pluralism, including strengthened acceptance that competing viewpoints, legitimate opposition, societal diversity, and coexistence among differently positioned groups can properly form part of the relevant civic or political order. *Boundary:* Tolerance or inclusion of one particular group is insufficient unless broader acceptance of pluralism is independently recoverable.
IdeologicalCommunityEngagement-Increase	E	Cues a broader logical outcome of increased joining, following, affiliation, continued participation, or deeper engagement with a community, channel, network, or group whose represented function is to sustain, develop, or reinforce a conspiratorial, anti-pluralist, extremist, or comparable ideological worldview. *Boundary:* Ordinary civic or party participation is coded under the applicable political AboutObject. Engagement with a represented violent extremist or terrorist organisation is coded under `ViolentExtremistEngagement-Increase`. Mere exposure to or reading of ideological material is insufficient unless community affiliation or participation itself is independently advanced.
IdeologicalCommunityEngagement-Decrease	E	Cues a broader logical outcome of reduced affiliation, participation, or continued engagement with a community, channel, network, or group whose represented function is to sustain, develop, or reinforce a conspiratorial, anti-pluralist, extremist, or comparable ideological worldview, including leaving, unfollowing, disengaging, or severing community ties. *Boundary:* Reduced agreement with a worldview does not automatically instantiate community disengagement, and community exit does not automatically instantiate worldview change.
HealthGuidanceAdherenceIncrease	F	Cues a broader logical outcome of increased adherence to an identifiable health recommendation, preventive measure, treatment instruction, public-health guidance, or comparable health-directed course of action represented in the message context. *Boundary:* The label does not imply that the represented guidance is medically correct, safe, evidence-based, or beneficial. Adherence to the represented health guidance itself must be the recoverable outcome.
HealthGuidanceAdherenceDecrease	F	Cues a broader logical outcome of reduced adherence to, rejection of, evasion of, or departure from an identifiable health recommendation, preventive measure, treatment instruction, public-health guidance, or comparable health-directed course of action. *Boundary:* Distrust of a source or institution does not by itself instantiate this label unless reduced adherence to the represented health guidance is independently advanced.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
HealthCareEngagementIncrease	F	Cues a broader logical outcome of increased seeking, use, continuation, or maintenance of engagement with health-care services or professional medical care, including seeking assessment, contacting providers, presenting for care, or remaining engaged with represented care. *Boundary:* Agreement with one health recommendation is insufficient unless engagement with care or health services itself is independently advanced.
HealthCareEngagementDecrease	F	Cues a broader logical outcome of reduced, delayed, avoided, refused, or discontinued engagement with health-care services or professional medical care, including recommendations to avoid providers, delay seeking care, withdraw from treatment relationships, or refuse represented professional care. *Boundary:* Reduced trust in health professionals or institutions is insufficient unless avoidance, refusal, delay, or withdrawal from care is independently recoverable.
SubstituteRemedyAdoptionIncrease	F	Cues a broader logical outcome of increased adoption or use of a represented remedy, treatment, practice, or intervention where that option is positioned within the message as an alternative or substitute for another represented health intervention, treatment, recommendation, or professional care. *Boundary:* No judgement is made about the substitute's medical safety, efficacy, evidential status, conventionality, or legitimacy. A treatment or remedy recommendation without a recoverable substitution relation is insufficient for this label.
SubstituteRemedyAdoptionDecrease	F	Cues a broader logical outcome of reduced, discontinued, rejected, or avoided use of a represented remedy, treatment, practice, or intervention that is positioned within the message as an alternative or substitute for another represented health intervention, treatment, recommendation, or professional care. *Boundary:* Reduced use of one treatment does not instantiate this label unless the relevant substitute-remedy relation is recoverable from the message.
CrisisProtectiveActionIncrease	F	Cues a broader logical outcome of increased adoption or performance of represented protective or response action in relation to an emergency, disaster, outbreak, hazard, or comparable crisis, including evacuation, sheltering, preparedness, protective conduct, or adherence to emergency response instructions. *Boundary:* The ontology does not determine whether the represented action is objectively protective or whether the underlying hazard claim is factually correct. The crisis-response behaviour itself must be independently advanced.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
CrisisProtectiveActionDecrease	F	Cues a broader logical outcome of reduced adoption or performance of represented protective or response action in relation to an emergency, disaster, outbreak, hazard, or comparable crisis, including discouraging, refusing, evading, delaying, or abandoning evacuation, sheltering, preparedness, or other represented crisis-response conduct. *Boundary:* Doubt, distrust, confusion, or criticism of emergency authorities is insufficient unless reduced protective or response action is independently recoverable.
EmergencyServiceCapacityIncrease	F	Cues a broader logical outcome of preserving, restoring, or increasing the operational availability or effective capacity of emergency-response services, including communication bearing toward keeping emergency channels, personnel, or response resources available for genuine response needs. *Boundary:* Ordinary advice about when to contact an emergency service does not automatically instantiate this label unless preservation or improvement of service capacity is itself independently recoverable.
EmergencyServiceCapacityDecrease	F	Cues a broader logical outcome of reducing, exhausting, obstructing, or overloading the operational availability or effective capacity of emergency-response services, including coordinated or mass unnecessary calls, false dispatch demands, resource diversion, or other communication bearing toward tying up emergency-response capacity. *Boundary:* A single erroneous request for assistance does not automatically instantiate this label. Reduced emergency-service capacity must itself be independently recoverable as part of the broader logical outcome.
EnvironmentalProtectionAction-Increase	G	Cues a broader logical outcome of increased individual, collective, organisational, institutional, or community behaviour represented as protecting, conserving, restoring, or reducing harm to the natural environment, including pollution reduction, waste reduction, habitat or species protection, resource stewardship, or comparable non-climate environmental action. *Boundary:* Where the identical behavioural endpoint is specifically represented as mitigation of climate change or adjustment to climate impacts, use the more specific `ClimateMitigationAction-*` or `ClimateAdaptationAction-*` label rather than duplicating the endpoint. Support for an environmental policy is coded as the applicable policy stance rather than environmental action unless behaviour itself is independently advanced.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
EnvironmentalProtectionAction-Decrease	G	Cues a broader logical outcome of reduced, discontinued, avoided, reversed, or discouraged individual, collective, organisational, institutional, or community behaviour represented as protecting, conserving, restoring, or reducing harm to the natural environment. *Boundary:* Opposition to an environmental claim or policy is insufficient unless reduced environmental-protection behaviour itself is independently recoverable.
ClimateChangeAcceptanceIncrease	G	Cues a broader logical outcome of increased acceptance of substantive represented claims concerning climate change, including its occurrence, anthropogenic contribution, or comparable core claims about the existence and causal character of climate change. *Boundary:* Trust in climate scientists, scientific expertise, or scientific institutions does not automatically instantiate this label. Acceptance of the climate-change proposition itself must be independently recoverable.
ClimateChangeAcceptanceDecrease	G	Cues a broader logical outcome of reduced acceptance or increased rejection of substantive represented claims concerning climate change, including its occurrence, anthropogenic contribution, or comparable core claims about the existence and causal character of climate change. *Boundary:* Criticism or distrust of scientists does not automatically instantiate this label unless rejection or reduced acceptance of the climate-change proposition itself is independently recoverable.
ClimatePolicySupport	G	Cues a broader logical outcome of increased support for an identifiable governmental, regulatory, institutional, or public-policy response to climate change, including represented mitigation, adaptation, resilience, energy, emissions, or comparable climate-response measures.
ClimatePolicyOpposition	G	Cues a broader logical outcome of increased opposition to, rejection of, resistance to, repeal of, or prevention of an identifiable governmental, regulatory, institutional, or public-policy response to climate change. *Boundary:* General acceptance or rejection of climate change does not automatically instantiate a policy stance. The represented climate policy or policy course must itself be independently supported or opposed.
ClimateMitigationActionIncrease	G	Cues a broader logical outcome of increased individual, collective, organisational, or institutional action represented as reducing, limiting, avoiding, or otherwise mitigating contributions to climate change or greenhouse-gas emissions. *Boundary:* Support for a mitigation policy does not automatically instantiate this behavioural outcome unless mitigation action itself is independently advanced.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
ClimateMitigationActionDecrease	G	Cues a broader logical outcome of reduced, discontinued, avoided, reversed, or rejected individual, collective, organisational, or institutional action represented as reducing, limiting, avoiding, or otherwise mitigating contributions to climate change or greenhouse-gas emissions. *Boundary:* Opposition to a climate policy is insufficient unless reduced mitigation action itself is independently recoverable.
ClimateAdaptationActionIncrease	G	Cues a broader logical outcome of increased individual, collective, organisational, institutional, or infrastructural adjustment to represented actual or expected climate effects in order to cope with, prepare for, or respond to those effects. *Boundary:* The label concerns adoption or performance of adaptation action rather than merely favourable evaluation of adaptation in the abstract.
ClimateAdaptationActionDecrease	G	Cues a broader logical outcome of reduced, discontinued, avoided, delayed, or rejected individual, collective, organisational, institutional, or infrastructural adjustment to represented actual or expected climate effects. *Boundary:* Criticism of one adaptation measure does not automatically instantiate this label unless reduced adaptation action itself is independently recoverable.
ClimateResilienceIncrease	G	Cues a broader logical outcome of strengthened capacity of a relevant social, institutional, economic, infrastructural, or ecological system to cope with, withstand, reorganise under, recover from, or continue essential functioning during climate-related hazards, trends, or disturbances while maintaining capacity for adaptation, learning, or transformation. *Boundary:* A single mitigation, adaptation, or protective action does not automatically instantiate this label. Strengthening broader system capacity or resilience must itself be independently recoverable.
ClimateResilienceDecrease	G	Cues a broader logical outcome of weakened capacity of a relevant social, institutional, economic, infrastructural, or ecological system to cope with, withstand, reorganise under, recover from, or continue essential functioning during climate-related hazards, trends, or disturbances. *Boundary:* Failure, refusal, or discouragement of one adaptation action is insufficient unless weakening of broader resilience capacity is itself independently recoverable.
EconomicConfidenceIncrease	H	Cues a broader logical outcome of increased confidence in the condition, stability, prospects, functioning, or capacity for continuation or recovery of the relevant economy. The ontology does not adjudicate the factual accuracy of represented economic conditions or forecasts.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
EconomicConfidenceDecrease	H	Cues a broader logical outcome of reduced confidence in the condition, stability, prospects, functioning, or capacity for continuation or recovery of the relevant economy, including representations bearing toward generalised economic insecurity, collapse expectations, or loss of confidence in economic functioning. *Boundary:* Criticism of a government, firm, policy, or institution is insufficient unless confidence in the broader economy itself is independently affected.
FinancialSystemConfidenceIncrease	H	Cues a broader logical outcome of increased confidence in the safety, stability, reliability, liquidity, solvency, or continued functioning of a represented bank, banking sector, deposit system, or wider financial system where that financial confidence itself constitutes the outcome.
FinancialSystemConfidenceDecrease	H	Cues a broader logical outcome of reduced confidence in the safety, stability, reliability, liquidity, solvency, or continued functioning of a represented bank, banking sector, deposit system, or wider financial system. *Boundary:* Reduced financial confidence does not automatically instantiate deposit-run participation. Withdrawal or run behaviour must be independently recoverable.
EssentialResourceStockpilingIncrease	H	Cues a broader logical outcome of increased accumulation, bulk acquisition, hoarding, or stockpiling of essential goods or resources in response to a represented shortage, disruption, crisis, scarcity, or expected loss of availability. *Boundary:* No inference is made that the behaviour is irrational, emotionally driven, excessive in an objective sense, or caused by panic. Ordinary purchasing or inventory accumulation without the represented shortage or stockpiling relation is insufficient.
EssentialResourceStockpilingDecrease	H	Cues a broader logical outcome of reduced, avoided, reversed, or discouraged stockpiling of essential goods or resources in a represented shortage, disruption, crisis, scarcity, or run context, including communication bearing toward maintenance or restoration of ordinary acquisition behaviour.
DepositRunParticipationIncrease	H	Cues a broader logical outcome of increased widespread, coordinated, or run-like withdrawal, transfer, or removal of deposits or comparable liquid financial holdings from a represented bank or financial system because continued availability, safety, or stability is placed in question. *Boundary:* Ordinary individual withdrawal or portfolio advice is insufficient. The message must bear toward participation in or amplification of a broader deposit-run or run-like withdrawal dynamic.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
DepositRunParticipationDecrease	H	Cues a broader logical outcome of reduced participation in, prevention of, or withdrawal from a represented deposit-run or run-like mass-withdrawal dynamic, including communication bearing toward keeping deposits in place or discouraging destabilising mass withdrawal. *Boundary:* Increased financial confidence alone does not automatically instantiate this behavioural outcome unless reduced run participation is independently advanced.
CriticalInfrastructureConfidence-Increase	H	Cues a broader logical outcome of increased confidence that represented critical infrastructure or essential services are dependable, available, capable of continued operation, or capable of maintaining or restoring their essential service functions.
CriticalInfrastructureConfidence-Decrease	H	Cues a broader logical outcome of reduced confidence that represented critical infrastructure or essential services are dependable, available, capable of continued operation, or capable of maintaining or restoring their essential service functions. *Boundary:* A report of one bounded outage or disruption is insufficient unless broader confidence in the relevant infrastructure or service continuity is independently affected.
CriticalInfrastructureResilienceIncrease	H	Cues a broader logical outcome of strengthened capacity of represented critical infrastructure or essential-service systems to withstand, absorb, adapt to, recover from, or maintain essential functioning during disruption, including increased redundancy, preparedness, recovery capacity, or other system-level resilience resources. *Boundary:* Increased confidence in infrastructure does not itself instantiate resilience increase. Strengthening system capacity must be independently recoverable.
CriticalInfrastructureResilience-Decrease	H	Cues a broader logical outcome of weakened capacity of represented critical infrastructure or essential-service systems to withstand, absorb, adapt to, recover from, or maintain essential functioning during disruption, including degradation of redundancy, preparedness, recovery capacity, or other system-level resilience resources. *Boundary:* Reduced confidence or prediction of failure is insufficient unless weakening of the underlying resilience capacity is itself independently recoverable.
InformationResilienceIncrease	I	Cues a strengthened capacity of persons, groups, institutions, or societies to recognise, critically assess, resist, or recover from manipulative, misleading, or otherwise challenging influence in the information environment. Evidence may include development of transferable recognition skills, verification practices, critical-evaluation habits, resistance strategies, or other resources that increase preparedness for subsequent influence attempts. Mere rebuttal of a particular claim or defence of a particular position is insufficient unless broader resistance capacity is independently advanced.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
InformationResilienceDecrease	I	Cues a weakened capacity of persons, groups, institutions, or societies to recognise, critically assess, resist, or recover from manipulative, misleading, or otherwise challenging influence in the information environment. Evidence may include discouraging or degrading transferable verification, evaluation, recognition, or resistance practices; undermining confidence in the usefulness of such defensive capacities; or otherwise orienting recipients toward reduced preparedness for subsequent influence attempts. The mere use of manipulation, deception, distraction, or another influence tactic does not by itself license this label.
InformationAvoidanceIncrease	I	Cues increased disengagement from, avoidance of, or deliberate non-exposure to relevant information or information sources, including orientations to stop following, reading, watching, seeking, receiving, or otherwise engaging with information. Reduced trust, low resilience, confusion, or overload alone is insufficient unless information withdrawal or avoidance is independently advanced as a substantive logical outcome.
InformationSeekingIncrease	I	Cues a broader logical outcome of increased active seeking, consulting, retrieving, comparing, monitoring, or otherwise acquiring relevant information or information sources, including orientations to obtain additional evidence, consult further sources, or actively verify a represented claim. *Boundary:* Mere receipt or exposure to information is insufficient. A specific information-seeking or verification behaviour does not automatically instantiate `InformationResilience-Increase`; resilience requires strengthening transferable capacity for subsequent influence situations. Likewise, increased information seeking does not require reduced `InformationAvoidance-Increase` because seeking and avoidance may coexist across different sources or information classes.
MilitaryForceSupportIncrease	J	Cues a broader logical outcome of increased support for the use, continuation, escalation, authorisation, or deployment of organised military force in an international or foreign-policy context, including military intervention or defence of another state or ally. *Boundary:* General support for a conflict party, alliance, foreign aid, or sanctions does not automatically instantiate this label. Support for the use of military force itself must be independently recoverable.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
MilitaryForceSupportDecrease	J	Cues a broader logical outcome of reduced support for the use, continuation, escalation, authorisation, or deployment of organised military force in an international or foreign-policy context, including opposition to intervention, calls for military withdrawal, or rejection of continued military operations. *Boundary:* Reduced support for one conflict party does not by itself instantiate reduced support for military force unless the military-policy outcome is independently advanced.
MilitaryServiceParticipationIncrease	J	Cues a broader logical outcome of increased enlistment, enrolment, reenlistment, reporting for duty, continued service, or comparable behavioural participation in the regular armed forces of a represented state or other recognised military authority. *Boundary:* Support for military intervention or war does not automatically instantiate military-service participation. The message must independently bear toward joining, entering, remaining in, or reporting for military service. Where the represented organisation is a violent-extremist or terrorist group rather than regular military service, use the applicable Domain C engagement label.
MilitaryServiceParticipationDecrease	J	Cues a broader logical outcome of reduced enlistment, refusal or avoidance of entry, withdrawal, non-renewal, departure, or reduced continued participation in the regular armed forces of a represented state or other recognised military authority. *Boundary:* Opposition to the use of military force does not automatically instantiate this label. Reduced military-service participation itself must be independently recoverable. No legal judgement concerning the status of refusal, withdrawal, or departure is implied.
ForeignAidSupportIncrease	J	Cues a broader logical outcome of increased support for providing, maintaining, expanding, or authorising foreign assistance to an identifiable foreign state, population, ally, conflict party, or other international recipient, including represented humanitarian, economic, military, or comparable assistance. *Boundary:* General sympathy or support for the recipient does not automatically instantiate this label unless provision of foreign assistance itself is independently advanced.
ForeignAidSupportDecrease	J	Cues a broader logical outcome of reduced support for, reduction, termination, suspension, withholding, or prevention of foreign assistance to an identifiable foreign state, population, ally, conflict party, or other international recipient. *Boundary:* Opposition to a recipient or conflict party does not automatically instantiate this label unless the foreign-assistance outcome is independently recoverable.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
SanctionsSupportIncrease	J	Cues a broader logical outcome of increased support for imposing, maintaining, expanding, or strengthening economic, diplomatic, financial, trade, or comparable sanctions against an identifiable foreign actor, state, entity, or target. *Boundary:* Hostility toward or opposition to the target is insufficient unless sanctions themselves are independently advanced as the relevant foreign-policy outcome.
SanctionsSupportDecrease	J	Cues a broader logical outcome of reduced support for, relaxation, removal, non-imposition, suspension, or weakening of economic, diplomatic, financial, trade, or comparable sanctions against an identifiable foreign actor, state, entity, or target.
ConflictPartySupportIncrease	J	Cues a broader logical outcome of increased approval, sympathy, political support, solidarity, identification, or alignment toward an identifiable state, government, movement, or other represented party to an international or geopolitical conflict. *Boundary:* Support for one policy benefiting the actor does not automatically instantiate this label unless broader support or alignment toward the conflict party itself is independently recoverable. Likewise, support for the actor does not automatically imply support for military force, foreign assistance, or sanctions.
ConflictPartySupportDecrease	J	Cues a broader logical outcome of reduced approval, sympathy, political support, solidarity, identification, or alignment toward an identifiable state, government, movement, or other represented party to an international or geopolitical conflict. *Boundary:* Opposition to one action or policy of the actor is insufficient unless reduced support or alignment toward the actor itself is independently recoverable.
AllianceSupportIncrease	J	Cues a broader logical outcome of increased support for establishing, maintaining, strengthening, participating in, fulfilling, or continuing an alliance or comparable international security alignment. *Boundary:* Confidence that an ally is reliable does not automatically instantiate alliance support. The alliance or alignment relationship itself must be independently advanced as an outcome.
AllianceSupportDecrease	J	Cues a broader logical outcome of reduced support for, weakening, withdrawal from, termination of, non-participation in, or rejection of an alliance or comparable international security alignment. *Boundary:* Doubt about an ally's reliability is insufficient unless weaker support for the alliance or alignment itself is independently recoverable.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
AllianceReliabilityConfidenceIncrease	J	Cues a broader logical outcome of increased confidence that an identifiable ally, security partner, or alliance relationship is reliable, dependable, capable of fulfilling relevant commitments, or likely to honour represented mutual-defence or security obligations. *Boundary:* This label concerns confidence in the reliability of the alliance relationship or partner rather than support for alliance membership or a particular military intervention.
AllianceReliabilityConfidenceDecrease	J	Cues a broader logical outcome of reduced confidence that an identifiable ally, security partner, or alliance relationship is reliable, dependable, capable of fulfilling relevant commitments, or likely to honour represented mutual-defence or security obligations. *Boundary:* Reduced confidence does not automatically instantiate `AllianceSupport-Decrease`, and opposition to an alliance does not automatically imply that its members are represented as unreliable.
PersonalPrivacyIncrease	K	Cues a broader logical outcome of strengthened privacy of identifiable persons or populations, including increased protection, confidentiality, anonymity, control over disclosure, or restriction of unwanted access to, dissemination of, or exposure of personal identifying information. *Boundary:* Generic privacy rhetoric is insufficient unless stronger protection or control of personal information or identifiability is itself independently recoverable as a logical outcome.
PersonalPrivacyDecrease	K	Cues a broader logical outcome of weakened privacy of identifiable persons or populations, including reduced confidentiality, anonymity, or control over personal identifying information; increased unwanted disclosure, dissemination, accessibility, identification, tracing, or exposure may provide evidence for this outcome. *Boundary:* A conditional threat to disclose, identify, track, or expose a person does not automatically instantiate this label where the threatened privacy loss functions only as leverage for another outcome. Reduced personal privacy itself must be independently recoverable as part of the broader logical outcome toward which the communication bears.
PersonalSecurityIncrease	K	Cues a broader logical outcome of strengthened security of identifiable persons or populations against targeting, intimidation, retaliation, harassment, exposure-related harm, or comparable threats to personal safety or security. *Boundary:* Reassurance, positive sentiment, or the absence of a threat is insufficient. Strengthening or protecting personal security itself must be independently recoverable as a logical outcome.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
PersonalSecurityDecrease	K	Cues a broader logical outcome of weakened security of identifiable persons or populations against targeting, intimidation, retaliation, harassment, exposure-related harm, or comparable threats to personal safety or security. *Boundary:* The presence of a threat, insult, hostile statement, or coercive Method does not automatically instantiate this label. Reduced personal security itself must be independently recoverable as an outcome, rather than merely functioning as a threatened consequence used to obtain a different result.
PublicExpressionIncrease	K	Cues a broader logical outcome of increased or sustained public expression, including speaking, publishing, reporting, communicating views or information, or otherwise participating in public discourse despite or in relation to represented pressure, intimidation, surveillance, censorship, or anticipated consequences. *Boundary:* General civic participation is insufficient unless public expression itself is independently advanced. Mere absence of censorship or intimidation does not constitute evidence of increased expression.
PublicExpressionDecrease	K	Cues a broader logical outcome of reduced, withdrawn, suppressed, deterred, or self-censored public expression, including communication bearing toward silencing, withholding views or information, ceasing publication or reporting, or avoiding public communication because of represented pressure, intimidation, surveillance, retaliation, or anticipated consequences. *Boundary:* Threat, harassment, criticism, reputational attack, or coercion alone is insufficient unless reduced public expression itself is independently recoverable as an outcome toward which the communication bears. Where the message instead seeks another substantive behaviour through coercion, code that corresponding AboutObject rather than inferring expression suppression.
CommercialInquiryIncrease	L	Cues a broader logical outcome of increased commercial inquiry or prospective- customer contact, including requesting information, a quotation, consultation, demonstration, sales contact, appointment, or comparable pre-purchase engagement with a commercial provider. *Boundary:* Mere exposure, awareness, website visitation, or information receipt is insufficient unless commercial inquiry or contact itself is independently advanced.
CommercialInquiryDecrease	L	Cues a broader logical outcome of reduced, avoided, cancelled, or discouraged commercial inquiry or prospective-customer contact, including orientations away from requesting information, quotations, demonstrations, consultations, appointments, or sales contact.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
PurchaseBehaviorIncrease	L	Cues a broader logical outcome of increased purchasing, ordering, acquisition, or commercial patronage of an identifiable product or service. *Boundary:* Inquiry, trial, subscription, installation, recommendation, or favourable evaluation does not automatically instantiate purchase behaviour unless purchase or patronage itself is independently advanced.
PurchaseBehaviorDecrease	L	Cues a broader logical outcome of reduced, avoided, discontinued, or rejected purchasing, ordering, acquisition, or commercial patronage of an identifiable product, service, provider, or brand. *Boundary:* Boycott, switching, avoidance, or anti-purchase appeals may provide evidence where reduced commercial patronage is independently recoverable; they do not require a separate boycott AboutObject.
SubscriptionEnrollmentIncrease	L	Cues a broader logical outcome of increased enrolment or opt-in to an ongoing commercial subscription, membership, recurring service, or comparable continuing commercial arrangement.
SubscriptionEnrollmentDecrease	L	Cues a broader logical outcome of reduced, avoided, rejected, or cancelled prospective enrolment or opt-in to an ongoing commercial subscription, membership, recurring service, or comparable continuing commercial arrangement. *Boundary:* Termination of an already active subscription is ordinarily a customer-retention outcome rather than subscription-enrollment decrease.
AppInstallationIncrease	L	Cues a broader logical outcome of increased installation or download of an identifiable commercial software application or comparable digital product.
AppInstallationDecrease	L	Cues a broader logical outcome of reduced, avoided, or rejected installation or download of an identifiable commercial software application or comparable digital product. *Boundary:* Removal of an already installed application is not treated as reduced installation unless the message independently bears on future installation behaviour; where removal terminates an ongoing customer relationship, the applicable retention outcome may instead apply.
TrialEnrollmentIncrease	L	Cues a broader logical outcome of increased enrolment, registration, or participation in a represented commercial trial, sample period, demonstration period, or comparable pre-paid or pre-conversion product experience.
TrialEnrollmentDecrease	L	Cues a broader logical outcome of reduced, avoided, rejected, or discontinued prospective enrolment in a represented commercial trial, sample period, demonstration period, or comparable pre-conversion product experience.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
CustomerRetentionIncrease	L	Cues a broader logical outcome of increased continuation or preservation of an existing commercial customer relationship, including reduced cancellation, defection, churn, switching away, or termination. *Boundary:* Continued patronage does not automatically instantiate `BrandLoyalty-Increase`. Retention may result from convenience, contractual constraints, switching costs, incentives, or other reasons without stronger preferential commitment to the brand.
CustomerRetentionDecrease	L	Cues a broader logical outcome of reduced continuation or increased termination, cancellation, defection, churn, or switching away from an existing commercial customer relationship. *Boundary:* Reduced loyalty or negative brand evaluation is insufficient unless termination or departure from the existing relationship itself is independently recoverable.
BrandTrustIncrease	L	Cues a broader logical outcome of increased trust or confidence in an identifiable commercial brand, provider, product, or service as reliable, dependable, honest, competent, or likely to fulfil represented commercial promises or obligations. *Boundary:* Brand trust is distinct from brand loyalty. A consumer may trust a brand without preferentially choosing it, and may remain with a brand because of switching costs, convenience, or contractual constraints without increased trust. Purchase, retention, or favourable evaluation alone is insufficient unless stronger commercial trust itself is independently recoverable.
BrandTrustDecrease	L	Cues a broader logical outcome of reduced trust or confidence in an identifiable commercial brand, provider, product, or service as reliable, dependable, honest, competent, or likely to fulfil represented commercial promises or obligations. *Boundary:* One complaint, rejected purchase, or reduced loyalty is insufficient unless weaker trust in the commercial brand or provider itself is independently recoverable. System-level confidence in banks, deposits, or the financial system is coded under Domain H rather than duplicated here for the identical endpoint.
CustomerReactivationIncrease	L	Cues a broader logical outcome of renewed commercial engagement by persons or customers represented as having previously lapsed, left, cancelled, become inactive, or otherwise ended the relevant commercial relationship. *Boundary:* Acquisition of a person not represented as a former customer is not reactivation. Continued engagement by an existing active customer is a retention rather than reactivation outcome.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
CustomerReactivationDecrease	L	Cues a broader logical outcome of reduced, avoided, or rejected return to a commercial relationship represented as previously lapsed, ended, cancelled, or inactive. *Boundary:* General avoidance of a brand is insufficient unless non-reactivation of a represented former customer relationship is itself independently recoverable.
BrandLoyaltyIncrease	L	Cues a broader logical outcome of strengthened durable preference, commitment, attachment, or propensity to continue favouring an identifiable brand, provider, product, or service over alternatives. *Boundary:* One purchase, continued enrolment, or one recommendation is insufficient unless stronger preferential commitment or loyalty itself is independently advanced.
BrandLoyaltyDecrease	L	Cues a broader logical outcome of weakened durable preference, commitment, attachment, or propensity to continue favouring an identifiable brand, provider, product, or service over alternatives. *Boundary:* One rejected purchase or isolated complaint is insufficient unless weakening the broader loyalty relationship is independently recoverable.
BrandRecommendationIncrease	L	Cues a broader logical outcome of increased recommending, referring, endorsing to prospective customers, or otherwise encouraging others to choose an identifiable commercial brand, product, or service. *Boundary:* General brand discussion, review production, or public brand advocacy does not automatically instantiate recommendation unless an interpersonal or audience-directed recommendation outcome is independently recoverable.
BrandRecommendationDecrease	L	Cues a broader logical outcome of reduced, withdrawn, discouraged, or rejected recommendation or referral of an identifiable commercial brand, product, or service to others.
ReviewGenerationIncrease	L	Cues a broader logical outcome of increased production or submission of public consumer ratings, reviews, testimonials, or comparable evaluative contributions concerning an identifiable commercial brand, product, service, or provider. *Boundary:* This label records review production rather than the truth, fairness, authenticity, or valence of the requested review.
ReviewGenerationDecrease	L	Cues a broader logical outcome of reduced, withheld, removed, discouraged, or prevented production or submission of public consumer ratings, reviews, testimonials, or comparable evaluative contributions.

Continued on next page

Supplementary Table S15: L3 label catalogue (continued)

L3 label	L2	Cue-based definition
BrandAdvocacyIncrease	L	Cues a broader logical outcome of increased active public advocacy for an identifiable commercial brand, product, service, or provider, including sustained endorsement, public defence, ambassador-like promotion, or active representation extending beyond an isolated purchase, recommendation, or review. *Boundary:* Ordinary favourable evaluation, recommendation, referral, or one review is insufficient unless broader active advocacy itself is independently advanced.
BrandAdvocacyDecrease	L	Cues a broader logical outcome of reduced, withdrawn, discouraged, or terminated active public advocacy, defence, ambassador-like promotion, or representation of an identifiable commercial brand, product, service, or provider.
`Unclear`	–	Use only where a broader civic or commercial influence-object orientation is recoverable from the message but the available message-internal evidence is insufficient to select a specific L3 label. Do not assign `Unclear` where no AboutObject is recoverable.

F.4 AboutObject: Evidence templates (L4) and selection rules

Interpretation (indicator logic). Evidence templates are measurement indicators: observable cue families that justify instantiating one or more L3 labels. Templates are not outcomes and are not lexical trigger lists. Multiple templates may be selected for the same message. Hybrid messages are represented by multiple instantiated L3 labels spanning civic domains (A–K) and commercial domain (L).

Annotation and selection rules. Annotators assign an AboutObject label only when the available message contains message-internal evidence of an orientation toward bringing about, maintaining, preventing, weakening, strengthening, or otherwise changing the outcome named by that label. The relevant relation may be expressed through an explicit call to action, behavioural instruction, stance, endorsement or rejection, trust or legitimacy positioning, or another representation whose outcome orientation is recoverable from the communication itself. Mere mention of an issue, actor, institution, behaviour, or outcome is insufficient. Likewise, the presence of a Method Tactic, TargetConstruct, NarrativeDevice, Polarization motif, or TopicDomain that could plausibly contribute to an outcome does not by itself license an AboutObject label. Annotators do not infer likely audience effects, downstream consequences, campaign objectives, or the communicator's private intention. Where several AboutObject outcomes are indepen-

dently recoverable, all corresponding L3 labels may be assigned. Do not add an additional label merely because it would be a plausible downstream consequence of another coded outcome. Where an evidence template is associated with more than one possible L3 label, instantiate only the label or labels whose defining directional relation is independently evidenced in the message. Quoted, reported, or mentioned material contributes AboutObject evidence only where the message deploys that material as part of its own outcome orientation; mere reporting, criticism, negation, or metalinguistic mention is insufficient. Use `Unclear` only where a broader influence-object orientation is recoverable from the message but the available evidence is insufficient to select a specific L3 label.

Supplementary Table S16: L4 evidence templates: selection guidance and typical cue families.

Evidence template (L4)	**Instantiates L3 label(s)**	**When to select / typical cues**
Voter registration mobilization	VoterRegistrationIncrease	Explicit encouragement to register or check registration; registration deadlines, locations, procedural guidance, or efficacy framing oriented toward registration (Brennen et al., 2020).
Voter registration demobilization	VoterRegistrationDecrease	Discouragement from registering; registration boycott or avoidance; procedural deterrence or claims oriented toward reducing registration. Do not infer turnout decrease unless voting participation is independently targeted.
Turnout mobilization	VoteTurnoutIncrease	Explicit encouragement to vote; voting dates, locations, procedures, mobilising efficacy cues, or calls to cast a ballot (Brennen et al., 2020).
Turnout demobilization	VoteTurnoutDecrease	"Your vote won't matter"; stay-home or abstention appeals; electoral boycott; fatalism; discouraging or deterrent procedural representations oriented toward reduced voting participation (Kim et al., 2018; Brennen et al., 2020).
Political-actor support positioning	PoliticalActorSupportIncrease; PoliticalActorSupportDecrease	Endorsements, voting recommendations, approval or rejection appeals, comparative evaluations, leadership-support or leadership-rejection cues, or other message-internal evidence bearing toward greater or lesser support for an identifiable domestic political leader, candidate, party, governing coalition, or comparable political actor. Policy discussion alone is insufficient unless support for the actor itself is independently advanced.

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Policy stance	PolicySupport; PolicyOpposition	Explicit or recoverable orientation toward adoption, maintenance, implementation, rejection, repeal, or resistance to a law, regulation, governmental measure, or policy course.
Protest mobilization / demobilization	ProtestParticipationIncrease; ProtestParticipationDecrease	Calls, logistics, encouragement, deterrence, withdrawal appeals, or other orientation toward increasing or decreasing participation in demonstrations, rallies, marches, strikes, or comparable protest action.
Broader civic engagement / withdrawal	CivicEngagementIncrease; CivicEngagementDecrease	Orientation toward broader participation in or withdrawal from civic and democratic channels, including contacting officials, petitions, civic or political organisations, volunteering, campaigning, public consultation, or comparable participation. Do not instantiate the broad label solely from a specific voting or protest cue unless broader engagement or disengagement is independently recoverable.
Election legitimacy positioning	ElectionLegitimacyIncrease; ElectionLegitimacyDecrease	Claims or representations strengthening or weakening acceptance of an electoral process or outcome as valid, fair, properly conducted, authoritative, or legitimate; includes certified-result acceptance as well as "rigged", "stolen", fraudulent, or invalid-election representations (Kavanagh and Rich, 2018).
Political-system legitimacy positioning	PoliticalSystemLegitimacyIncrease; PoliticalSystemLegitimacyDecrease	Claims or representations bearing toward stronger or weaker acceptance of the relevant political system, regime, constitutional order, or form of government as rightful, appropriate, entitled to govern, or worthy of continued political recognition. Approval of an incumbent, legitimacy of one election, or trust in one institution is insufficient unless legitimacy of the broader political order is independently recoverable (Anderson and Just, 2013).

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Policy compliance / non-compliance	PolicyComplianceIncrease; PolicyComplianceDecrease	Instructions, appeals, recommendations, encouragement, discouragement, evasion guidance, refusal appeals, or comparable evidence bearing toward greater or lesser behavioural conformity with an identifiable public rule, law, regulation, requirement, obligation, or policy implementation measure. Policy stance alone is insufficient (Dowling and Legrand, 2023).
Collective-identity strengthening / weakening	CollectiveIdentityStrengthening; CollectiveIdentityWeakening	Message-internal evidence that stronger or weaker collective identification, belonging, attachment, or solidarity is itself a broader logical outcome. Mere shared-identity rhetoric or group reference is insufficient.
Collective-identity reconfiguration	CollectiveIdentityReconfiguration	Claims redefining who belongs to a collective or what membership means, requires, permits, or obliges, where alteration of collective identity itself is the recoverable outcome.
Polarization increase / decrease	PolarizationIncrease; PolarizationDecrease	Message-internal evidence bearing toward harder or softer intergroup division, zero-sum opposition, incompatibility, coexistence, dialogue, cooperation, or compromise. A Polarization-dimension Divide, Cohere, or Bridge annotation is supporting but not sufficient evidence by itself (Kavanagh and Rich, 2018; EUvsDisinfo, n.d.).
Outgroup hostility increase / decrease	OutgroupHostilityIncrease; OutgroupHostilityDecrease	Message-internal evidence bearing toward greater or lesser fear, hatred, contempt, aversion, antipathy, empathy, warmth, or animosity toward an identifiable outgroup. Mere negative or positive wording about a group is insufficient unless alteration of hostility toward the group is itself the broader logical outcome (United Nations, 2019).
Societal cohesion increase / decrease	SocietalCohesionIncrease; SocietalCohesionDecrease	Message-internal evidence bearing toward stronger or weaker society-level solidarity, mutual trust, shared belonging, reciprocal commitment, cooperation, or integrity of the broader social fabric. Ingroup solidarity alone is insufficient where the broader society is simultaneously fragmented or excluded.

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Group inclusion / exclusion	GroupInclusionIncrease; GroupInclusionDecrease	Message-internal evidence bearing toward increased or decreased acceptance of an identifiable group's belonging, membership, participation, standing, rights, protection, or legitimate place within the relevant community. Hostility and humanization outcomes are coded separately where independently recoverable.
Group humanization / dehumanization	GroupHumanizationIncrease; GroupHumanizationDecrease	Message-internal evidence bearing toward increased or decreased recognition of an identifiable group's full humanity, dignity, mental complexity, moral standing, or ordinary human experience. A dehumanizing representation is not sufficient by itself unless reduced recognition of the group's humanity is also recoverable as the broader logical outcome (United Nations, 2019).
Political-violence support positioning	PoliticalViolenceSupportIncrease; PoliticalViolenceSupportDecrease	Representations increasing or decreasing the acceptability, justification, normalisation, endorsement, or legitimacy of physical violence as a political, ideological, civic, or intergroup means. Mere depiction of violence is insufficient.
Violent-action mobilization / demobilization	PoliticalViolenceParticipationIncrease; PoliticalViolenceParticipationDecrease	Calls, instructions, mobilization, encouragement, deterrence, cessation appeals, disarmament appeals, or comparable representations bearing toward increased or decreased participation in physical political or intergroup violence. Support for violence alone is insufficient unless violent conduct is independently advanced.
Terrorism support / rejection	TerrorismSupportIncrease; TerrorismSupportDecrease	Glorification, hero or martyr framing, celebration, justification, normalisation, condemnation, rejection, or comparable representations where increased or decreased support for terrorist violence, causes, acts, or represented terrorist actors is itself the broader logical outcome. Mere reporting or quotation is insufficient.

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Violent-extremism support positioning	ViolentExtremismSupportIncrease; ViolentExtremismSupportDecrease	Representations increasing or decreasing approval, normalisation, justification, endorsement, or support for violent extremist groups, movements, orientations, or their use of violence. Ideological worldview and organisational engagement are coded separately where independently recoverable.
Extremist recruitment / engagement	ViolentExtremistEngagement-Increase	Join, follow, contact, onboard, train, contribute, remain active, or otherwise increase behavioural involvement with violent extremist or terrorist actors, networks, groups, or activity (McCauley and Moskalenko, 2017).
Extremist disengagement	ViolentExtremistEngagement-Decrease	Leave, withdraw, cease participation, sever organisational ties, stop contributing, reject renewed involvement, or otherwise reduce behavioural engagement with violent extremist or terrorist actors, networks, groups, or activity. Behavioural disengagement does not require ideological conversion.
Media trust positioning	MediaTrustIncrease; MediaTrustDecrease	Claims or representations positioning journalism or news media as reliable, credible, honest, dependable, corrupt, deceptive, fabricated, or generally untrustworthy. Select the directional label independently from the represented orientation (Ognyanova et al., 2020; Marwick and Lewis, 2017).
Science trust positioning	ScienceTrustIncrease; ScienceTrustDecrease	Claims or representations positioning scientific expertise, scientists, scientific institutions, or scientific consensus as reliable and credible or as dishonest, corrupt, fabricated, systematically unreliable, or unworthy of trust (Lazer et al., 2018).
Institutional trust positioning	InstitutionalTrustIncrease; InstitutionalTrustDecrease	Claims or representations positioning public or civic institutions as reliable, competent, honest, procedurally dependable, captured, corrupt, incompetent, deceptive, or otherwise trustworthy or untrustworthy (Kavanagh and Rich, 2018).

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Institutional legitimacy positioning	InstitutionalLegitimacyIncrease; InstitutionalLegitimacyDecrease	Claims affirming or denying an institution's rightful authority, standing, entitlement to govern, adjudicate, regulate, represent, or otherwise exercise its institutional role. Mere performance criticism does not license a legitimacy label unless rightful standing or authority is independently at issue.
Generalized epistemic confidence	GeneralizedEpistemicTrustIncrease; GeneralizedEpistemicTrustDecrease	Broad claims about whether reliable knowledge, authentic evidence, or dependable means of distinguishing warranted from unwarranted information are available at all; includes "nothing is real", "nobody can know", total source-collapse claims, or affirmative claims that reliable verification and knowledge remain possible.
Authenticity / synthetic-media uncertainty	GeneralizedEpistemicTrustDecrease	Broad use of synthetic-media ambiguity, unverifiability, or authenticity collapse to advance the conclusion that reliable distinctions between authentic and fabricated information cannot be made. Mere uncertainty about the authenticity of one item is insufficient unless the broader epistemic outcome is independently recoverable (Diel et al., 2024).
Responsibility attribution reassignment	ResponsibilityAttributionShift	Explicit movement of responsibility from one represented actor toward another; counter-attribution; "they did it", "inside job", "they framed us", or false-flag representations where the communicated outcome is reassignment of who is understood as responsible (Kavanagh and Rich, 2018; EUvsDisinfo, n.d.).
Conspiratorial-worldview strengthening / weakening	ConspiratorialWorldviewIncrease; ConspiratorialWorldviewDecrease	Message-internal evidence bearing toward stronger or weaker generalised interpretation of consequential events through hidden coordination, secret plots, covert powerful actors, or systematic concealed control. A Narrative conspiracy/hidden-hand structure or one specific conspiracy allegation is insufficient unless broader worldview change is independently recoverable (EUvsDisinfo, n.d.).

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Pluralist / anti-pluralist worldview positioning	AntiPluralistWorldviewIncrease; AntiPluralistWorldviewDecrease	Representations bearing toward greater or lesser rejection of legitimate political or societal pluralism, including whether competing viewpoints, opposition, diversity, and coexistence are represented as legitimate features of the relevant civic order. Strong ideology or disagreement alone is insufficient (Medzihorsky and Lindberg, 2024).
Ideological-community engagement / disengagement	IdeologicalCommunityEngagementIncrease; IdeologicalCommunityEngagementDecrease	Join, follow, subscribe, enter, remain, participate, deepen involvement, leave, unfollow, or sever ties where engagement with a worldview-reinforcing ideological community is itself the recoverable outcome. Ordinary civic participation and violent-extremist organisational engagement are coded in their respective domains.
Health-guidance adherence / non-adherence	HealthGuidanceAdherenceIncrease; HealthGuidanceAdherenceDecrease	Message-internal evidence bearing toward greater or lesser adherence to an identifiable health recommendation, preventive measure, treatment instruction, or public-health course of action. Trust or distrust alone is insufficient unless adherence itself is independently recoverable (Nascimento et al., 2022).
Health-care engagement / withdrawal	HealthCareEngagementIncrease; HealthCareEngagementDecrease	Recommendations or representations bearing toward seeking, continuing, delaying, avoiding, refusing, or withdrawing from health-care services or professional medical care. Treatment-position disagreement alone is insufficient unless care engagement itself is affected.
Substitute-remedy adoption / rejection	SubstituteRemedyAdoptionIncrease; SubstituteRemedyAdoptionDecrease	Promotion, adoption, rejection, discontinuation, or avoidance of a represented remedy or treatment where it is positioned as a substitute for another represented health intervention, treatment, recommendation, or professional care. Do not adjudicate safety, efficacy, or evidential status.
Crisis protective-action mobilization / demobilization	CrisisProtectiveActionIncrease; CrisisProtectiveActionDecrease	Evacuate, shelter, prepare, isolate, protect, comply, refrain, remain, withdraw, or comparable crisis-response orientations where increased or decreased protective action itself is recoverable (Hilberts et al., 2025).

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Emergency-service capacity preservation / disruption	EmergencyServiceCapacity-Increase; EmergencyServiceCapacity-Decrease	Communication bearing toward preserving, restoring, obstructing, exhausting, diverting, or overloading emergency-response availability or resources, including instructions concerning mass unnecessary calls or deliberate resource diversion. Mere use of emergency services is insufficient.
Environmental-protection action mobilization / demobilization	EnvironmentalProtectionAction-Increase; EnvironmentalProtectionAction-Decrease	Calls, recommendations, commitments, discouragement, withdrawal, or comparable representations bearing toward increased or decreased conservation, pollution-reduction, waste-reduction, habitat or species protection, resource stewardship, or other environmental-protection behaviour. Use the more specific climate-action label for an identical explicitly climate-related endpoint (Chatterjee et al., 2024).
Climate-change acceptance positioning	ClimateChangeAcceptance-Increase; ClimateChangeAcceptance-Decrease	Message-internal evidence bearing toward greater or lesser acceptance of substantive claims about climate change, including its occurrence or anthropogenic contribution. Claims about the honesty, credibility, or reliability of scientists are coded under Domain D unless acceptance of the climate proposition is independently affected (Lewandowsky, 2021).
Climate-policy stance	ClimatePolicySupport; ClimatePolicyOpposition	Message-internal evidence bearing toward support for or opposition to an identifiable governmental, regulatory, institutional, or public-policy response to climate change. Climate belief alone is insufficient; the policy outcome must be independently recoverable.
Mitigation-action mobilization / demobilization	ClimateMitigationActionIncrease; ClimateMitigationActionDecrease	Calls, recommendations, commitments, discouragement, withdrawal, or comparable representations bearing toward increased or decreased individual, collective, organisational, or institutional mitigation behaviour. Policy stance alone is insufficient.
Adaptation-action mobilization / demobilization	ClimateAdaptationActionIncrease; ClimateAdaptationActionDecrease	Calls, recommendations, preparedness measures, infrastructural or behavioural adjustments, discouragement, delay, or withdrawal bearing toward increased or decreased adjustment to represented climate impacts.

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Climate-resilience strengthening / weakening	ClimateResilienceIncrease; ClimateResilienceDecrease	Message-internal evidence bearing toward stronger or weaker system capacity to cope with, withstand, reorganise under, recover from, or maintain essential function during climate-related hazards or disturbances. A single adaptation measure is insufficient unless broader resilience capacity is independently advanced or weakened.
Economic-confidence positioning	EconomicConfidenceIncrease; EconomicConfidenceDecrease	Representations bearing toward increased or decreased confidence in the condition, stability, prospects, functioning, or recovery capacity of the relevant economy. Institutional criticism alone is insufficient unless economic confidence itself is independently affected.
Financial-system confidence positioning	FinancialSystemConfidence-Increase; FinancialSystemConfidence-Decrease	Representations bearing toward increased or decreased confidence in the safety, stability, reliability, liquidity, solvency, or continued functioning of banks, deposits, or the wider financial system. Withdrawal behaviour is coded separately where independently advanced.
Essential-resource stockpiling / normalization	EssentialResourceStockpiling-Increase; EssentialResourceStockpiling-Decrease	Bulk-acquisition, stockpiling, hoarding, restraint, normalisation, or anti-stockpiling orientations involving essential goods or resources in a represented scarcity, shortage, disruption, or crisis context. Do not infer panic or irrationality (Ali et al., 2023).
Deposit-run mobilization / containment	DepositRunParticipationIncrease; DepositRunParticipationDecrease	Calls, recommendations, warnings, reassurance, coordination, or comparable representations bearing toward participation in or prevention of widespread run-like deposit withdrawal. Financial confidence and withdrawal behaviour are coded separately where both are independently recoverable (Sandri et al., 2023).
Critical-infrastructure confidence positioning	CriticalInfrastructureConfidence-Increase; CriticalInfrastructureConfidence-Decrease	Representations bearing toward increased or decreased confidence in the dependability, continuity, availability, or recoverability of critical infrastructure or essential services. A bounded outage report alone is insufficient.

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Critical-infrastructure resilience strengthening / weakening	CriticalInfrastructureResilienceIncrease; CriticalInfrastructureResilienceDecrease	Message-internal orientation toward strengthening or weakening system-level capacity to withstand, absorb, adapt to, recover from, or maintain essential functions during disruption, including redundancy, preparedness, continuity, or recovery resources. Confidence alone is insufficient (OECD, 2019).
Resilience strengthening	InformationResilienceIncrease	Message-internal orientation toward building transferable capacity to recognise, evaluate, resist, or recover from subsequent influence attempts; examples include teaching recognition of manipulation techniques, developing verification or source-evaluation practices, rehearsing resistance strategies, or otherwise strengthening preparedness beyond a single immediate claim. Mere rebuttal or disagreement is insufficient unless broader resistance capacity is independently advanced (Roozenbeek et al., 2022a).
Resilience weakening	InformationResilienceDecrease	Message-internal orientation toward degrading transferable capacities, practices, norms, or resources for recognising, evaluating, resisting, or recovering from subsequent influence; examples include discouraging verification, undermining defensive or critical-evaluation practices, or promoting abandonment of independent resistance capacities. The presence or successful use of an influence tactic alone is insufficient.
Information withdrawal / avoidance	InformationAvoidanceIncrease	Recommendations, conclusions, or represented orientations to stop reading, watching, following, seeking, receiving, or otherwise engaging with relevant information or information sources. Distrust or reduced resilience alone does not license this label unless withdrawal or avoidance is independently recoverable (Hicks et al., 2025).

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Information seeking / verification	InformationSeekingIncrease	Recommendations or orientations to actively seek, consult, retrieve, compare, monitor, or verify relevant information or sources. A one-off seeking act does not establish information resilience unless transferable defensive capacity is also independently advanced (Hicks et al., 2025).
Military-force support positioning	MilitaryForceSupportIncrease; MilitaryForceSupportDecrease	Representations bearing toward greater or lesser support for military intervention, defence of another state or ally, continuation or escalation of military operations, withdrawal, non-intervention, or comparable international use-of-force choices. Support for a conflict party or alliance alone is insufficient (Krishnarajan and Tolstrup, 2023).
Military-service mobilization / demobilization	MilitaryServiceParticipationIncrease; MilitaryServiceParticipationDecrease	Enlist, reenlist, report for duty, remain in service, avoid enlistment, decline entry, leave, or otherwise increase or decrease behavioural participation in regular military service. Support for the use of military force is coded separately (Pretorius, 2019).
Foreign-assistance support positioning	ForeignAidSupportIncrease; ForeignAidSupportDecrease	Representations bearing toward provision, expansion, continuation, reduction, withholding, termination, or prevention of foreign humanitarian, economic, military, or comparable assistance. Support for the recipient alone is insufficient.
Sanctions support positioning	SanctionsSupportIncrease; SanctionsSupportDecrease	Representations bearing toward imposing, maintaining, strengthening, weakening, suspending, removing, or avoiding sanctions against a represented foreign actor or target. Hostility toward the target alone is insufficient (Geissler et al., 2023).
Conflict-party alignment	ConflictPartySupportIncrease; ConflictPartySupportDecrease	Representations bearing toward greater or lesser approval, sympathy, solidarity, political support, or geopolitical alignment with an identifiable party to an international conflict. Concrete policy instruments are coded separately where independently recoverable.

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Alliance support positioning	AllianceSupportIncrease; AllianceSupportDecrease	Representations bearing toward establishing, maintaining, strengthening, weakening, leaving, rejecting, or otherwise changing support for an alliance or comparable international security alignment. Alliance reliability and specific use-of-force decisions are coded separately.
Alliance reliability positioning	AllianceReliabilityConfidenceIncrease; AllianceReliabilityConfidenceDecrease	Representations bearing toward greater or lesser confidence that an ally, security partner, or alliance relationship will reliably honour, sustain, or perform represented security commitments. Alliance support does not follow automatically.
Personal-privacy strengthening / weakening	PersonalPrivacyIncrease; PersonalPrivacyDecrease	Message-internal evidence bearing toward stronger or weaker privacy, confidentiality, anonymity, control over disclosure, or protection from unwanted identification, tracing, access, dissemination, or exposure. Doxxing, publication of identifying information, removal of such information, anonymisation, concealment, or disclosure restrictions may provide evidence. A disclosure threat used only as leverage does not by itself establish a privacy AboutObject (Chen et al., 2019; Eckert and Metzger-Riftkin, 2020).
Personal-security strengthening / weakening	PersonalSecurityIncrease; PersonalSecurityDecrease	Message-internal evidence bearing toward stronger or weaker security of identifiable persons or populations against targeting, intimidation, retaliation, harassment, exposure-related harm, or comparable personal-security threats. A coercive or threatening Method alone is insufficient.
Public-expression mobilization / suppression	PublicExpressionIncrease; PublicExpressionDecrease	Communication bearing toward continued or increased speaking, publishing, reporting, opinion expression, or other public communication, or toward silencing, deterrence, withdrawal, withholding, or self-censorship. Threats, harassment, surveillance, reputational pressure, or other coercive means may provide evidence where the expression outcome is independently recoverable (Björkenfeldt, 2026; Marwick and Lewis, 2017).

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Commercial inquiry / lead generation	CommercialInquiryIncrease; CommercialInquiryDecrease	Requests or discouragement concerning quotations, demonstrations, consultations, sales contact, appointments, product information, or comparable pre-purchase commercial inquiry. Mere awareness or browsing is insufficient.
Purchase mobilization / avoidance	PurchaseBehaviorIncrease; PurchaseBehaviorDecrease	Purchase, order, checkout, buy-now, patronage, non-purchase, switching, boycott, or comparable market-behaviour orientations where increased or decreased purchasing itself is independently recoverable (Lahtinen et al., 2020).
Subscription enrollment / avoidance	SubscriptionEnrollmentIncrease; SubscriptionEnrollmentDecrease	Subscribe, opt in, enrol, join a recurring commercial service, or avoid or reject prospective enrolment. Cancellation of an already active relationship is coded under customer retention.
Application installation / avoidance	AppInstallationIncrease; AppInstallationDecrease	Install, download, avoid installation, reject download, or comparable orientation toward installation of an identifiable commercial application or digital product.
Trial enrollment / avoidance	TrialEnrollmentIncrease; TrialEnrollmentDecrease	Start, register for, avoid, reject, or refrain from a represented commercial trial, sample period, demonstration period, or comparable pre-conversion experience.
Customer retention / churn	CustomerRetentionIncrease; CustomerRetentionDecrease	Stay, renew, continue, do not cancel, cancel, leave, defect, switch away, or otherwise preserve or terminate an existing commercial customer relationship. Loyalty is coded separately where preferential commitment itself is affected.
Customer reactivation / non-reactivation	CustomerReactivationIncrease; CustomerReactivationDecrease	Return, reactivate, resume, come back, remain lapsed, or reject return where the relevant commercial relationship is represented as having previously ended or become inactive.

Continued on next page

Supplementary Table S16: L4 evidence templates (continued)

Evidence template (L4)	Instantiates L3 label(s)	When to select / typical cues
Brand-trust strengthening / weakening	BrandTrustIncrease; BrandTrustDecrease	Representations bearing toward stronger or weaker confidence in the reliability, honesty, dependability, competence, or promise-fulfilment of an identifiable commercial brand, product, service, or provider. Purchase, retention, favourable evaluation, and loyalty are insufficient unless commercial trust itself is independently recoverable. System-level confidence in banks or the financial system is coded under Domain H (Chaudhuri and Holbrook, 2001).
Brand-loyalty strengthening / weakening	BrandLoyaltyIncrease; BrandLoyaltyDecrease	Representations bearing toward stronger or weaker durable preference, commitment, attachment, or propensity to continue favouring a brand or provider. Retention or repeat behaviour alone is insufficient unless loyalty itself is independently recoverable.
Brand recommendation / referral	BrandRecommendationIncrease; BrandRecommendationDecrease	Recommend, refer, tell others to choose, withdraw recommendations, or discourage recommending an identifiable commercial brand, product, or service (Buttle, 1998).
Review generation / suppression	ReviewGenerationIncrease; ReviewGenerationDecrease	Requests or discouragement concerning submission, publication, removal, or withholding of public ratings, reviews, testimonials, or comparable consumer evaluations. Review truth or valence is not adjudicated.
Brand advocacy strengthening / weakening	BrandAdvocacyIncrease; BrandAdvocacyDecrease	Message-internal evidence bearing toward stronger or weaker sustained public endorsement, defence, ambassador-like promotion, or active representation of a brand beyond an isolated recommendation or review.

F.5 AboutObject: Evidence map (construct → role → sources)

Supplementary Table S17: Traceability of literature use in the AboutObject specification.

Construct / design choice	Role in this specification	Key sources
Truth/intent-agnostic coding constraint	Methodological constraint preventing annotation from depending on factual verification, private communicative intention, or observed audience effects	(Kavanagh and Rich, 2018; Marwick and Lewis, 2017)
Message-internal outcome orientation	Supports use of observable message content, including claims, evaluations, recommendations, calls to action, and represented relations, as evidence for the broader influence object while distinguishing that object from downstream empirical effects	(Kavanagh and Rich, 2018; Marwick and Lewis, 2017)
Politics: participation, actor support, legitimacy, and compliance	Motivates Domain A distinctions among electoral and civic participation, support for political actors, legitimacy of elections and the broader political system, policy stance, and behavioural policy compliance	(Kim et al., 2018; Brennen et al., 2020; Kavanagh and Rich, 2018; Anderson and Just, 2013; Dowling and Legrand, 2023)
Identity, intergroup relations, and social cohesion	Supports Domain B treatment of collective identity, polarization, outgroup hostility, inclusion/exclusion, humanization/dehumanization, and wider social cohesion as related but non-identical group-relational outcomes	(Kavanagh and Rich, 2018; EUvsDisinfo, n.d.; United Nations, 2019)
Political violence and violent-extremist outcomes	Supports Domain C separation of attitudinal support for violence from violent participation and of support for violent-extremist or terrorist activity from behavioural organisational engagement	(Fortify Rights, 2018; United Nations, 2019; McCauley and Moskalenko, 2017)
Epistemic trust, institutional legitimacy, and attribution	Motivates Domain D distinctions among source-specific trust, institutional trust, institutional legitimacy, generalised epistemic confidence, and responsibility attribution; also supports separation of localised source doubt from generalised collapse in confidence that reliable knowledge is available	(Ognyanova et al., 2020; Lazer et al., 2018; Kavanagh and Rich, 2018; Diel et al., 2024)

Supplementary Table S17: Traceability of literature use (continued)

Construct / design choice	Role	Key sources
Method–AboutObject distinction for epistemic erosion	Illustrates that Method-level doubt-seeding or deceptive/manipulative operations do not automatically instantiate an epistemic-trust AboutObject, while they may provide evidence where broader trust erosion is independently recoverable	(Linden, 2022; Lewandowsky, 2021)
Conspiratorial and anti-pluralist worldviews	Supports Domain E distinction between specific conspiracy narratives and a generalised conspiratorial worldview, and between strong ideological commitment and broader rejection of legitimate political or social pluralism	(EUvsDisinfo, n.d.; Malcolm, 2023; Medzihorsky and Lindberg, 2024)
Public-health and civil-protection behaviour	Motivates Domain F treatment of health-guidance adherence, health-care engagement, substitute-remedy behaviour, crisis protective action, and emergency-response capacity as substantive outcomes without requiring medical truth or efficacy adjudication	(Nascimento et al., 2022; Hilberts et al., 2025)
Environmental and climate behaviour	Supports Domain G distinctions among environmental-protection behaviour, climate-change acceptance, climate-policy stance, mitigation and adaptation action, and broader climate-resilience capacity	(Lewandowsky, 2021; Chatterjee et al., 2024)
Economic confidence, runs, and infrastructure resilience	Supports Domain H separation of economic or financial-system confidence from stockpiling and deposit-run behaviour, and of infrastructure confidence from system-level infrastructure resilience	(Ali et al., 2023; Sandri et al., 2023; OECD, 2019)
Information and influence resilience	Motivates `InformationResilience-*` as a transferable capacity to recognise, evaluate, resist, or recover from subsequent influence attempts, rather than the success or failure of one immediate rebuttal or influence episode	(Roozenbeek et al., 2022a)

Supplementary Table S17: Traceability of literature use (continued)

Construct / design choice	Role	Key sources
Information seeking and avoidance	Supports treatment of active information seeking and information avoidance as information practices that need not be exact directional inverses and need not instantiate broader resilience automatically	(Hicks et al., 2025)
Geopolitical alignment and international use of force	Supports Domain J separation of public support for military force, conflict parties, alliance relationships, alliance reliability, foreign assistance, sanctions, and military-service participation	(Krishnarajan and Tolstrup, 2023; Geissler et al., 2023; Tomz and Weeks, 2021; Pretorius, 2019)
Personal privacy, security, and public expression	Supports Domain K separation of privacy loss, personal-security degradation, and silencing or self-censorship from the threatening, harassing, coercive, or doxxing operations that may be used to produce those outcomes	(Chen et al., 2019; Eckert and Metzger-Riftkin, 2020; Björkenfeldt, 2026; Marwick and Lewis, 2017)
Commercial acquisition and customer lifecycle	Motivates Domain L distinction among inquiry, purchase and enrolment behaviours, customer retention and reactivation, and later relationship or communication outcomes such as loyalty, recommendation, reviews, and advocacy	(Lahtinen et al., 2020; Gordon, 2012; Buttle, 1998; Friedman, 1991)
Brand trust versus loyalty	Supports treatment of trust in a commercial brand or provider as a distinct commercial relationship outcome rather than a synonym for purchase, retention, or durable preferential loyalty	(Chaudhuri and Holbrook, 2001)

G Computational Verification and Reproducibility Package

The operational codebook in Appendices A–F is the normative specification of the Influence Ontology. To verify that its structural commitments can be represented consistently and enforced computationally, we additionally constructed an OWL 2 reference representation and a corresponding SHACL constraint graph. These formal artefacts are used for specification-level verification; they are not requirements for downstream annotation, classification, or deployment.

G.1 Formal verification strategy

OWL and SHACL serve complementary purposes. OWL reasoning operates under the Open World Assumption and is used here to assess whether the formal reference representation can be classified consistently under OWL 2 DL semantics. SHACL provides closed-world validation of structural requirements that are more naturally expressed as graph constraints, including required assignments and cardinality restrictions (Knublauch and Kontokostas, 2017).

The formalisation preserves the distinction between interpretive annotation and deterministic derivation. Message evidence is used to assign the canonical base labels specified by the codebook. Determinism applies only to relations explicitly fixed by the specification. Thus, once a Tactic is assigned, its Method family and TargetConstruct profile are fixed; once a Polarization motif is assigned, its direction and axis set are fixed; NarrativeArchetypes are derived from specified device/attribute signatures evaluated within each assigned NarrativeConfiguration; and a specific AboutObject label deterministically inherits its higher-level domain and sphere. TopicDomain remains independently assigned.

G.2 Verification artefacts

The accompanying formal verification package contains the artefacts required to inspect the formalisation and reproduce the reported OWL and SHACL checks:

- `influence_ontology_reference.ttl`: the OWL/RDF reference representation of the ontology;
- `influence_ontology_shapes.ttl`: the SHACL constraint graph;
- `tests/golden_annotations.ttl`: a test graph constructed to satisfy the specified SHACL constraints;
- `tests/negative_annotations.ttl`: a deliberately malformed test graph used to verify constraint detection;
- `tests/narrative_archetype_tests.json`: minimal-signature tests for NarrativeArchetype derivation; and
- `tests/narrative_configuration_tests.json`: regression tests for NarrativeConfiguration-scoped derivation.

The human-readable codebook remains authoritative for construct definitions, annotation rules, and canonical mappings.

G.3 Validation procedure

The RDF/Turtle artefacts were first checked for successful parsing. The conforming test graph was constructed to satisfy the formal constraints, whereas the nonconforming graph deliberately instantiated twelve structural violations covering assignment cardinality and NarrativeConfiguration membership, parentage, message consistency, local identifiers, and configuration membership requirements.

Closed-world conformance was evaluated using pySHACL 0.40.1 with RDFLib 7.6.0. The same SHACL constraint graph was applied to both test graphs.

NarrativeArchetype derivation was tested separately using twelve positive minimal-signature tests and six NarrativeConfiguration-scope regression tests. The latter test within-configuration composition, prohibition of cross-configuration composition, existential matching among multiple same-type device instances, same-instance conjunction, multiple archetypes within one configuration, and non-derivation from partial signatures.

OWL 2 DL classification was evaluated separately in Protégé using the HermiT reasoner (Musen and Protégé Team, 2015; Glimm et al., 2014). The generated OWL reference ontology was classified and its inferred class hierarchy inspected for named unsatisfiable classes.

G.4 Results

Both RDF/Turtle formal artefacts parsed successfully. Standards-based SHACL validation produced the expected discrimination: the conforming test graph satisfied the constraint graph (`Conforms: True`; 0 validation results), whereas the deliberately nonconforming graph did not (`Conforms: False`; 12 validation results).

All twelve positive NarrativeArchetype minimal-signature tests and all six NarrativeConfiguration-scope regression tests completed successfully.

HermiT completed OWL 2 DL classification of the formal reference ontology without error. Inspection of the inferred class hierarchy identified no named unsatisfiable classes.

Taken together, these checks show that the OWL reference representation can be processed coherently under the tested description-logic reasoner and that the tested closed-world structural constraints are enforceable through SHACL. The Narrative regression tests additionally confirm the specified configuration-scoped derivation behaviour once NarrativeDevice instances, applicable attributes, and configuration memberships have been fixed. SHACL does not establish whether a human annotator selected the semantically correct NarrativeConfiguration. These checks do not test whether independent annotators reliably identify the underlying message-level constructs, whether those constructs produce audience effects, or whether they

support downstream predictive performance; those are separate empirical validation questions.

H Coverage check: cited frameworks against ontology architecture

This appendix provides supporting evidence for the paper's coverage claim relative to cited adjacent frameworks. The exercise is a post hoc representational coverage check; it is not the procedure by which the ontology's canonical categories were generated. Coverage is evaluated against the ontology's message-internal analytical architecture. Four core annotation branches are relevant: Method/Tactic, Narrative, Polarization, and AboutObject. Within the Method branch, Tactic is the assigned canonical operation and its Method family is inherited deterministically. TopicDomain is an additional independently assigned supplementary substantive index. TargetConstruct is treated separately as a deterministic analytical profile inherited from an independently licensed Tactic rather than as an independently annotated dimension.

Three verdict categories are used. *Covered* indicates that the message-internal analytical distinctions relevant to the comparison can be represented by the ontology, including where a source category must be decomposed into several more specific ontology constructs. *Partial overlap (scope)* indicates that the framework contains message-internal constructs that are representable but also includes actor-, campaign-, platform-, infrastructure-, distribution-, contextual-, or observed-effect dimensions outside the ontology's unit of analysis. *Scoped out (design)* indicates that the framework's principal distinction depends on judgements or information that the ontology deliberately does not infer from message-internal evidence, such as external veracity, private intent, or operational attribution.

Framework selection is purposive rather than exhaustive. The comparison set follows the principal neighbouring units of analysis identified in the related-work section: information-disorder and veracity/intent typologies; message-level propaganda and persuasion-technique inventories; campaign-, incident-, and operation-level influence frameworks; and framing/narrative frameworks. Historical anchor frameworks are included alongside widely used computational or institutional operationalisations and recent integrative proposals. The resulting claim is deliberately bounded: the tables establish coverage relative to the frameworks examined here, not relative to every taxonomy or theory in the broader influence literature.

The coverage check is also intended to be extensible. A candidate framework is informative for ontology development when it introduces an independently recoverable message-internal distinction that cannot be represented through Method/Tactic, Narrative, Polarization, AboutObject, or the sup-

plementary TopicDomain index. Constructs requiring excluded levels of analysis or external evidence instead delimit the ontology's scope rather than constituting missing message-level categories.

The crosswalks below are representational correspondences, not annotation rules. A broad source-framework category may map to more than one ontology construct because the ontology separates communicative operations that the source framework combines. Conversely, a correspondence shown in Narrative, Polarization, AboutObject, or TopicDomain does not mean that assignment of a Method/Tactic automatically licenses that annotation. Each corresponding branch must satisfy its own codebook criteria independently. TargetConstruct values are inherited only from an independently licensed Tactic.

Supplementary Table S18: Coverage check: summary of framework verdicts.

Framework	Primary contribution	Ontology correspondence / scope	Verdict
Wardle & Derakhshan 2017	Information-disorder framework distinguishing mis-, dis-, and mal-information through relations among falsity/genuineness, harmful use, and intent	The ontology deliberately does not determine external truth status or private harmful/deceptive intent from message content; messages in any of these categories may nevertheless be analysed using the message-internal ontology [b]	Scoped out (design)
ENISA/EEAS FIMI 2022	Analytical framework linking Foreign Information Manipulation and Interference with cybersecurity threat-landscape analysis	Message-internal manipulative content is representable, while the framework also addresses threat actors, incidents, cybersecurity relations, distribution behaviour, attribution, and other operation-level properties [a]	Partial overlap (scope)
Da San Martino et al. 2020	SemEval-2020 Task 11: span identification and classification using a 14-class propaganda-technique inventory in news articles	The 14 source classes are representable primarily through Method/Tactic, with decomposition into multiple canonical Tactics where the shared-task taxonomy merges distinct operations; see Table S19	Covered
Dimitrov et al. 2021	SemEval-2021 Task 6: persuasion-technique detection in memes using 20 textual techniques and 22 techniques in the multimodal task	The persuasion-technique distinctions are representable through the canonical Method/Tactic inventory and conditional independent branches; text–image modality is treated as an evidence/implementation property rather than an additional influence dimension[c]	Covered
Piskorski et al. 2023	SemEval-2023 Task 3 persuasion subtask: paragraph-level identification of 23 persuasion techniques in multilingual online news	The 23 persuasion categories are representable through Method/Tactic, including decomposition where a source category combines operations; the shared task's separate genre and generic-frame subtasks are not treated as part of this technique crosswalk; see Table S21	Covered
Dimitrov et al. 2024	SemEval-2024 Task 4: multilingual persuasion detection in memes using a hierarchical organisation of 22 persuasion techniques, of which 20 apply to textual content and two additional techniques are reserved for multimodal evidence	The hierarchy provides a further construct-level coverage check against the canonical Method/Tactic inventory; modality and benchmark hierarchy do not constitute additional independently annotated influence dimensions [c]	Covered
DISARM (DISARM Foundation, n.d.; Newman, 2022)	Tactics, techniques, and procedures for describing influence-operation behaviour across an incident or campaign lifecycle	Message-content practices overlap with Method/Tactic and Narrative, while actor, infrastructure, asset, amplification, distribution, coordination, and campaign-sequencing behaviour lies outside the message-internal unit of analysis[a]	Partial overlap (scope)
Johnson et al. 2021	Semantic ontology for Cyber Influence Campaigns linking physical and cyber actions, events, and their causal and structural relations	Message content represented within a campaign can be analysed independently by the present ontology, while cross-realm event structure, campaign linkages, social-network relations, and other physical/cyber operational properties lie outside the message-internal unit of analysis [a]	Partial overlap (scope)
Pamment & Smith 2022	Attribution framework for information influence operations	Operational and actor attribution requires evidence beyond message-internal form and is deliberately not inferred by the ontology [a,b]	Scoped out (design)

Table S18 (continued)

Framework	Primary contribution	Ontology correspondence / scope	Verdict
OII Computational Propaganda	Campaign-level study of computational amplification, automation, coordination, and political communication	Message content can be represented where relevant, while automated distribution, amplification, account behaviour, and coordination remain outside scope[a]	Partial overlap (scope)
Tudela et al. 2025	Influence Operation Ontology comprising Threat, Channel, and Social domains for modelling influence-operation entities and relations	Message and Narrative entities provide direct content-level overlap, while Threat Actors, Campaigns, Incidents, Attack Patterns, Channels, User Accounts, communities, locations, and related operational relations extend well beyond the message-internal ontology[a]	Partial overlap (scope)
Media Manipulation Casebook;	Campaign-lifecycle documentation spanning content creation, actor activity, platform tactics, distribution, and coordination	Message-level communicative content is representable, while lifecycle, platform, distributional, and actor-coordination properties remain outside scope[a]	Partial overlap (scope)
Lenk 2026	Four-level framework linking semantic frames, strategic narratives, master-narratives, and meta-frames	Text-internal semantic frames and strategic narratives overlap with Framing Tactics and Narrative structure; master-narratives and meta-frames require historical, ideological, political, or geopolitical contextual interpretation beyond the ontology's strictly message-internal evidence rule [d]; see Table S24	Partial overlap (scope)
François 2019	Actors–Behaviors–Content framework for analysing disinformation and influence operations	Content and some message-realised behaviours overlap with the ontology; actor identity, coordinated operational behaviour, and other campaign-level relations remain outside scope[a]	Partial overlap (scope)
Pamment 2020	ABCDE framework extending Actors–Behaviors–Content with Degree and Effect	Message-internal content and some communicative behaviours overlap with the ontology. Actor, reach/degree, campaign behaviour, and observed effects are outside scope; message-internally represented outcomes may instead be captured by AboutObject without asserting that the effect occurred [a]	Partial overlap (scope)
Nimmo & Hutchins 2023	Online Operations Kill Chain: ten observation-based tactical phases for online operations	Some posting, engagement, and content-related behaviours intersect with message-level evidence, while asset creation, infrastructure, evasion, coordination, targeting, distribution, and other operation-level behaviours remain outside the ontology's unit of analysis[a]	Partial overlap (scope)
Institute for Propaganda Analysis 1939	Historical seven-device propaganda inventory	The seven devices can be represented by canonical Method/Tactic operations, sometimes by decomposition rather than one-to-one equivalence; see the historical-device crosswalk below	Covered
Entman 1993	Framing through problem definition, causal interpretation, moral evaluation, and treatment recommendation	The four functions are representable by decomposition across TopicDomain, Narrative, applicable Method/Tactic operations, Polarization, and AboutObject; no branch is automatically activated merely by the source framing-function label; see Table S26	Covered
Benford & Snow 2000	Core collective-action framing tasks: diagnostic, prognostic, and motivational framing	The three core tasks can be decomposed across Narrative, TopicDomain, AboutObject, and independently licensed Method/Tactic and Polarization constructs; TargetConstruct enters only through the fixed profile of a licensed Tactic; see Table S27	Covered

[a] Actor-, campaign-, infrastructure-, platform-, distribution-, reach-, coordination-, and sequencing-level properties are outside the ontology's message-internal unit of analysis unless the relevant relation is itself represented as message content.

[b] External truth status, private communicative intent, and operational attribution are not inferred as annotation prerequisites.

[c] The SemEval meme tasks vary the modality from which persuasion-technique evidence is detected. The present coverage verdict concerns conceptual label coverage rather than performance on a particular textual or visual input modality.

[d] Lenk explicitly distinguishes text-extractable semantic frames and strategic narratives from higher-level master-narratives and meta-frames requiring broader historical, ideological, political, or geopolitical interpretation.

H.1 SemEval propaganda and persuasion shared-task lineage

The SemEval propaganda/persuasion tasks provide a useful longitudinal coverage check because the associated technique inventory changes in granularity, modality, hierarchy, language, and domain while retaining a common concern with message-level persuasive operations.

SemEval-2020 Task 11 used a 14-class technique-classification inventory in English-language news articles (Da San Martino et al., 2020). That inventory was itself derived from the earlier 18-technique fine-grained propaganda inventory used in the NLP4IF-2019 lineage: several rare categories were merged for the 2020 shared task and one was removed from the shared-task label set. SemEval-2021 Task 6 subsequently extended the analysis to memes, using 20 techniques for the text-only subtasks and 22 for the multimodal subtask (Dimitrov et al., 2021). SemEval-2023 Task 3 returned to multilingual online news and separated genre classification, generic framing, and a 23-technique persuasion subtask (Piskorski et al., 2023b). SemEval-2024 Task 4 again focused on memes and organised a 22-technique inventory hierarchically; 20 techniques apply to textual content and two additional techniques are available where visual evidence is considered (Dimitrov et al., 2024).

The crosswalks below compare the source-framework categories with the ontology's final canonical specification. They do not assume one-to-one equivalence. Where a source category combines observably different communicative operations, it is decomposed. Where a source category is defined partly by presumed intent or audience effect, only the message-internally recoverable operation is represented. Narrative, Polarization, AboutObject, and TopicDomain annotations remain independent unless explicitly licensed by their own evidence rules.

H.2 SemEval-2020 Task 11: 14-class technique crosswalk

The official SemEval-2020 technique-classification task contains 14 classes. Two of these classes deliberately merge several techniques: *Whataboutism / Straw man / Red herring* and *Bandwagon / Reductio ad Hitlerum*. The ontology restores these distinctions where their constitutive operations differ.

Supplementary Table S19: SemEval-2020 Task 11 propaganda-technique classes crosswalked to the canonical ontology.

SemEval-2020 class	Canonical representation	Crosswalk rationale
Loaded language	*Loaded language* / Framing	Direct correspondence where evaluatively or affectively charged lexical choice itself supplies the frame.

Supplementary Table S19: SemEval-2020 Task 11 crosswalk (continued)

SemEval-2020 class	Canonical representation	Crosswalk rationale
Name calling or labeling	*Name-calling / labeling* / Framing	Direct correspondence where a compact evaluative epithet or categorical characterisation frames an identifiable target.
Repetition	*Repetition* / Persuasion	Direct correspondence where the same or substantially equivalent proposition recurs within the available message or recoverable message sequence and the recurrence functions as reinforcement. Campaign-level posting frequency or amplification remains outside scope.
Exaggeration or minimization	*Exaggeration / overstatement*; *Minimization / understatement* / Deception	The source class combines opposite scalar directions. The ontology separates upscaling from downscaling because each has a distinct observable relation to a recoverable comparison point or baseline.
Doubt	*Doubt-seeding* / Persuasion; *Ad hominem / personal-credibility attack* / Distraction where applicable	The source definition broadly concerns questioning the credibility of someone or something. The ontology separates epistemic doubt directed at an external claim, source, evidence base, consensus, or inference from person-directed credibility undercutting tied to an identifiable contribution. Mere criticism does not automatically instantiate either Tactic.
Appeal to fear / prejudice	*Appeal to fear* / Persuasion; *Threat framing* / Framing; other applicable group-evaluative Tactic	The source class combines fear-based support with prejudice-based evaluation. The ontology distinguishes a represented threat relation from use of threatened harm as a warrant for an identifiable protective response. Prejudicial representation is coded through whichever specific evaluative, group-directed, or dehumanising operation is independently evidenced rather than through a generic prejudice Tactic.
Flag-waving	*Unity (shared identity)* / Persuasion; other applicable identity or temporal Tactic	National or group identification maps to *Unity* when shared membership itself functions as the influence warrant. Symbols or patriotic language alone are insufficient. A positively valued collective past may instead support *Collective nostalgic framing*, and longstandingness used as a warrant may support *Appeal to tradition*.
Causal oversimplification	*Causal oversimplification*; *Scapegoating* / Framing where independently satisfied	The source definition includes both restricted causal explanation and transfer of blame to a person or group. The ontology separates these: monocausal explanation does not require a scapegoat, and concentrated blame does not require a monocausal claim.

Supplementary Table S19: SemEval-2020 Task 11 crosswalk (continued)

SemEval-2020 class	Canonical representation	Crosswalk rationale
Slogans	No standalone Method/Tactic by form; conditional Narrative *Symbolic condensation*	Being a brief or striking slogan is not itself a unique influence operation. A slogan may instantiate *Symbolic condensation* where it compresses a broader narrative, and it may independently instantiate a Tactic where that Tactic's own conditions are satisfied.
Appeal to authority	*Appeal to authority* / Persuasion; other source-warrant Tactic where applicable	Direct correspondence where expertise, credentials, professional standing, or recognised authority functions as the warrant. The source taxonomy also includes testimonial-like cases involving non-experts; these do not automatically satisfy *Appeal to authority* and are represented by another canonical warrant only where its own conditions are met.
Black-and-white fallacy / dictatorship	*False dilemma* / Framing; other conduct-directed Tactic where independently satisfied	Restricted alternatives map directly to *False dilemma*. The source category's "dictatorship" extreme—telling an audience what action to take—does not by itself define a separate canonical operation. Authority, choice-architecture, scarcity, or coercive labels require their own additional evidence.
Thought-terminating cliché	No standalone canonical Tactic by form; applicable Tactic or NarrativeDevice where independently satisfied	The source category combines a short conventional form with the presumed effect of discouraging critical thought. The ontology does not infer that effect from cliché form alone. Loaded language, agenda shifting, obfuscation, or *Symbolic condensation* may be coded only where their respective observable conditions are independently present.
Whataboutism / Straw man / Red herring	*Whataboutism*; *Straw man*; *Red herring* / Distraction	The shared task merged three low-frequency classes. The ontology restores the distinctions between comparative counteraccusation, substitution of an opponent's position, and tangential diversion.
Bandwagon / Reductio ad Hitlerum	*Social proof (bandwagon)*; *Associative transfer / guilt-or-honor by association* / Persuasion	The shared task merged two low-frequency techniques. Bandwagon corresponds to social prevalence functioning as a warrant. Reductio ad Hitlerum instead uses association with a negatively evaluated actor, group, or concept to transfer negative standing to the focal target.

H.3 SemEval-2021 Task 6 and SemEval-2024 Task 4: meme-persuasion crosswalk

SemEval-2021 introduced an extended 22-technique meme inventory. The first 20 techniques are applicable to textual as well as visual realisation; *Appeal to (Strong) Emotions* and *Transfer* were reserved for visual evidence in the multimodal task. SemEval-2024 re-used this 22-technique inventory in a multilingual setting and introduced a hierarchical organisation over the technique labels (Dimitrov et al., 2021; Dimitrov et al., 2024).

The SemEval-2024 hierarchy groups the leaf techniques under broader organisational categories such as Ethos, Pathos, and Logos, with additional intermediate nodes including Ad Hominem, Justification, Reasoning, Distraction, and Simplification. These hierarchy nodes are treated here as taxonomy organisation rather than as additional message-level labels. The coverage comparison therefore concerns the 22 leaf techniques.

Supplementary Table S20: SemEval-2021/2024 meme-persuasion techniques crosswalked to the canonical ontology.

Source technique	**Canonical representation**	**Crosswalk rationale**
Loaded Language	*Loaded language* / Framing	Direct correspondence.
Name Calling or Labeling	*Name-calling / labeling* / Framing	Direct correspondence where the condensed evaluative characterisation satisfies the canonical labeling criterion.
Doubt	*Doubt-seeding* / Persuasion; *Ad hominem / personal-credibility attack* / Distraction where applicable	The broad source category is decomposed according to the epistemic target and communicative relation rather than mapped mechanically to a single Tactic.
Exaggeration or Minimisation	*Exaggeration / overstatement*; *Minimization / understatement* / Deception	Directional split of the source category.
Appeal to Fear or Prejudices	*Appeal to fear* / Persuasion; *Threat framing* / Framing; other applicable group-evaluative Tactic	Fear, threat construction, and prejudicial representation are distinct operations in the ontology and are therefore separated.
Slogans	No standalone Method/Tactic by form; conditional Narrative *Symbolic condensation*	Slogan form alone does not license a Tactic. Narrative symbolic condensation is coded only when the slogan compresses broader narrative meaning.
Whataboutism	*Whataboutism* / Distraction	Direct correspondence.
Flag-Waving	*Unity (shared identity)* / Persuasion; other applicable identity/temporal Tactic	Shared group identity must function as the warrant; patriotic imagery alone does not automatically license *Unity (shared identity)*.
Misrepresentation of Someone's Position (Straw Man)	*Straw man* / Distraction	Direct correspondence.

Supplementary Table S20: SemEval-2021/2024 meme-persuasion crosswalk (continued)

Source technique	Canonical representation	Crosswalk rationale
Causal Oversimplification	*Causal oversimplification*; *Scapegoating* / Framing where independently satisfied	The source definition explicitly includes both monocausal explanation and possible blame concentration; the ontology distinguishes the two relations.
Appeal to Authority	*Appeal to authority* / Persuasion; other applicable source-warrant Tactic	Expert/authority warrants correspond directly; testimonial-like cases involving non-authorities require separate functional assessment.
Thought-Terminating Cliché	No standalone canonical Tactic by form; applicable Tactic or NarrativeDevice where independently satisfied	The source label partly characterises an intended cognitive/discussion effect. The ontology instead requires an observable canonical operation and does not infer thought termination from cliché form alone.
Black-and-White Fallacy or Dictatorship	*False dilemma* / Framing; other conduct-directed Tactic where independently satisfied	False alternatives map directly; directive force, authority, constrained choice, or coercive leverage is not inferred merely from the source category's "dictatorship" formulation.
Reductio ad Hitlerum	*Associative transfer / guilt-or-honor by association* / Persuasion	The source operation transfers negative standing from a disfavoured associate to the focal target.
Repetition	*Repetition* / Persuasion	Direct correspondence subject to message-internal or recoverable-sequence recurrence.
Obfuscation, Intentional Vagueness, Confusion	*Strategic vagueness / equivocation* / Deception; *Obfuscation (jargon/complexity)* / Distraction	The source category combines underdetermination of communicative content, linguistic opacity, and a presumed audience state. The ontology separates vagueness/equivocation from complexity-based obfuscation and does not treat confusion itself as a Tactic.
Presenting Irrelevant Data (Red Herring)	*Red herring* / Distraction	Direct correspondence where irrelevant material diverts from an established focal issue.
Bandwagon	*Social proof (bandwagon)* / Persuasion	Direct correspondence.
Smears	No one-to-one Tactic; commonly applicable *Name-calling / labeling*, *Loaded language*, *Ad hominem / personal-credibility attack*, or another independently satisfied operation	"Smear" is broader than a single observable communicative operation. Negative reputation positioning is decomposed according to whether the message labels, evaluatively frames, discounts a contribution through a personal attack, transfers negative standing, or performs another canonical operation.

Supplementary Table S20: SemEval-2021/2024 meme-persuasion crosswalk (continued)

Source technique	Canonical representation	Crosswalk rationale
Glittering Generalities (Virtue)	No one-to-one Tactic; applicable *Loaded language*, *Moral framing*, *Associative transfer / guilt-or-honor by association*, *Unity (shared identity)*, or other specific operation	The source class groups positively valued words and symbols. The ontology classifies the function performed by that value relation rather than retaining a generic positive-symbol category.
Appeal to (Strong) Emotions	No generic emotion Tactic; applicable *Appeal to fear*, *Appeal to anger / indignation*, *Appeal to pity / sympathy*, *Affective associative cueing*, or another specific operation	The source category groups visual emotional appeals by affective strength. The ontology instead distinguishes the communicative operation through which the affectively significant material functions. Emotional intensity alone is not a canonical Tactic.
Transfer	*Associative transfer / guilt-or-honor by association* / Persuasion	Direct functional correspondence where represented qualities, commitments, credibility, prestige, stigma, blame, approval, or other standing are transferred through association.

H.4 SemEval-2023 Task 3: 23-technique persuasion-subtask crosswalk

SemEval-2023 Task 3 contains three distinct subtasks: article genre classification, generic-frame detection, and persuasion-technique detection. The present table concerns only the third subtask, whose taxonomy contains 23 fine-grained persuasion techniques grouped under six source categories: Attack on Reputation, Justification, Distraction, Simplification, Call, and Manipulative Wording (Piskorski et al., 2023b). The six source-level groups organise the benchmark taxonomy; the crosswalk is made at the 23 fine-grained technique labels.

Supplementary Table S21: SemEval-2023 Task 3 persuasion techniques crosswalked to the canonical ontology.

SemEval-2023 technique	Canonical representation	Crosswalk rationale
Attack on Reputation		
Name Calling or Labelling	*Name-calling / labeling* / Framing	Direct correspondence.
Guilt by Association	*Associative transfer / guilt-or-honor by association* / Persuasion	Negative standing is transferred to the focal target through association.

Supplementary Table S21: SemEval-2023 persuasion-technique crosswalk (continued)

SemEval-2023 technique	Canonical representation	Crosswalk rationale
Casting Doubt	*Ad hominem / personal-credibility attack* / Distraction; *Doubt-seeding* / Persuasion where independently satisfied	The SemEval category concerns questioning character or personal attributes in order to undermine credibility or quality. Person-directed contribution-undercutting maps to *Ad hominem*. Where uncertainty or source-reliability concerns instead function as reasons to weaken confidence in an external epistemic target, *Doubt-seeding* applies. The labels are not interchangeable.
Appeal to Hypocrisy	*Whataboutism*; *Ad hominem / personal-credibility attack*; *Induced hypocrisy / self-inconsistency pressure*, according to function	An inconsistency allegation can redirect scrutiny through a comparative countercase, discount another person's contribution, or confront the recipient with their own standard–conduct discrepancy. The ontology separates these operations rather than retaining hypocrisy as one Tactic.
Questioning the Reputation	No one-to-one Tactic; applicable *Name-calling / labeling*, *Loaded language*, *Moral framing*, *Ad hominem / personal-credibility attack*, or another specific operation	The source category is organised around damage to reputation or moral standing rather than one unique rhetorical relation. Canonical assignment therefore depends on how that negative characterisation is performed.
Justification		
Flag Waving	*Unity (shared identity)* / Persuasion; other applicable identity/temporal operation	Group pride or shared membership maps to *Unity* only when identity itself warrants the advocated judgement or conduct.
Appeal to Authority	*Appeal to authority* / Persuasion	Direct correspondence where authority or expertise functions as the supporting warrant.
Appeal to Popularity	*Social proof (bandwagon)* / Persuasion	Direct correspondence where prevalence, approval, or conduct of others functions as social evidence.
Appeal to Values	No one-to-one Tactic; commonly *Moral framing* / Framing or another value-specific canonical operation	A broad value appeal is underdetermined. Moral values map to *Moral framing*; tradition, authority, shared identity, novelty, and other specific warrants retain their respective Tactics.
Appeal to Fear / Prejudice	*Appeal to fear* / Persuasion; *Threat framing* / Framing; other applicable group-evaluative operation	The source category combines fear/repulsion with prejudice. The ontology separates threat representation, fear-based protective influence, and specific group-evaluative operations.
Distraction		
Strawman	*Straw man* / Distraction	Direct correspondence.
Red Herring	*Red herring* / Distraction	Direct correspondence.

Supplementary Table S21: SemEval-2023 persuasion-technique crosswalk (continued)

SemEval-2023 technique	**Canonical representation**	**Crosswalk rationale**
Whataboutism	*Whataboutism* / Distraction	Direct correspondence.
Simplification		
Causal Oversimplification	*Causal oversimplification* / Framing	Direct correspondence where one cause is represented as sufficient, overriding, or exhaustive of a more complex causal explanation.
False Dilemma or No Choice	*False dilemma* / Framing; *Inevitability framing* where independently satisfied	Restricted alternatives map to *False dilemma*. A contingent course represented as unavoidable or effectively without viable alternatives maps to *Inevitability framing*. Direct instruction alone licenses neither automatically.
Consequential Oversimplification	*Slippery-slope / consequential escalation* / Framing	The source definition explicitly concerns an initial event or action leading through a domino-like chain toward increasingly consequential or improbable outcomes. This corresponds to the ontology's consequential-escalation operation rather than to *Causal oversimplification*.
Call		
Slogans	No standalone Method/Tactic by form; conditional Narrative *Symbolic condensation*	A slogan may compress a broader narrative but does not automatically perform one unique Method-level operation.
Conversation Killer	No standalone canonical Tactic; applicable Tactic or NarrativeDevice where independently satisfied	The source category is the successor of thought-terminating-cliché style labels and partly characterises suppression of discussion or critical thought. The ontology does not infer that effect from a phrase alone. Agenda shifting, loaded language, obfuscation, symbolic condensation, or another construct is coded only where its own observable conditions are present.
Appeal to Time	No one-to-one Tactic; *Scarcity appeal* only where limited time/availability is itself the warrant; conditional Narrative *Stakes & urgency cues* where independently satisfied	The SemEval definition centres on the claim that the time has come for an action. Mere temporal exhortation does not satisfy the ontology's scarcity criterion, which requires limited continuing availability, nor the Narrative urgency criterion without independently represented time-critical stakes.
Manipulative Wording		
Loaded Language	*Loaded language* / Framing	Direct correspondence.
Obfuscation, Intentional Vagueness, Confusion	*Strategic vagueness / equivocation* / Deception; *Obfuscation (jargon/complexity)* / Distraction	The source label combines materially underdetermined communication with opacity and a presumed state of confusion. The ontology separates the observable operations and does not annotate confusion as a standalone audience state.

Supplementary Table S21: SemEval-2023 persuasion-technique crosswalk (continued)

SemEval-2023 technique	Canonical representation	Crosswalk rationale
Exaggeration or Minimisation	*Exaggeration / overstatement*; *Minimization / understatement* / Deception	Directional split of the source category.
Repetition	*Repetition* / Persuasion	Direct correspondence subject to the ontology's message-internal recurrence criterion.

SemEval coverage result. Across the four SemEval shared-task iterations considered here, changes in genre, language, modality, taxonomy granularity, and hierarchical organisation do not introduce an additional independently annotated message-internal dimension. The source technique inventories are either directly represented by canonical Tactics, decomposed into distinct canonical operations, or respecified where the source category is primarily an expressive form or is partly defined through inferred intent or audience effect. Multimodal realisation changes the evidence from which an operation may be recognised; it does not by itself define a separate influence dimension.

H.5 ENISA/EEAS FIMI (2022): scope comparison

The joint ENISA/EEAS report on Foreign Information Manipulation and Interference (FIMI) and cybersecurity does not provide a closed message-technique inventory analogous to the SemEval taxonomies. Rather, it proposes an analytical framework, consistent with the ENISA Threat Landscape methodology, for examining the relation between information manipulation and cybersecurity (European Union Agency for Cybersecurity (ENISA) and European External Action Service (EEAS), 2022). FIMI accordingly spans a broader unit of analysis than the present ontology: the report addresses the creation and dissemination of manipulative activity together with cybersecurity relations, incident analysis, attribution, and information sharing.

This difference in scope is substantively important. FIMI emphasises manipulative behaviour rather than making truthfulness of content the sole organising criterion. That orientation is compatible with the present ontology's truth- and intent-agnostic analysis of message-internal communicative operations. Where a FIMI case contains recoverable message content, that content can be analysed through Method/Tactic, Narrative, Polarization, AboutObject, and TopicDomain according to their respective codebook criteria. Operational attribution, threat-actor relations, cybersecurity activity, coordinated dissemination, incident structure, and other campaign-level properties are not thereby represented.

FIMI is therefore classified as *Partial overlap (scope)* rather than *Covered*. No category-by-category crosswalk is asserted here because the ENISA/EEAS framework is not treated as a closed taxonomy of message-internal influence techniques.

H.6 Tudela et al. (2025): Influence Operation Ontology crosswalk

The Influence Operation Ontology (IOO) is a cyber-threat-intelligence-oriented ontology for modelling influence operations across the information environment (Cayuela Tudela et al., 2025). Its architecture comprises three domains: *Threat*, *Channel*, and *Social*. This differs substantially from the present ontology's message-level measurement architecture and therefore provides a useful scope comparison rather than a one-to-one ontology alignment.

Supplementary Table S22: Influence Operation Ontology domains crosswalked to the message-internal ontology architecture.

IOO domain	Principal IOO constructs	Correspondence and boundary	Verdict
Threat	Incident; Attack Pattern; Campaign; Threat Actor	This domain models who conducts an influence operation, the incidents and campaigns in which actions occur, and the attack patterns through which they are conducted. Threat Actor, Campaign, and Incident have no direct message-internal equivalent. An IOO Attack Pattern may intersect with a canonical Method/Tactic when the relevant action is itself instantiated as a recoverable communicative operation in a message, but the broader TTP, campaign, attribution, and actor relations remain outside scope.	Partial overlap
Channel	Message; Channel; User Account and associated communication/channel properties	The IOO *Message* provides direct agreement on the importance of a message-level observable, but the two ontologies model it for different purposes. The present ontology analyses the communicative content of the message. IOO additionally models the account and channel through which it is published and related information-environment properties. User-account, platform, channel, amplification, and provenance metadata are therefore outside the message-internal analytical branches unless represented as content within the message itself.	Partial overlap

Supplementary Table S22: Influence Operation Ontology crosswalk (continued)

IOO domain	Principal IOO constructs	Correspondence and boundary	Verdict
Social	Cyber Persona; Community; Narrative; Event; Location	IOO models social and contextual entities that may participate in or become targets of an influence operation. Its *Narrative* construct provides conceptual overlap with the present ontology's interest in narrative meaning, but it is not equivalent to a canonical NarrativeDevice or NarrativeArchetype. Those labels describe message-internal narrative organisation under specific operational criteria. Cyber Personas, Communities, Events, and Locations may occur as represented entities in message content, but the present ontology does not maintain them as actor/context entities. AboutObject similarly records an independently recoverable civic or commercial outcome orientation rather than the identity of an IOO target entity.	Partial overlap

The resulting verdict is *Partial overlap (scope)*. IOO and the present ontology address complementary levels of representation: IOO models the broader socio-technical information environment and operational relations, whereas the present ontology decomposes the communicative structure of the message itself. Their common use of message and narrative concepts therefore does not imply equivalence between their respective classes.

H.7 DISARM: object-level scope crosswalk

DISARM is designed to describe and understand disinformation and influence incidents using tactics, techniques, and procedures modelled in a form inspired by information-security frameworks (DISARM Foundation, n.d.; Newman, 2022). Its object model contains higher-level phases, tactics, techniques, tasks, incident descriptions, and responder-side objects including counters and response types. The Red Team framework organises creator behaviour by operational stage, while the Blue Team framework organises countermeasures.

Accordingly, the relevant comparison is not whether every DISARM object maps to an ontology label, but which DISARM objects can contain a message-internal communicative operation of the kind measured here.

Supplementary Table S23: DISARM object types crosswalked to the message-internal ontology.

DISARM object / level	DISARM function	Ontology correspondence and boundary
Phases and tactics	Higher-level groupings and operational stages in the conduct of an information-manipulation incident	These are generally above the message level. A content-related tactic can nevertheless intersect with the present ontology where it entails production or use of recoverable message content. Such overlap does not make the DISARM operational stage itself a Method family or Tactic.
Techniques and tasks	Activities used to carry out an operational tactic, and tasks that need to be performed at a particular stage	Some techniques concern content creation and therefore admit message-internal analysis; others concern accounts, infrastructure, targeting, dissemination, amplification, coordination, or other operational behaviour outside scope. For example, DISARM technique *Develop New Narratives* under the *Develop Narratives* tactic concerns narrative development, while any individual resulting message must still satisfy the present codebook's own Narrative and Method/Tactic conditions.
Incidents	Descriptions of observed disinformation or influence incidents	Incident identity, temporal organisation, participating actors, and the combination of TTPs used across an operation are not message-level annotations. Messages occurring within an incident remain independently analysable.
Actor-related objects	Resources or actor types associated with conducting or responding to information operations	Actor identity, organisational affiliation, capability, coordination, and attribution are not inferred as message-internal labels. Representations of actors within message content may nevertheless participate in Narrative, Polarization, or AboutObject evidence where their respective criteria are met.
Counters, response types, and related responder objects	Countermeasures and categories for responding to observed influence TTPs	These describe interventions against influence operations rather than features of the influence message being annotated and are therefore outside the ontology's unit of analysis.

DISARM is consequently classified as *Partial overlap (scope)*. The framework contains genuine message-content intersections, but its principal purpose is the representation of operational behaviour across an incident lifecycle. Importantly, campaign-level dissemination or repeated posting in DISARM does not imply the canonical *Repetition* Tactic. Repetition is licensed in the present ontology only when the same or substantially equivalent proposition recurs within the available message or recoverable message sequence and that recurrence itself functions as reinforcement.

H.8 Lenk (2026): multilevel framing–narrative correspondence

Lenk proposes a four-level framework comprising semantic frames, strategic narratives, master-narratives, and meta-frames (Lenk, 2026). The levels are ordered by analytical abstraction rather than by message-internal annotation depth. This distinction is important for the present comparison: Lenk treats semantic frames and strategic narratives as directly extractable from a given text, whereas master-narratives and meta-frames require interpretation in relation to broader historical, ideological, political, or geopolitical contexts.

The correspondence is therefore strongest at Lenk's first two levels. Higher-level similarities with the present ontology's NarrativeArchetypes are conceptual rather than class-equivalent: a message may satisfy a canonical archetype signature without thereby licensing inference of a Lenk master-narrative or meta-frame.

Supplementary Table S24: Lenk's four analytical levels crosswalked to the message-internal ontology architecture.

Lenk level	**Closest ontology correspondence**	**Correspondence and boundary**
Level 1: Semantic frames	Applicable Method/Tactic and Narrative constructs	Lenk's semantic frames are lexical or semantic resources used to make particular aspects of an issue salient within a strategic narrative. They are not equivalent to the ontology's Framing family. A semantic realisation may independently satisfy canonical Tactics such as *Loaded language*, *Metaphor / analogy*, *Moral framing*, *Threat framing*, or another Framing operation where the respective codebook conditions are met. It may likewise supply evidence for a NarrativeDevice such as *Metaphorical mapping* or *Moral valuation*. No such annotation is inherited merely because Lenk identifies a semantic frame.
Level 2: Strategic narratives	Narrative; conditional Method/Tactic and AboutObject	This level has the closest correspondence with the ontology's Narrative branch because it concerns coherent message-level story organisation. Applicable messages may instantiate *Role-casting*, *Causal emplotment*, *Conflict–resolution arc*, *Temporal sequencing & contrast Moral valuation*, *Stakes & urgency cues*, or other NarrativeDevices. A deterministic NarrativeArchetype is assigned only where its own minimal signature is satisfied. Method/Tactic and AboutObject remain independently annotated and are not inherited from the presence of a strategic narrative.

Supplementary Table S24: Lenk level crosswalk (continued)

Lenk level	Closest ontology correspondence	Correspondence and boundary
Level 3: Master-narratives	No direct message-level equivalent	Lenk uses master-narratives to connect recurring strategic narratives through broader storylines that extend across texts and require contextual interpretation. The present ontology deliberately does not infer such a cross-message contextual category from a single message. Individual messages participating in a master-narrative may nevertheless instantiate one or more canonical NarrativeArchetypes, but those archetypes describe recoverable message structure rather than membership in the higher-level master-narrative itself.
Level 4: Meta-frames	No direct message-level equivalent	Meta-frames capture highly abstract and comparatively stable themes that organise master-narratives across cases. Their identification may require historical, ideological, political, or geopolitical contextual knowledge. Such contextual abstraction lies outside the ontology's message-internal annotation rule. A particular message may contain independently recoverable moral, identity, metaphorical, or group-positioning operations related to a meta-frame, but those annotations do not establish the meta-frame itself.

Lenk is consequently classified as *Partial overlap (scope)*. The first two levels provide substantial message-internal correspondence, whereas the upper two levels deliberately employ contextual abstraction beyond the present ontology's unit of analysis.

H.9 Institute for Propaganda Analysis (1939): seven-device crosswalk

The Institute for Propaganda Analysis (IPA) provides an important historical anchor for technique-based propaganda analysis. Its seven commonly cited devices are *Name-calling*, *Glittering generalities*, *Transfer*, *Testimonial*, *Plain folks*, *Card stacking*, and *Bandwagon* (Institute for Propaganda Analysis, 1939). The historical devices are broader and less operationally differentiated than the present Tactic inventory. Their coverage is therefore assessed by communicative function rather than by label similarity.

Supplementary Table S25: Institute for Propaganda Analysis devices crosswalked to canonical Method/Tactic operations.

IPA device	Canonical representation	Crosswalk rationale
Name-calling	*Name-calling / labeling* / Framing	Direct correspondence where a negative epithet or condensed evaluative label characterises an identifiable target. The canonical Tactic does not require the annotator to infer whether the audience accepts the condemnation.
Glittering generalities	No one-to-one Tactic; commonly *Loaded language*, *Moral framing*, or another value-specific operation	The historical device groups favourable "virtue" words and broad positive associations. The ontology instead distinguishes the function performed: evaluatively positive lexical choice can instantiate *Loaded language*; an identifiable moral value or principle can instantiate *Moral framing*; other specific warrants retain their own Tactics. Generic positivity alone does not define a canonical operation.
Transfer	*Associative transfer / guilt-or-honor by association* / Persuasion	This is the closest direct correspondence. Both concern use of an association with a positively or negatively regarded actor, institution, symbol, group, or other entity to transfer represented standing to a focal target. Shared-group identity used as a warrant is coded separately as *Unity (shared identity)* where independently satisfied.
Testimonial	*Appeal to authority*; *Associative transfer / guilt-or-honor by association*; *Source / channel credibility*, according to warrant	The IPA device groups endorsements by respected or disfavoured persons. The ontology decomposes testimonial by why the endorsement is supposed to carry weight. Expertise, credentials, or recognised authority license *Appeal to authority*; prestige, stigma, or evaluative standing transferred through association license *Associative transfer / guilt-or-honor by association*; and reliability or institutional provenance of the information source or channel licenses *Source / channel credibility*. Endorsement alone does not automatically license all three.
Plain folks	*Ingratiation* / Persuasion; conditional *Unity (shared identity)*	Common-person self-presentation is represented as *Ingratiation* where ordinariness, relatability, similarity, informality, or an "I am like you" presentation functions to cultivate a more favourable evaluation of the source. *Unity (shared identity)* applies instead or additionally only where membership in an identifiable shared collective or social category itself functions as the influence warrant.

Supplementary Table S25: Institute for Propaganda Analysis crosswalk (continued)

IPA device	Canonical representation	Crosswalk rationale
Card stacking	*Selective evidence presentation (cherry-picking)* / Deception; other applicable Deception or Distraction Tactic	The historical device encompasses a broad range of selective facts, omissions, illustrations, distractions, and arguments used to construct a one-sided case. The closest canonical correspondence is *Selective evidence presentation (cherry-picking)* where a systematically favourable subset is selected from a broader evidential or consideration set that is recoverable from the available message context. The broader IPA category may also instantiate another Deception or Distraction Tactic, but a merely one-sided argument or externally inferred omission is insufficient for cherry-picking.
Bandwagon	*Social proof (bandwagon)* / Persuasion	Direct correspondence where the beliefs, approval, choices, or conduct of multiple others function as social evidence for corresponding judgement or conduct.

The IPA inventory is therefore classified as *Covered*. This does not mean that the historical devices are reproduced as seven modern labels: several are decomposed because they combine distinct warranting or presentation relations that the ontology treats separately.

H.10 Entman (1993): framing-function crosswalk

Entman conceptualises framing as selecting aspects of perceived reality and making them salient in ways that may promote a problem definition, causal interpretation, moral evaluation, and/or treatment recommendation (Entman, 1993). These are framing functions rather than four labels that must co-occur: a particular text or frame may perform only a subset of them. The ontology therefore represents their message-internal realisations by decomposition rather than introducing four additional framing categories.

Supplementary Table S26: Entman's framing functions crosswalked to the ontology architecture.

Entman function	Ontology correspondence	Correspondence and boundary
Problem definition	TopicDomain; applicable Method/Tactic, Narrative, and AboutObject	TopicDomain indexes the substantive issue but does not itself assert that the issue is a "problem." The way a condition is constituted as problematic may instead satisfy an applicable Framing Tactic or NarrativeDevice—for example *Threat framing*, *Loaded language*, *Role-casting*, or *Stakes & urgency cues*—where their own conditions are met. AboutObject is added only where a broader civic or commercial outcome orientation is independently recoverable. There is therefore no automatic ProblemDefinition ontology label.
Causal interpretation	Narrative *Causal emplotment*; conditional Method/Tactic and AboutObject	Recoverable causal organisation can be represented through *Causal emplotment*. A restricted or exhaustive causal account additionally licenses *Causal oversimplification*; concentrated blame may license *Scapegoating*. The AboutObject *ResponsibilityAttribution-Shift* applies only where the message explicitly moves responsibility from one represented actor toward another, not to ordinary causal attribution.
Moral evaluation	Narrative *Moral valuation*; conditional *Moral framing* and Polarization	Narrative *Moral valuation* represents moral organisation of actors, actions, events, or outcomes within the story. The Method Tactic *Moral framing* applies where an identifiable moral value, right, obligation, virtue, vice, justice criterion, or prohibition performs the framing operation. Polarization is added only where the same message independently satisfies a canonical group-positioning motif; moral evaluation alone does not imply polarization.
Treatment recommendation	AboutObject; conditional Method/Tactic and TopicDomain	Where a proposed treatment bears toward a recoverable broader civic or commercial outcome, the applicable specific AboutObject can represent that orientation. TopicDomain indexes the substantive policy or issue field. Persuasive, framing, manipulative, coercive, or other Tactics are annotated only where their own communicative operations are independently satisfied. A treatment recommendation therefore does not directly assign *Behavior* or *Agency/Choice*; TargetConstruct values arise only through the fixed profile of a licensed Tactic.

Entman's four functions are therefore classified as *Covered*, with representation achieved by decomposition. The correspondence concerns their influence-relevant message-internal realisations rather than one-to-one reproduction of Entman's framing-function terminology.

H.11 Benford & Snow (2000): core collective-action framing tasks

Benford and Snow identify three "core framing tasks" in collective-action framing: diagnostic framing, prognostic framing, and motivational framing (Benford and Snow, 2000). Diagnostic framing concerns identification of a problematic condition and attribution of responsibility; prognostic framing articulates proposed solutions, plans, and strategies; and motivational framing supplies a rationale or "call to arms" for collective action. These tasks are broader discursive functions rather than canonical influence techniques and are therefore crosswalked by their message-internal realisation.

Supplementary Table S27: Benford and Snow's core framing tasks crosswalked to the ontology architecture.

Core framing task	Ontology correspondence	Correspondence and boundary
Diagnostic framing	Narrative; TopicDomain; conditional Method/Tactic, Polarization, and AboutObject	Problem identification can be indexed substantively through TopicDomain and organised narratively through devices such as *Role-casting* and *Causal emplotment*. Attribution of responsibility does not automatically constitute *Scapegoating* or *Causal oversimplification*; those Tactics require concentrated blame or restricted causal explanation respectively. Polarization applies only where a group-positioning motif is independently recoverable. *ResponsibilityAttribution-Shift* applies only where responsibility is explicitly moved from one actor toward another.
Prognostic framing	AboutObject; TopicDomain; conditional Narrative and Method/Tactic	A proposed solution, alternative arrangement, plan, or strategy can bear toward a specific AboutObject when the corresponding broader civic or commercial outcome is independently recoverable. TopicDomain indexes the substantive field. Narrative *Conflict–resolution arc* may apply where the proposed solution functions as resolution of an established conflict or disruption. Any persuasive or other influence Tactic must be licensed independently; prognostic framing itself does not directly assign a TargetConstruct.
Motivational framing	Conditional Method/Tactic and Narrative	A call or rationale for collective action may instantiate one or more conduct-directed canonical Tactics according to the warrant actually used. Narrative *Stakes & urgency cues* may apply where severe, escalating, time-critical, catastrophic, or existential stakes are independently represented. Other vocabularies of motive—including efficacy, propriety, severity, or urgency—are not themselves TargetConstruct labels. *Behavior*, *Agency/Choice*, or another TargetConstruct enters the representation only through the fixed mapping of an independently licensed Tactic.

The three core collective-action framing tasks are consequently classified as *Covered*, with representation achieved by decomposition. The verdict concerns the message-internal framing functions themselves; Benford and Snow's broader theory of social movement framing, frame resonance, movement organisation, mobilisation, and context is not claimed to be reproduced by the ontology.

H.12 Campaign- and operation-level frameworks: scope notes

Several comparison frameworks operate primarily above the message level. Detailed label-by-label crosswalks are therefore neither necessary nor appropriate for these frameworks. Their relevance to the coverage check lies instead in establishing which analytical objects overlap with message content and which require actor-, network-, campaign-, distribution-, contextual-, or effect-level evidence unavailable under the present ontology's unit of analysis.

Supplementary Table S28: Scope comparison with campaign- and operation-level frameworks.

Framework	Principal unit of analysis	Message-level correspondence and boundary
François ABC (François, 2019)	Manipulative Actors, Deceptive Behaviors, and Harmful Content	The *Content* vector provides the closest direct overlap with the present ontology because communicative content can be decomposed using Method/Tactic, Narrative, Polarization, AboutObject, and TopicDomain. The *Actor* vector requires actor identity and related evidence outside message form. The *Behavior* vector includes operational practices such as coordinated or inauthentic activity that generally require account, network, or campaign evidence. A behaviour can overlap with a canonical Tactic only where the relevant communicative operation is itself recoverable from the message. The ontology does not inherit the ABC framework's "manipulative," "deceptive," or "harmful" judgements merely from framework membership.
Pamment ABCDE (Pamment, 2020)	Actor, Behavior, Content, Degree, and Effect	*Content* again provides the closest direct correspondence. *Actor* and most *Behavior* properties require information about the source or operation; *Degree* concerns distribution, reach, scale, targeted audiences, and engagement; and *Effect* concerns the assessed impact or harms of the case. These are not message-internal annotation branches. In particular, AboutObject represents a broader civic or commercial outcome toward which the message bears; it does not assert that this outcome occurred. Likewise, TargetConstruct represents a fixed audience-state dimension constitutive of an independently licensed Tactic and must not be interpreted as an observed campaign effect.

Supplementary Table S28: Campaign- and operation-level scope comparison (continued)

Framework	Principal unit of analysis	Message-level correspondence and boundary
Johnson et al. CIC ontology (Johnson et al., 2021)	Semantic modelling of Cyber Influence Campaigns across physical and cyber actions, events, and their linkages	The CIC ontology is designed to represent campaign mechanics and causal or structural relations across physical and cyber realms rather than to provide a closed taxonomy of message-level influence operations. Where a represented cyber event contains an analysable message, its communicative content can be annotated independently using the present ontology. The cross-event, cross-realm, network, temporal, and campaign relations modelled by the CIC ontology remain outside the message-internal annotation unit.
OII Computational Propaganda (Oxford Internet Institute, n.d.)	Computationally mediated political manipulation involving automation, algorithms, social-media infrastructure, amplification, and organised actors	Computational-propaganda research provides important evidence about how content is produced, targeted, distributed, amplified, and coordinated. Those properties generally concern accounts, networks, algorithms, audiences, and campaign behaviour rather than the internal structure of one message. The content carried by such an operation remains independently analysable through the ontology, but bot use, algorithmic amplification, microtargeting, posting scale, or coordinated distribution does not automatically instantiate a canonical Tactic. Nor does inclusion in a computational-propaganda dataset establish external falsity or deceptive intent for message-level annotation.
Media Manipulation Casebook (Media Manipulation Casebook, n.d.; Technology and Social Change Project, n.d.)	Five-stage Media Manipulation Life Cycle and case-level variables describing campaign origins, dissemination, responses, mitigation, and adaptation	The Casebook explicitly studies campaigns through their order, scale, scope, actors, platforms, exploited vulnerabilities, tactics, outcomes, mitigation, and subsequent adaptation. Message content occurring within a case can be analysed through the present ontology, and a Casebook tactic may overlap with a canonical Tactic where the same observable communicative operation is independently satisfied. Life-cycle stage, campaign strategy, network terrain, attribution, dissemination, observable outcome, mitigation, and adaptation are not message-internal labels. A Casebook campaign therefore cannot be mechanically converted into an ontology annotation.
Nimmo & Hutchins Online Operations Kill Chain (Nimmo and Hutchins, 2023)	Observation-based ten-link tactical model of online operations, decomposing each top-level tactic into techniques and procedures	The kill chain describes observable operational behaviour ranging from asset acquisition and disguise through targeting, engagement, execution, and measures that enable operational longevity. Some techniques involve posting, engagement, or content and may therefore produce messages analysable by the present ontology. Nevertheless, a kill-chain tactic describes a stage or approach in an operation rather than a message-internal Method family. Infrastructure, account acquisition, cross-platform behaviour, coordination, evasion, targeting, and persistence remain outside scope.

Supplementary Table S28: Campaign- and operation-level scope comparison (continued)

Framework	Principal unit of analysis	Message-level correspondence and boundary
Pamment & Smith attribution framework (Pamment and Smith, 2022)	Attribution of information influence operations using technical, behavioural, and contextual evidence together with legal and ethical assessment	Message content may contribute contextual evidence within an attribution investigation, but attribution itself asks who is responsible for an operation and with what degree of evidentiary support. The framework combines message/context analysis with technical traces, behavioural TTP evidence, and data sources not available from message content alone. The present ontology therefore does not derive actor identity, organisational responsibility, state affiliation, operational intent, or an attribution confidence judgement. These are complementary analytical outputs rather than missing message-level ontology dimensions.

H.13 Coverage synthesis

The comparison produces a consistent division by unit of analysis. The message-level propaganda and persuasion inventories examined above map primarily to Method/Tactic, with decomposition where a source category combines observably different operations. Framing and narrative frameworks introduce broader discursive functions that can be represented through combinations of Narrative, applicable Method/Tactic operations, Polarization, AboutObject, and TopicDomain without requiring those branches to co-occur automatically. Campaign- and incident-level frameworks add analytical objects—including actors, infrastructure, platforms, coordination, dissemination, temporal sequencing, attribution, scale, and observed effects—that deliberately lie outside the ontology's message-internal unit of analysis.

Across the bounded comparison set, the coverage check did not identify an independently recoverable message-internal analytical distinction that could not be represented through the existing architecture and therefore required adding another peer annotation branch. Apparent gaps instead fell into four recurrent categories:

1. external judgements of truth, falsity, private intent, legality, or attribution;

2. actor-, account-, infrastructure-, platform-, network-, distribution-, coordination-, or campaign-level properties;

3. cross-message or contextual abstractions requiring historical, ideological, political, geopolitical, or other evidence beyond the available message; and

4. observed reach, audience response, campaign impact, or downstream effects.

These exclusions are substantive boundary decisions rather than residual "other" categories. They preserve the distinction between what can be supported from message-internal evidence and what requires additional information about the source, operation, audience, context, or external world.

The coverage result should therefore be interpreted as evidence of *bounded representational coverage*: within the frameworks examined, the ontology can represent the relevant message-internal distinctions without introducing an additional independently annotated dimension. It does not establish that the ontology exhausts every possible future framework, that the constructs are empirically valid, that independent annotators will identify them reliably, that represented outcomes occur in audiences or society, or that the resulting annotations support predictive performance. Those questions require separate empirical evaluation.

Future candidate frameworks can be evaluated using the same criterion. A framework warrants extension of the message-level architecture only if it introduces an independently recoverable message-internal distinction that cannot be represented through the existing canonical branches or by principled decomposition. Differences arising solely from terminology, granularity, modality, campaign context, actor information, external truth judgements, or observed effects do not by themselves establish a missing message-level dimension.

I Development-Stage Annotation Exercise and Specification-Revision Cases

Purpose and scope

This appendix reports a development-stage annotation exercise conducted to examine whether an analyst previously unfamiliar with the ontology could complete the specified annotation workflow using the then-current codebook and annotation interface, and to obtain descriptive information about the effort associated with that process.

The exercise preceded final specification freeze. It was not designed to assess annotation accuracy, inter-annotator agreement, construct validity, or predictive performance. Only one analyst participated, no independently adjudicated reference annotations were available, and the analyst's individual label assignments were not subsequently validated for correctness. The resulting annotations therefore should not be interpreted as examples of correct application of either the development-stage or final specification.

Subsequent specification review refined parts of the canonical vocabulary, boundary rules, and deterministic mappings. Empirical assessment of

annotation reliability under the final specification remains future work.

I.1 Data and procedure

The exercise used 34 social-media messages selected for development-stage annotation. Two additional warm-up items preceded the exercise and are excluded from the descriptive summary below. The analyst had prior experience with qualitative content annotation but no prior exposure to the ontology.

All items were annotated during a single session using a purpose-built annotation interface and the development-stage operational codebook. The interface recorded active annotation time automatically. Apparent inactivity periods longer than two minutes were excluded from active-time calculations.

For publication, the message excerpts retained below are lightly paraphrased and de-identified versions of the source messages used in the exercise. These editorial changes were made to reduce identifiability and avoid verbatim reproduction while preserving the rhetorical and linguistic material relevant to the comparison. The analyst's original selections refer to the original source messages; the publication-safe excerpts are used only to make the specification-revision cases interpretable to the reader.

I.2 Descriptive annotation-effort profile

Table S29 reports descriptive timing and message-length information for the 34-item exercise. Because the exercise involved one analyst using a development-stage instrument, these quantities describe only that annotation session and should not be generalised to the final specification.

Supplementary Table S29: Descriptive effort profile for the 34-item development-stage annotation exercise.

Measure	Mean	Median	Min	Max
Active time per message (s)	795	750	109	1,504
Text length (characters)	260	255	47	411
Total active annotation time (min)	450 (approx. 7.5 hours)			

The exercise shows that completion of the development-stage workflow was time-intensive in this setting: mean active annotation time was approximately 13.3 minutes per message, with substantial message-level variation. No inference is made from these observations about the time requirements of the final specification. Message characteristics, analyst decisions, familiarity with the instrument, interface design, and subsequent specification changes may all affect annotation time.

I.3 Development-stage annotation interface

The exercise used a purpose-built span-oriented annotation interface. A persistent panel displayed the message text, span highlights, recorded selections, and an event log. A tabbed annotation panel exposed the then-current codebook categories across the major annotation branches. The analyst selected a message span and then assigned one or more available labels; the interface stored span offsets, label identifiers, and annotation events in machine-readable form. Active and paused time were tracked automatically.

The interface additionally displayed the then-current narrative-signature information to assist navigation of that part of the codebook. Its role was procedural rather than adjudicative: the interface exposed available choices and recorded the analyst's selections but did not determine whether those selections were correct.

This interface description is retained to document how the development-stage timing observations were produced. It does not prescribe the implementation of the final specification or a future production annotation workflow.

I.4 Specification-revision cases

Three messages from the development-stage exercise are retained below to illustrate the distinction between *annotation migration* and *annotation adjudication*. The purpose of these cases is not to evaluate the analyst as an individual, but to show how specification change and annotation error can produce superficially similar differences in recorded labels.

For each case, three representations are distinguished:

1. **Development-stage analyst record.** This summarises the selections actually recorded by the analyst under the codebook available during the exercise. Deprecated vocabulary is not reproduced exhaustively; historical selections are described only to the level needed to understand the comparison.

2. **Schema-migrated representation.** This translates the analyst's recorded selections into the final canonical vocabulary and recomputes deterministic relations where a valid mapping exists. It deliberately preserves the analyst's underlying interpretive choices. Migration therefore answers the question, "What would this recorded annotation look like under the current schema if we did not reconsider whether the analyst's selections were substantively correct?"

3. **Final-codebook adjudication.** This is a new assessment of the publication-safe message against the frozen operational definitions and boundary rules. Labels unsupported by the final evidence criteria

are removed; labels independently supported by the message may be added. Only this third representation is intended as a normative example of how the final specification should be applied.

The development-stage analyst record is summarised in the surrounding prose; the comparison tables juxtapose its schema-migrated representation with the fresh final-codebook adjudication.

The distinction is necessary because a difference between the original record and the final adjudication can arise for at least two reasons. A label or deterministic relation may have changed during specification refinement, or the analyst may have selected a construct that the message evidence did not adequately support. Mechanical migration addresses the former; fresh adjudication addresses the latter.

Case 1: specification migration versus substantive adjudication.

> *"They're calling this government measure 'transitional,' but rights you surrender almost never come back."*

The development-stage analyst marked the second clause as a dense multi-branch cue. The case is useful because some differences under the final specification arise mechanically from changed derivation rules, whereas others arise only after the message is freshly adjudicated against the final evidence criteria.

Interpretation. The migration and adjudication columns answer different questions. Under migration, the analyst's original fear and exaggeration selections are preserved, but their current deterministic consequences change: *Appeal to fear* now contributes only *Agency/Choice* and *Behavior*, while *Exaggeration / overstatement* contributes *Belief* and *Attitude*. The old Polarization selections cannot be converted into valid final annotations because the final architecture requires an independently assigned motif before direction and axis values are derived.

Fresh adjudication changes the substantive label set. The warning that surrendered rights are unlikely to return supplies a recoverable harmful consequence for the implied protective response of not surrendering those rights, supporting *Appeal to fear*. The excerpt does not, however, supply the recoverable comparison point or evidential baseline required for *Exaggeration / overstatement*; the strength of "almost never" is insufficient by itself. The message instead frames the measure through the moral criterion of rights and uses "surrender" as independently evaluatively loaded wording, supporting *Moral framing* and *Loaded language*.

At the Narrative level, the present-to-future relation between surrender and non-restoration supports *Temporal sequencing & contrast*, but the excerpt does not require a distinct causal emplotment or a severe, escalating,

Supplementary Table S30: Case 1: schema migration of the development-stage analyst record and final-codebook adjudication.

Branch	**Schema-migrated analyst record**	**Final-codebook adjudication**
Method / Tactic	*Appeal to fear* → Persuasion; *Exaggeration / overstatement* → Deception	*Appeal to fear* → Persuasion; *Moral framing* → Framing; *Loaded language* → Framing
TargetConstruct	*Appeal to fear*: {Agency/Choice, Behavior}; *Exaggeration / overstatement*: {Belief, Attitude}	Derived only from the adjudicated Tactics: {Agency/Choice, Behavior, Belief, Attitude}
Narrative	*Causal emplotment*; *Temporal sequencing & contrast*; *Stakes & urgency cues*	*Temporal sequencing & contrast*
Polarization	The development-stage record contained two Divide direction/axis annotations. No current motif can be recovered from those values by mechanical migration.	None. No salient social or political group-relational positioning operation is recoverable from the displayed message.
AboutObject	Historical institutional-delegitimization selection normalised, without re-adjudication, to `InstitutionalLegitimacy-Decrease` → Epistemic Trust, Legitimacy & Attribution → CivicOutcome	`Policy-Opposition` → Politics → CivicOutcome
TopicDomain	No value is reconstructed solely from the historical record shown here.	Not recoverable with sufficient substantive specificity from the publication-safe excerpt. The excerpt establishes that the target is a governmental measure but does not preserve the policy field to which that measure belongs.

catastrophic, existential, or time-critical stakes structure. No salient intergroup relation supports a Polarization motif. Finally, opposition to the represented government measure is recoverable as the substantive AboutObject, whereas criticism of the measure does not by itself question the rightful standing or authority of an institution.

Case 2: branch independence and group-boundary adjudication.

> *"Local law enforcement official claims the commentator 'disguised himself as a person of faith' in online posts. Source: regional news outlet. This is the TRUTH and his supporters are furious they've been called out on it. Their entire movement is built on a LIE—you cannot be a person of faith and also spread HATE."*

The development-stage record assigned multiple Persuasion and Framing Tactics, several NarrativeDevices, repeated Polarization annotations, and group-level civic outcomes. The final specification materially changes the interpretation because several superficially plausible labels no longer satisfy their operational boundaries and because group-positioning is now represented through canonical motifs rather than directly selected direction/axis combinations.

Supplementary Table S31: Case 2: schema migration of the development-stage analyst record and final-codebook adjudication.

Branch	**Schema-migrated analyst record**	**Final-codebook adjudication**
Method / Tactic	*Source / channel credibility*; *Appeal to authority*; *Moral framing*; *Two-sided with refutation*; *Loaded language*; *Exaggeration / overstatement*; *False dilemma*	*Appeal to authority* → Persuasion; *Moral framing* → Framing; *Loaded language* → Framing
TargetConstruct	Recomputed from the migrated Tactics rather than preserving the analyst's development-stage TargetConstruct selections.	Derived from the adjudicated Tactics: {Belief, Attitude}

Branch	Schema-migrated analyst record	Final-codebook adjudication
Narrative	*Symbolic condensation*; *Moral valuation*; *Role-casting (actants)*	*Moral valuation*
Polarization	The development-stage record contained several directly selected Divide/Cohere direction-and-axis combinations. These cannot be converted mechanically into valid final Polarization annotations because the final specification requires an independently assigned canonical motif.	*Membership denial* → Divide → {Affective/Identity}; *Moral condemnation* → Divide → {Affective/Identity}; *Homogenization* → Divide → {Perception/Distance}
AboutObject	The development-stage polarization-increase outcome has a direct current counterpart, `Polarization-Increase`. The historical group-vilification outcome has no unique one-to-one successor in the final AboutObject vocabulary and is therefore not mechanically resolved.	`GroupInclusion-Decrease`; `CollectiveIdentity-Reconfiguration` Both derive through Identity, Intergroup Relations & Social Cohesion → CivicOutcome.
TopicDomain	No development-stage Topic selection is reconstructed here.	*Religion*

Interpretation. Mechanical migration preserves the analyst's underlying selections wherever a valid current counterpart exists, even when a fresh application of the final codebook would reject those selections. Thus *Source / channel credibility*, *Two-sided with refutation*, *Exaggeration / overstatement*, and *False dilemma* remain visible in the migrated representation but are removed by adjudication.

The final codebook distinguishes the authority of the person making a claim from the provenance of the channel through which that claim is re-

ported. In this message, the source's status as a local law-enforcement official is foregrounded as evidential support for the attributed claim rather than serving only as source identification, supporting *Appeal to authority*. By contrast, the bare statement "Source: regional news outlet" does not independently establish that the outlet's reputation or reliability is being used as a credibility warrant. Likewise, assertion that a proposition is "the TRUTH" does not constitute *Two-sided with refutation* without an identifiable opposing proposition and substantive rebuttal, and the categorical phrase "their entire movement" does not satisfy *Exaggeration / overstatement* without a recoverable comparison or evidential baseline.

The final clause is also not treated as a *False dilemma*. Its principal operation is to define an authentic group-membership boundary: spreading hate is represented as incompatible with being a "person of faith." This supports *Moral framing* at the Method level and, independently, the Polarization motif *Membership denial*. The broader claims that the movement is founded on a lie and morally associated with hate additionally support *Moral condemnation*, while the generalisation to the "entire movement" supports *Homogenization*. All three motifs derive Divide, but their axis profiles remain motif-specific.

At the outcome level, the same membership claim independently bears toward reduced acceptance of the represented movement and supporter constituency as authentically belonging within the religious collective and toward a normative redefinition of what such membership entails, supporting `GroupInclusion-Decrease` and `CollectiveIdentity-Reconfiguration`. Neither `Polarization-Increase` nor an outgroup-hostility outcome is inherited merely because Divide motifs are present.

The TopicDomain is *Religion* because the substantive issue being organised and contested is authentic religious identity and the normative conditions of membership; the law-enforcement official and news outlet function as sources of information rather than as the substantive topic of the message.

Case 3: dehumanizing representation across independent branches.

> *"Over 100 House Democrats refused to vote for the resolution condemning the commentator's killing. [partisan news outlet/. . .] [58 FILTH voted against; the rest were too spineless to vote at all—or hid behind 'present'. . . LEFTISTS: A PLAGUE UPON THIS NATION]"*

The development-stage record treated this message primarily as dense multi-tactic stacking involving shaming, personal attack, moral and causal framing, selective evidence, and other operations. Fresh adjudication under the final specification yields a different structure: the recoverable Method

operations are concentrated in the Framing family, while the contamination representation independently licenses Narrative and Polarization structure.

Supplementary Table S32: Case 3: schema migration of the development-stage analyst record and final-codebook adjudication.

Branch	**Schema-migrated analyst record**	**Final-codebook adjudication**
Method / Tactic	*Shaming / derogation*; *Loaded language*; *Moral framing*; *Causal oversimplification*; *Name-calling / labeling*; *Selective evidence presentation (cherry-picking)*; *Ingratiation*. The development-stage personal-attack selection has no lossless current counterpart; *Ad hominem / personal-credibility attack* is narrower and therefore requires re-adjudication.	*Loaded language*; *Name-calling / labeling*; *Moral framing*; *Metaphor / analogy*; *Dehumanizing representation*; *Threat framing* All six derive to *Framing*.
TargetConstruct	Recomputed from current mappings of the mechanically migrated Tactics rather than preserving the analyst's development-stage TargetConstruct union.	Derived from the adjudicated Tactics: {Belief, Attitude}
Narrative	*Moral valuation*; *Role-casting (actants)*	*Moral valuation*; *Metaphorical mapping*, `Type=disease/parasite/contamination`; derived NarrativeArchetype: *Contamination / disease*
Polarization	The development-stage record contained directly selected Divide direction-and-axis combinations. These do not identify a unique final motif and therefore cannot be mechanically migrated as valid final Polarization annotations.	*Dehumanization* → Divide → {Affective/Identity}; *Identity-based ridicule / disparagement* → Divide → {Affective/Identity}; *Homogenization* → Divide → {Perception/Distance}

Continued on next page

Supplementary Table S32: Case 3: schema migration and final-codebook adjudication (continued)

Branch	**Schema-migrated analyst record**	**Final-codebook adjudication**
AboutObject	The historical polarization-increase selection has a direct current counterpart, `Polarization-Increase`; the historical group-vilification outcome has no unique one-to-one successor. The historical institutional-delegitimization selection can be normalised mechanically to `InstitutionalLegitimacy-Decrease` without implying that it is substantively correct.	`PoliticalActorSupport-Decrease` → Politics; `OutgroupHostility-Increase`, `GroupHumanization-Decrease`, and `Polarization-Increase` → Identity, Intergroup Relations & Social Cohesion. All four inherit `CivicOutcome`.
TopicDomain	No development-stage Topic selection is reconstructed here.	*Elections & Governance*

Interpretation. The final adjudication removes several development-stage selections for different reasons. Retrospective insults directed at legislators do not constitute *Shaming / derogation* unless negative self- or social evaluation functions as pressure on recipient conduct. Nor do they constitute *Ad hominem / personal-credibility attack* unless a person-directed characterisation is used to discount that person's claim, argument, testimony, or other contribution. No recoverable restricted causal account licenses *Causal oversimplification*, no recoverable broader evidential set licenses *Selective evidence presentation (cherry-picking)*, and no source–audience affinity operation licenses *Ingratiation*.

The strongest final Method evidence lies instead in the message's evaluative and nonhuman representation of the political targets. "FILTH" and "PLAGUE" function as condensed evaluative labels and as contaminating or nonhuman characterisations, while "too spineless" and "hid behind" provide independently loaded and morally evaluative framing. The statement "LEFTISTS: A PLAGUE UPON THIS NATION" additionally maps a disease/contamination domain onto the political group–nation relation and represents the group as a source of harm to the nation, independently supporting *Metaphor / analogy* and *Threat framing*.

The same material has different consequences in the Narrative branch. The disease/contamination metaphor supports *Metaphorical mapping* with `Type = disease/parasite/contamination`; that attribute satisfies

the fixed minimal signature for the *Contamination / disease* NarrativeArchetype, which is therefore derived deterministically.

At the Polarization level, the group-level contamination representation independently satisfies *Dehumanization*; categorical contempt toward the political identity supports *Identity-based ridicule / disparagement*; and the generalised characterisation of "LEFTISTS" as a plague supports *Homogenization*. These motifs derive their respective direction and axis sets. The Method-level *Threat framing* annotation does not create the Polarization motif *Existential threat*: the displayed message represents the outgroup as harmful but does not represent the continued physical existence or defining identity of the national collective as being destroyed or erased. Likewise, the Method-level *Dehumanizing representation* Tactic does not mechanically create either the Polarization motif *Dehumanization* or the AboutObject `GroupHumanization-Decrease`; each is retained here because its additional group-relational or broader-outcome criterion is independently satisfied.

Finally, the AboutObject labels are adjudicated independently from those message-level operations. The sweeping attack on House Democrats bears toward reduced support for the represented political actors; the contemptuous outgroup characterisation independently bears toward increased hostility, while the sustained categorical representation of the political outgroup as "FILTH" and "A PLAGUE UPON THIS NATION" supports the broader outcome of reduced recognition of that group's ordinary human standing; and the outgroup-as-contaminant relation bears toward a harder antagonistic intergroup condition. None of these outcomes is inherited merely from the corresponding Method or Polarization label.

I.5 Lessons from specification revision

The three cases demonstrate why development-stage annotation records, specification migration, and final-codebook adjudication must be kept analytically distinct.

First, *schema compatibility is not annotation validity.* A historical selection may have a recognisable counterpart in the final vocabulary and may therefore be migrated mechanically, while still failing the final construct's evidence requirements when the message is freshly adjudicated. Conversely, a final-codebook annotation may be supported by message evidence even though the corresponding construct was absent from the development-stage record.

Second, *deterministic derivation does not eliminate interpretive judgement at the point of base-label assignment.* Once a canonical Tactic is licensed, its Method family and TargetConstruct profile are fixed. Once a Polarization motif is licensed, its direction and axis set are fixed. Once a qualifying NarrativeDevice attribute configuration satisfies an archetype signature, the corresponding NarrativeArchetype is fixed; and once a specific

AboutObject is assigned, its higher-level domain and sphere are fixed. The difficult annotation decision nevertheless remains whether the message evidence satisfies the relevant canonical base definition in the first place.

Third, *identical derived output can conceal different underlying analyses.* In Case 1, the mechanically migrated and freshly adjudicated Tactic sets differ substantively while producing exactly the same aggregate TargetConstruct profile: {Belief, Attitude, Agency/Choice, Behavior}. Cases 2 and 3 illustrate the same many-to-one property across different Tactic configurations: three adjudicated Tactics in Case 2 and six in Case 3 both yield the aggregate profile {Belief, Attitude}. Derived constructs therefore cannot be used retrospectively to validate the canonical base annotations from which they were generated.

Fourth, *semantic overlap across branches does not imply inheritance.* Case 2 shows that moral and identity-relevant language can independently support Method, Narrative, Polarization, and AboutObject annotations without any one branch mechanically licensing the others. Case 3 makes the same principle especially visible: a contamination representation can satisfy the Method-level *Dehumanizing representation* Tactic, a Narrative *Metaphorical mapping* device and resulting *Contamination / disease* archetype, and the Polarization *Dehumanization* motif, but only because the evidence conditions for each construct are independently satisfied.

Finally, the revision cases illustrate the role of increasingly explicit boundary rules in reducing uncontrolled label proliferation. Several development-stage selections that were superficially plausible—for example generic personal attack, shaming, exaggeration, causal oversimplification, or directly selected polarization directions—do not survive the final specification because their canonical constructs now require more specific observable relations. The refinement therefore changes not only terminology but also the discriminant structure of the measurement instrument.

J Ontology provenance and decision ledger

Purpose and status. This appendix documents the conceptual provenance and principal design decisions underlying the ontology's controlled vocabularies. Its purpose is to make the transition from literature constructs and candidate categories to the final operational specification inspectable and traceable. The status-coded decision tables record 127 row-level decisions linking candidate constructs or overlaps to the final specification, including 11 REJECT dispositions concerning candidates that were redundant, insufficiently distinct, dependent on unavailable evidence, or outside the message-internal scope.

The operational appendices remain authoritative for annotation. The present appendix does not create additional annotation labels and does not supersede their definitions, minimum-evidence conditions, boundary rules,

or mappings. Where wording in this provenance record is abbreviated, the corresponding canonical entry in the operational appendix governs annotation.

Decision-status vocabulary. In provenance and decision-ledger tables containing a *Status / qualifier* column, each entry is assigned one of six primary statuses. A short qualifier may further describe the transformation, boundary, or operational condition; qualifiers are explanatory and do not constitute additional status values.

- **KEEP**: a literature-supported construct or influence operation is retained as a canonical category, potentially with operational refinement or canonical renaming for message-internal annotation;
- **MERGE**: a candidate is not retained as a separate category because its relevant content is represented by one or more existing canonical categories. This includes cases described as decomposition where the candidate is distributed across existing operations rather than retained independently;
- **SPLIT**: a broader or ambiguous candidate construct is divided because distinct observable communicative operations, relations, layers, or endpoints must be preserved separately;
- **RESCOPE**: a construct is retained after narrowing or changing the unit, evidential basis, or conditions under which it is recoverable from message-internal evidence;
- **REJECT**: a candidate is not retained as a canonical category because it is outside the ontology's scope or unit of analysis, fails the inclusion criteria, or otherwise cannot be represented as an admissible canonical construct;
- **BOUNDARY**: no candidate disposition is being made; the entry instead records a rule distinguishing already represented constructs, preventing inheritance, specifying co-coding or specificity conditions, or otherwise constraining how canonical categories relate.

Statuses classify the specific decision recorded in a ledger entry rather than a literature term globally. A candidate may therefore appear in more than one table where distinct decisions are being documented, for example when a family-level representation or boundary decision is recorded separately from the cross-family decision about whether the candidate is retained as a standalone canonical Tactic. **REJECT** concerns non-retention of the candidate as the stated standalone canonical category; it does not imply that particular observable manifestations cannot be represented by other canonical categories.

Verification scope. Provenance transparency and specification reproducibility are distinct from empirical annotation reliability or predictive validity. The tables below permit a reader to inspect the literature basis, category transformation, canonical destination, and principal boundary rationale for the ontology specification. They do not by themselves establish that independent annotators achieve a particular level of agreement. Empirical annotation reliability and predictive validity are outside the scope of the present specification study and are reserved for future validation work.

TargetConstruct mappings. TargetConstruct values reported here reproduce the fixed mappings in Appendices A and B. They are not independently selected or re-derived in this appendix. The mapping reflects the constitutive operation of each Tactic rather than every psychological mechanism or possible downstream effect.

J.1 Method and Tactic provenance

The Method dimension contains the canonical message-internal influence operations specified in Appendix B. Provenance is reported at the Tactic level because Tactic is the annotated operational category; Method family and TargetConstruct values are inherited deterministically after Tactic assignment.

The provenance table distinguishes the literature construct or research tradition motivating each Tactic from the ontology's operational interpretation of that construct. A literature construct is not reproduced mechanically: where necessary, it is narrowed to observable message-internal evidence, separated from neighbouring operations, or integrated with another candidate category under the ontology's final boundary rules. Method-family closure and completeness of the 71 canonical Tactic mappings are verified jointly with the fixed TargetConstruct mappings under Section J.2.

J.1.1 Persuasion

Supplementary Table S33: Provenance of canonical Persuasion Tactics.

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Foot-in-the-door (FITD)	Beh	Foot-in-the-door sequential-request compliance	**KEEP.** Completed compliance with a smaller request is used as the message-internally recoverable step toward a subsequent larger request. Mere presentation of differently sized requests is insufficient.	(Freedman and Fraser, 1966; Burger, 1999; Cialdini and Goldstein, 2004)

Supplementary Table S33: Provenance of canonical Persuasion Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Door-in-the-face (DITF)	Beh	Door-in-the-face and reciprocal-concession procedures	**KEEP.** A rejected larger request is followed by a smaller target request presented as a retreat or concession. The rejection–retreat sequence distinguishes the Tactic from FITD and ordinary request reduction.	(Cialdini et al., 1975; O'Keefe and Hale, 1998; Cialdini and Goldstein, 2004)
Appeal to authority	Bel	Expertise, authority, and heuristic source cues	**KEEP; FUNCTIONALLY SPLIT.** Expertise, credentials, professional status, or recognised authority are retained here when they function as an epistemic warrant for accepting a claim, judgement, or recommendation. Authority used instead as a warrant for conduct is represented separately by *Authority-based directive*.	(Cialdini, 2007; Petty and Cacioppo, 1986; Chaiken, 1980)
Authority-based directive	Beh	Authority, legitimate power, and obedience	**SPLIT / ADD.** Represented institutional, hierarchical, legal, organisational, professional, or comparable authority is used to direct identifiable conduct. The authority relation itself must warrant compliance; a title, rank, or imperative alone is insufficient.	(Raven, 1993; Tyler, 2006; Milgram, 1963; Cialdini and Goldstein, 2004)
Social proof (bandwagon)	Bel; Beh	Informational and normative social influence; social proof	**KEEP.** Information about others' beliefs, approval, choices, or behaviour functions as social evidence for corresponding judgement or conduct. Shared identity is not required and is represented separately by *Unity (shared identity)*.	(Deutsch and Gerard, 1955; Cialdini et al., 1990; Schultz et al., 2007; Goldstein et al., 2008; Cialdini, 2007)
Appeal to tradition	Bel; Att	Argument or persuasion from longstanding customary status	**KEEP.** Historical persistence, inherited practice, or customary status must itself function as a warrant for correctness, legitimacy, preference, or retention. Historical reference alone is insufficient.	(Michaud, 2018)
Appeal to novelty	Bel; Att	Argument or persuasion from novelty, modernity, or innovation	**KEEP.** Newness, recency, modernity, innovation, or technological advancement must itself function as a warrant for correctness, credibility, desirability, or superiority. It is separated from inevitability, which concerns represented unavoidability rather than newness.	(Biržietienė and Gabrėnaitė, 2014)

Supplementary Table S33: Provenance of canonical Persuasion Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Ingratiation	Att	Ingratiation, other-enhancement, opinion conformity, similarity, and favourable self-presentation	**KEEP + MERGE.** Strategic interpersonal presentation is retained where it cultivates favourable evaluation of the source. Historical "plain-folks" or common-person self-presentation is absorbed here when ordinariness, similarity, or relatability serves that source-favourability operation. Shared collective identity as a warrant remains *Unity*.	(Jones, 1964; Gordon, 1996; Cialdini and Goldstein, 2004)
Appeal to fear	A/C; Beh	Fear appeals and protection-motivation models	**KEEP; BOUNDARY REFINED.** A represented threat or harmful consequence must function as a reason for an identifiable protective response. Mere danger characterisation is handled by *Threat framing*; prospective harm appropriated as contingent leverage over conduct is handled by Coercion.	(Rogers, 1975; Witte and Allen, 2000; Tannenbaum et al., 2015)
Appeal to anger / indignation	Bel; Att; Beh	Anger appeals, appraisal tendencies, and anger-based persuasion	**KEEP; TARGET PROFILE REFINED.** An identifiable offence, injustice, betrayal, violation, or comparable anger-relevant appraisal must be recruited as persuasive leverage for a proposition, evaluation, judgement, position, or response. Angry tone or likely audience anger alone is insufficient.	(Walter et al., 2019; Skurka, 2019; Turner et al., 2020; Lerner and Keltner, 2001)
Storytelling / narrative packaging	Bel; Att	Narrative persuasion and transportation	**KEEP; CROSS-LAYER BOUNDARY.** Story form is a Persuasion Tactic only where an observable narrative sequence functions as the vehicle for a claim or evaluation. Narrative structure that merely describes how a message is organised remains representable in the Narrative dimension without automatically instantiating this Tactic.	(Green and Brock, 2000; Braddock and Dillard, 2016; Green and Appel, 2024)

Supplementary Table S33: Provenance of canonical Persuasion Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Two-sided with refutation	Bel	Two-sided messages, refutational messages, and content-based inoculation	**KEEP.** An identifiable opposing proposition or argument is presented and substantively rebutted. Preemptive timing alone does not create a separate operation; prospective preparation for later influence is represented independently by *Preemptive inoculation / prebunking*.	(Allen, 1991; O'Keefe, 1999; Banas and Rains, 2010)
Preemptive inoculation / prebunking	Bel; Att; A/C	Inoculation theory and prebunking / misinformation-resistance interventions	**KEEP; PROSPECTIVE OPERATION RETAINED.** The defining distinction is an explicitly recoverable orientation toward preparing the recipient to resist, discount, scrutinise, recognise, or counter anticipated later influence. Technique-based prebunking can satisfy the Tactic without rebutting a particular proposition.	(Banas and Rains, 2010; Compton et al., 2021; Lewandowsky and Linden, 2021; Roozenbeek et al., 2022b)
Doubt-seeding	Bel	Manufactured uncertainty, production of doubt, and uncertainty-based persuasion	**KEEP; EPISTEMIC WARRANT SPECIFIED.** Uncertainty, disagreement, evidential limitations, source-reliability concerns, or procedural limitations are used as reasons for weakening confidence in or suspending judgement toward an identifiable epistemic target. The constitutive operation therefore bears directly on epistemic confidence and is represented as a Persuasion Tactic with fixed TargetConstruct *Belief*.	(Michaels, 2006; Stocking and Holstein, 2009; Biddle and Leuschner, 2015; Gustafson and Rice, 2020)
Presuppositional embedding	Bel	Linguistic presupposition and accommodation	**KEEP; FORMS CONSOLIDATED.** Material content not already established in the recoverable context is introduced as presupposed or taken for granted rather than focal assertion. "Loaded" or complex questions are treated as possible interrogative realisations of this operation rather than as a separate Tactic.	(Thoma et al., 2023; Macagno, 2024; Lombardi Vallauri, 2021)

Supplementary Table S33: Provenance of canonical Persuasion Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Repetition	Bel	Repeated-exposure and illusory-truth research	**RESCOPE / KEEP.** Repetition is retained where the same or substantially equivalent proposition recurs within the available message or recoverable message sequence and recurrence itself functions as reinforcement. Campaign-level scheduling, amplification, or repetition that is not message-internally recoverable remains outside scope.	(Hasher et al., 1977; Fazio et al., 2015; Ye et al., 2026)
Statistical evidence	Bel	Statistical and quantitative evidence in persuasion	**KEEP.** Numerical or statistical information must function as evidential support for a separate proposition, relation, estimate, or inference. Numbers or visualised quantities alone do not instantiate the Tactic.	(Allen and Preiss, 1997; Hornikx, 2005; Hoeken and Hustinx, 2009)
Source / channel credibility	Bel	Source, media, and channel credibility	**KEEP; SOURCE TYPE DISTINGUISHED.** The provenance, reliability, institutional standing, publication venue, outlet, or channel through which information is communicated functions as the credibility warrant. Expertise or authority of an endorsing actor is instead represented by *Appeal to authority*.	(Hovland and Weiss, 1951; Metzger et al., 2003; Kiousis, 2001; Sanz-Menéndez and Cruz-Castro, 2019)
Associative transfer / guilt-or-honor by association	Bel; Att	Argument by association, propaganda transfer, and evaluative conditioning	**KEEP + MERGE: directions.** An association functions as a warrant for transferring represented commitments, qualities, credibility, prestige, stigma, blame, approval, or other standing to the focal target. "Guilt" and "honor" by association are treated as negative and positive realisations of the same association-based operation rather than separate Tactics.	(Mohammed, 2023; Jowett and O'Donnell, 2018; Moran et al., 2023)
Reciprocity	Beh	Norm of reciprocity and reciprocal compliance	**KEEP.** A benefit, favour, gift, service, or concession already provided to the recipient functions as a warrant for a corresponding return act. Prospective reward contingencies and DITF request sequences are represented separately unless a reciprocal return warrant is independently present.	(Gouldner, 1960; Regan, 1971; Cialdini and Goldstein, 2004)

Supplementary Table S33: Provenance of canonical Persuasion Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Commitment / consistency	Beh	Commitment, consistency, and preference for consistency	**KEEP.** A prior explicit choice, pledge, promise, or commitment is itself invoked as a reason for corresponding follow-through. Prior behaviour that merely facilitates later compliance does not automatically instantiate this Tactic.	(Cialdini, 2007; Cialdini et al., 1995)
Low-ball	Beh	Low-ball compliance procedure	**KEEP.** Agreement to the same focal decision precedes a worsening of its terms while continued compliance is sought. This is distinguished from FITD's small-to-larger request sequence and DITF's rejected-large-to-smaller request sequence.	(Cialdini et al., 1978; Burger and Petty, 1981; Burger and Caputo, 2015)
Unity (shared identity)	Bel; Att; Id; Beh	Unity principle, shared social identity, and identity-based social influence	**KEEP; HISTORICAL FORMS ABSORBED.** Shared ingroup membership or collective "we-ness" itself functions as the warrant for a claim, evaluation, or conduct. Patriotic or flag-waving appeals are represented here when national or collective identity performs that warranting function; symbols or group language alone are insufficient.	(Cialdini, 2016; Spears, 2021)
Scarcity appeal	Att; A/C; Beh	Scarcity effects on valuation and action	**KEEP.** Limited quantity, time, supply, access, or continuing availability itself functions as a reason to value the focal opportunity or act before it is lost. Scarcity is distinguished from adverse-consequence coercion and from prospective rewards.	(Lynn, 1991; Worchel et al., 1975; Barton et al., 2022)
Altercasting	Id; Beh	Altercasting and identity/role-based influence	**KEEP.** The recipient is assigned, projected into, or invited to assume a role, category, or identity position carrying an associated behavioural expectation. Shared collective membership alone is handled by *Unity*; positive self-evaluation without role assignment is handled by the applicable ego or pride operation.	(Weinstein and Deutschberger, 1963; Pratkanis, 2000; Miller et al., 1975; Turner et al., 2010)

Target abbreviations: Bel = Belief; Att = Attitude; Id = Identity; A/C = Agency/Choice; Beh = Behavior. No Persuasion Tactic in the final specification inherits Percep-

tion/Attention.

Persuasion decision ledger. Table S34 records transformations for which the final category cannot be understood as simple retention of a literature label. These decisions are included because they materially affect reproducibility of the inventory and its boundaries.

Supplementary Table S34: Principal Persuasion decision ledger.

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Broad authority influence	SPLIT	*Appeal to authority*; *Authority-based directive*	Authority cues perform at least two independently recoverable functions. Expertise, credentials, or recognised authority may operate as an epistemic warrant for accepting a proposition, whereas represented institutional, hierarchical, legal, organisational, or professional standing may itself warrant compliance with conduct. The final ontology therefore distinguishes epistemic authority (Belief) from conduct-directing authority (Behavior). Authority alone does not imply coercion (Cialdini, 2007; Raven, 1993; Tyler, 2006; Milgram, 1963).
Plain-folks / common-person presentation	MERGE	*Ingratiation*, or *Unity (shared identity)* where independently satisfied	The historical "plain-folks" form does not require a separate influence operation. Where ordinariness, relatability, similarity, informality, or common-person presentation cultivates favourable evaluation of the source, it is represented as *Ingratiation*. Where shared membership in an identifiable collective instead functions as the warrant, *Unity* applies. Ordinary clothing, informal language, familiar settings, or humble background alone are insufficient (Institute for Propaganda Analysis, 1939; Jones, 1964; Gordon, 1996; Cialdini, 2016).
Fear / threat-based persuasion	SPLIT: functional	*Appeal to fear*; *Threat framing*; applicable Coercion Tactic	A dangerous or harmful representation may perform different operations. *Appeal to fear* requires threatened harm to warrant an identifiable protective response. *Threat framing* concerns danger or threat as an interpretive framing operation without requiring such a response. Coercion requires prospective harm to be communicatively appropriated as contingent leverage over conduct. The categories may co-occur only where those functions are independently recoverable (Rogers, 1975; Witte and Allen, 2000; Tannenbaum et al., 2015).

Supplementary Table S34: Principal Persuasion decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Content-based inoculation / advance refutation	BOUNDARY: allow overlap	*Two-sided with refutation*; *Preemptive inoculation / prebunking*	Presenting and substantively rebutting an opposing proposition constitutes *Two-sided with refutation*. Preparing the recipient to resist an anticipated later influence attempt constitutes *Preemptive inoculation / prebunking*. Conventional inoculation messages may instantiate both; technique-based prebunking may instantiate the latter without a specific opposing proposition (Banas and Rains, 2010; Compton et al., 2021; Lewandowsky and Linden, 2021).
Manufactured uncertainty / doubt	KEEP: canonicalised	*Doubt-seeding* / Persuasion	Literature on manufactured uncertainty describes the strategic use of uncertainty, disagreement, evidential limitations, or source-reliability concerns to weaken confidence in an epistemic target. Under the ontology's constitutive-operation criterion, this is represented as a persuasive epistemic warrant: the recoverable operation directly bears on acceptance or confidence in a claim, source, evidence base, consensus, or inference. Accordingly, *Doubt-seeding* is represented in the final specification as Persuasion with fixed TargetConstruct *Belief* (Michaels, 2006; Stocking and Holstein, 2009; Biddle and Leuschner, 2015).
Repetition as campaign scheduling / repeated exposure	RESCOPE	*Repetition* / Persuasion where message-internally recoverable	Campaign-level scheduling, repeated distribution, and amplification remain outside the message-internal ontology. Repetition is nevertheless retained as a Tactic where the same or substantially equivalent proposition recurs within the available message or recoverable message sequence and that recurrence itself performs reinforcement. Thus the final inclusion criterion concerns recoverable propositional repetition rather than unobserved campaign frequency (Hasher et al., 1977; Fazio et al., 2015; Ye et al., 2026).

Supplementary Table S34: Principal Persuasion decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Loaded / complex question	MERGE: as form	*Presuppositional embedding*	A question is not a separate Tactic merely because it is rhetorically "loaded." Where its formulation introduces an unestablished determinate proposition as taken for granted, it is an interrogative realisation of *Presuppositional embedding*. Other interrogative forms receive no special label unless another canonical operation is independently satisfied (Thoma et al., 2023; Macagno, 2024).
Guilt by association / honor by association	MERGE: directions	*Associative transfer / guilt-or-honor by association*	Positive and negative transfer through association are treated as directional realisations of one operation. The defining relation is that an association functions as a warrant for transferring commitments, properties, credibility, prestige, stigma, approval, blame, or evaluative standing to a focal target (Mohammed, 2023; Moran et al., 2023).
Patriotic / flag-waving appeal	MERGE: as form	*Unity (shared identity)*	National symbols, patriotic references, or collective language do not constitute a separate Tactic. They instantiate *Unity* where salient national, political, or other ingroup membership itself functions as the influence warrant. Mere display of a flag or reference to a nation is insufficient (Cialdini, 2016; Spears, 2021).

J.1.2 Framing

Supplementary Table S35: Provenance of canonical Framing Tactics.

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Loaded language	Att	Evaluative language, dysphemism/euphemism, and propaganda loaded-language categories	**KEEP + MERGE FORMS.** Evaluatively or affectively charged lexical choice is retained where connotation itself positions a referent favourably or unfavourably without requiring a new substantive proposition. Historical "glittering generalities" and generic positive-value wording are represented here when their operative feature is positive evaluative connotation; more specific warrants such as moral principle, tradition, authority, novelty, or shared identity retain their corresponding canonical Tactics.	(Da San Martino et al., 2020; Allan and Burridge, 1991; Capraro and Vanzo, 2019)
Name-calling / labeling	Att	Propaganda name-calling, evaluative labeling, and social categorisation	**KEEP; BOUNDARY REFINED.** A compact epithet, evaluative label, or categorical characterisation frames who or what an identifiable target is. The condensed characterisation distinguishes this Tactic from more general evaluatively loaded wording. A descriptive social category alone is insufficient.	(Da San Martino et al., 2019; Jowett and O'Donnell, 2018; Carnaghi et al., 2008; Bianchi et al., 2019)
Ridicule / mockery	Att	Political ridicule, mockery, derisive humour, parody, and target-directed comic devaluation	**KEEP; EXPRESSIVE FORMS CONSOLIDATED.** Mockery, parody, caricature, sarcastic re-presentation, ironic quotation, comic incongruity, or comparable forms instantiate the Tactic only where they render an identifiable target laughable, foolish, absurd, contemptible, or otherwise diminished. Satire, sarcasm, parody, humour, or irony are not separate Tactics merely by virtue of form.	(D'Errico and Poggi, 2016; Wang and Silva, 2018; Zappavigna, 2022; Verhulsdonk et al., 2022)
False dilemma	Bel; A/C	False-dilemma / black-and-white reasoning and restricted-alternative representation	**KEEP.** A restricted set of alternatives is communicatively represented as exhaustive, foreclosing additional, intermediate, combined, or neutral possibilities that remain recoverable from the message. The presence of two alternatives alone is insufficient.	(Macagno, 2024; Da San Martino et al., 2020)

Supplementary Table S35: Provenance of canonical Framing Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Causal oversimplification	Bel	Causal oversimplification and monocausal explanatory framing	**KEEP.** A complex outcome, condition, or problem is represented through one factor treated as sufficient, overriding, or exhaustive of the relevant causal explanation. Ordinary causal attribution does not instantiate the Tactic unless causal restriction or exhaustivity is itself recoverable from the communication.	(Da San Martino et al., 2020; Dimitrov et al., 2021)
Slippery-slope / consequential escalation	Bel; Att	Slippery-slope arguments, consequential progression, precedent, and cascade reasoning	**KEEP; DIRECTION GENERALIZED.** An initial action, event, policy, or condition is construed through a projected downstream progression toward a materially more consequential endpoint. Negative and structurally analogous positive escalatory trajectories are treated as directional realisations of the same operation. A single direct consequence is insufficient.	(Walton, 1992; Walton, 2015; Haigh et al., 2016; Piskorski et al., 2023a)
Scapegoating	Bel; Att	Scapegoating, displaced blame, and concentrated responsibility	**KEEP; SEPARATED FROM MONOCAUSALITY.** Blame or responsibility for a broader negative condition or set of harms is concentrated on an identifiable person or group. This differs from *Causal oversimplification*, whose defining operation is a restricted causal model and need not target a person or group. Both are coded where causal exhaustivity and concentrated blame are independently recoverable.	(Rothschild et al., 2012; Glick, 2002)
Moral framing	Bel; Att	Moral framing and moral-value reframing	**KEEP; GENERIC VALUE APPEALS ABSORBED WHERE APPLICABLE.** An issue, actor, policy, outcome, or action is construed or evaluated through an identifiable moral value, principle, virtue, vice, right, obligation, or prohibition. Generic positive or negative wording and isolated morally connoted terms are insufficient unless a moral criterion performs the interpretive or evaluative operation.	(Feinberg and Willer, 2015; Kodapanakkal et al., 2022; Mobayed and Sanders, 2022)

Supplementary Table S35: Provenance of canonical Framing Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Metaphor / analogy	Bel	Structure mapping, analogy, metaphorical framing, and metaphor-induced reasoning	**KEEP; FUNCTIONAL THRESHOLD.** Relational structure from a source or comparison domain must be mapped onto a target so that the imported structure organises interpretation or inference. Figurative language, idiom, imagery, or comparison alone does not instantiate the Tactic where the cross-domain mapping performs no substantive framing work.	(Gentner, 1983; Bowdle and Gentner, 2005; Thibodeau and Boroditsky, 2011)
Dehumanizing representation	Bel; Att	Animalistic, mechanistic, objectifying, and other forms of dehumanization	**KEEP; REPRESENTATIONAL OPERATION ISOLATED.** An identifiable human target is represented as less than fully human, deficient in human qualities, or through an animalistic, object-like, mechanistic, contaminating, parasitic, biological, or comparable materially nonhuman category. Hostility, insult, metaphor, threat, or group reference alone is insufficient. Polarization is annotated independently rather than inherited from dehumanization.	(Haslam, 2006; Kteily et al., 2015; Enock and Over, 2023)
Collective nostalgic framing	Bel; Att; Id	Collective nostalgia, shared past, group continuity, and restorative temporal framing	**KEEP; SEPARATED FROM TRADITION AND UNITY.** A positively valued shared ingroup past functions as a temporal frame for interpreting or evaluating the present or future through continuity, decline, loss, or restoration. Historical persistence used as a warrant is instead *Appeal to tradition*; shared identity used directly as an influence warrant is *Unity (shared identity)*.	(Boym, 2001; Smeekes et al., 2021; Schreurs, 2021; Wildschut et al., 2014)
Inevitability framing	Bel; A/C	Technological or institutional inevitability discourse and "there-is-no-alternative" framing	**KEEP.** A contingent future, development, trajectory, institutional change, policy course, or outcome is represented as unavoidable, predetermined, necessary, unstoppable, or effectively without viable alternatives. Prediction, confidence, advocacy, or novelty alone is insufficient.	(Leonardi, 2008; Séville, 2017)

Supplementary Table S35: Provenance of canonical Framing Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Threat framing	Bel	Threat construction and danger/risk framing	**KEEP; FUNCTIONALLY SEPARATED FROM FEAR APPEALS AND COERCION.** A person, group, event, condition, development, or other phenomenon is represented as standing in a relation of danger, risk, hazard, or potential harm to an identifiable valued target. The Tactic does not require a protective response or a conduct–sanction contingency.	(Eriksson, 2020)
Gain/loss goal framing	Beh	Goal framing and gain/loss persuasive framing of behavioural consequences	**KEEP AS DISTINCT FRAMING TYPE.** An identifiable focal behaviour is promoted through desirable consequences of performing it or undesirable consequences of failing to perform it, while the behavioural goal itself remains constant. This differs from attribute framing and from risky-choice framing because the represented consequences are organised around performance of the focal behaviour.	(Levin et al., 1998; Rothman and Salovey, 1997; Meyerowitz and Chaiken, 1987; O'Keefe and Jensen, 2007)
Risky-choice framing	Beh	Prospect theory and risky-choice framing	**KEEP AS DISTINCT FRAMING TYPE.** Choice alternatives involving uncertain or probabilistic outcomes are represented as gains or losses relative to a reference point, thereby structuring selection among the available alternatives. Neither risk language nor the mere presence of two options is sufficient.	(Tversky and Kahneman, 1981; Levin et al., 1998; Kühberger, 1998)
Attribute framing	P/A; Att	Attribute framing and valenced complementary descriptions	**KEEP AS DISTINCT FRAMING TYPE.** An equivalent underlying attribute is presented through one positively or negatively valenced pole of a directly recoverable complementary description, making that representation relatively focal and evaluatively marked. Positive or negative wording without the complementary-attribute relation is insufficient.	(Levin and Gaeth, 1988; Levin et al., 1998)

Target abbreviations: P/A = Perception/Attention; Bel = Belief; Att = Attitude; Id =

Identity; A/C = Agency/Choice; Beh = Behavior.

Framing decision ledger. Table S36 records transformations in which a historical technique label, broader framing construct, or neighbouring candidate required decomposition, consolidation, or functional separation before entering the canonical Framing inventory.

Supplementary Table S36: Principal Framing inclusion, merge, split, and boundary decisions.

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Glittering generalities / generic positive-value language	MERGE: decomposed across existing categories	*Loaded language*, or applicable specific warrant	A generic positive-value category does not identify a unique influence operation. Where positively evaluative lexical connotation itself supplies the frame, the communication is represented as *Loaded language*. Where an identifiable moral principle, tradition, shared identity, novelty, authority relation, or another specific warrant performs the influence operation, the applicable canonical Tactic is used instead. This preserves the historical phenomenon without introducing a generic residual "appeal-to-values" category (Institute for Propaganda Analysis, 1939; Da San Martino et al., 2020; Capraro and Vanzo, 2019).
Loaded language versus name-calling	BOUNDARY: distinguish	*Loaded language*; *Name-calling / labeling*	Both may involve evaluative language, but their minimum operations differ. *Loaded language* depends on marked connotative lexical choice applied to a referent; *Name-calling / labeling* requires a condensed evaluative label, epithet, or categorical characterisation of what or who the target is. Co-coding is permitted only where both operations are independently recoverable (Da San Martino et al., 2020; Da San Martino et al., 2019; Carnaghi et al., 2008).
Satire / sarcasm / parody / caricature	MERGE: as form	*Ridicule / mockery*, where target-directed devaluation is present	Comic or ironic forms do not constitute separate influence Tactics merely because of their expressive form. They instantiate *Ridicule / mockery* where the communication renders an identifiable target laughable, foolish, absurd, contemptible, or otherwise diminished. Humor used instead to displace substantive engagement is represented by *Humor as deflection* in the Distraction family (D'Errico and Poggi, 2016; Zappavigna, 2022; Verhulsdonk et al., 2022).

Supplementary Table S36: Principal Framing decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Black-and-white fallacy	MERGE: canonicalised	*False dilemma*	The historical black-and-white formulation is treated as one realisation of the broader restricted-alternative operation. Annotation therefore requires that the message represent a restricted set of alternatives as exhaustive; binary wording alone is insufficient (Da San Martino et al., 2020; Macagno, 2024).
Causal oversimplification plus scapegoating	SPLIT: allow overlap	*Causal oversimplification*; *Scapegoating*	Restricted causal explanation and concentrated blame are related but not identical operations. *Causal oversimplification* requires one factor to be represented as sufficient, overriding, or exhaustive of the relevant causal explanation. *Scapegoating* requires concentrated blame or responsibility for a broader negative condition to be directed at an identifiable person or group. They co-occur where both relations are independently recoverable (Da San Martino et al., 2020; Dimitrov et al., 2021; Rothschild et al., 2012).
Broad appeal to values	MERGE: by criterion	*Moral framing*, or another applicable specific Tactic	A broad "values" label is underdetermined because values can perform different influence operations. An identifiable moral value, right, obligation, virtue, vice, justice criterion, or prohibition used to construe or evaluate a target is represented by *Moral framing*. Tradition, authority, shared identity, novelty, and other independently specified warrants retain their respective Tactics (Feinberg and Willer, 2015; Kodapanakkal et al., 2022).
Tradition / nostalgia	SPLIT: across operations	*Appeal to tradition* / Persuasion; *Collective nostalgic framing* / Framing	Longstandingness or customary status used as a reason that a practice, institution, belief, or arrangement should be accepted or retained is a persuasive warrant and is represented as *Appeal to tradition*. A positively valued shared ingroup past used as a temporal reference frame for interpreting the present or future is instead *Collective nostalgic framing*. The categories may co-occur where both operations are independently present (Michaud, 2018; Smeekes et al., 2021; Wildschut et al., 2014).

Supplementary Table S36: Principal Framing decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Flag-waving / patriotic appeal	MERGE: decomposed across existing operations	*Unity (shared identity)*; *Collective nostalgic framing*; other independently satisfied operations	Patriotic symbols or national references do not define one unique message-internal influence operation. Shared national or political membership used directly as an influence warrant is represented as *Unity (shared identity)*. Invocation of a positively valued shared national past as a temporal frame may instead instantiate *Collective nostalgic framing*. Symbols or patriotic vocabulary alone are insufficient for either label (Cialdini, 2016; Spears, 2021; Smeekes et al., 2021).
Fear-mongering / broad threat appeal	SPLIT: functional	*Threat framing*; *Appeal to fear*; applicable Coercion Tactic	Threat-related communication is decomposed by communicative relation. *Threat framing* represents a phenomenon as dangerous or harmful; *Appeal to fear* uses threatened harm as a reason for an identifiable protective response; Coercion uses prospective adverse consequences as contingent leverage over conduct. A single danger premise does not mechanically instantiate all three (Eriksson, 2020; Rogers, 1975; Witte and Allen, 2000).
Generic gain/loss framing	SPLIT	*Attribute framing*; *Risky-choice framing*; *Gain/loss goal framing*	The framing literature distinguishes different structural operations that would be conflated by a single gain/loss category. *Attribute framing* selects a valenced pole of an equivalent complementary attribute; *Risky-choice framing* structures choice among alternatives involving uncertain outcomes relative to a gain/loss reference point; and *Gain/loss goal framing* organises consequences around performance or nonperformance of the same focal behaviour. The final ontology preserves these distinct message-internal structures (Levin and Gaeth, 1988; Levin et al., 1998; Tversky and Kahneman, 1981; Rothman and Salovey, 1997).

Supplementary Table S36: Principal Framing decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Dehumanizing epithet / metaphor / threat	BOUNDARY: allow overlap	*Dehumanizing representation*; *Name-calling / labeling*; *Metaphor / analogy*; *Threat framing*	Nonhuman representation is not inferred from every insult, metaphor, animal term, disease term, or hostile characterisation. The communication must materially construe an identifiable human target as less than fully human or through a nonhuman category. A dehumanising expression may additionally instantiate labeling, metaphorical mapping, or threat framing where those operations are independently recoverable; none is inherited automatically from another (Haslam, 2006; Kteily et al., 2015; Enock and Over, 2023).
Thought-terminating clichés	MERGE: decomposed across existing categories	Applicable *Loaded language*; *Obfuscation*; Narrative symbolic condensation; or other independently satisfied operation	A formulaic or clichéd expression does not identify a sufficiently unique message-internal influence relation to warrant a separate Tactic. Its observable operation is represented through the canonical dimension that it actually instantiates: evaluative lexical framing, obstruction of substantive understanding, narrative symbolic condensation, or another independently recoverable operation. The phrase's familiarity or formulaicity alone is not sufficient.

J.1.3 Manipulation

Supplementary Table S37: Provenance of canonical Manipulation Tactics.

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Guilt trip	Bel; Att; Beh	Guilt, responsibility, interpersonal obligation, and guilt-based compliance	**KEEP; LEVERAGE RELATION SPECIFIED.** Responsibility, obligation, indebtedness, implication, inequity, transgression, or failure attributed to the recipient or a relevant ingroup must function as pressure on conduct. Guilt is not inferred merely from emotional language or discussion of wrongdoing. The message-internal responsibility or obligation relation is the licensing evidence.	(Baumeister et al., 1994; Tangney et al., 2007; Boster et al., 2016)

Supplementary Table S37: Provenance of canonical Manipulation Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Induced hypocrisy / self-inconsistency pressure	P/A; Beh	Induced hypocrisy, self-inconsistency, cognitive dissonance, and behavioural alignment	**KEEP; OBSERVABLE DISCREPANCY IS THE TACTIC.** The recipient's endorsed standard or position and their own inconsistent conduct are placed into direct comparison, and that discrepancy is used as pressure for subsequent alignment. Cognitive dissonance, discomfort, guilt, or shame are treated as possible mechanisms rather than annotation requirements. Both sides of the discrepancy must be recoverable.	(Aronson et al., 1991; Stone et al., 1994; Stone et al., 1997; Priolo et al., 2019)
Appeal to pity / sympathy	Bel; Att; Beh	Argumentum ad misericordiam, sympathy, compassion, leniency, and pity-mediated judgement	**KEEP; TARGET SCOPE BROADENED BEYOND ACTION.** Another person's or group's suffering, vulnerability, hardship, misfortune, or need functions as a warrant for accepting a judgement, adopting a more sympathetic or lenient evaluation, or taking a favourable action. The emotion itself is not a TargetConstruct; the targeted judgement, evaluation, or conduct is.	(Walton, 1997; Wood et al., 2014; Condon and DeSteno, 2011)
Love-bombing	Bel; Att; Beh	Love-bombing, accelerated intimacy, intense affiliative attention, and relational influence	**KEEP; INTENSITY AND ACCELERATION REQUIRED.** Unusually intense, dense, accelerated, or repeated affection, praise, attention, gifts, intimacy cues, or future-oriented commitment are used to establish or accelerate trust, attachment, or relational commitment that can support subsequent influence. Ordinary affection, compliments, gifts, or rapport building are insufficient.	(Strutzenberg et al., 2017; Arabi, 2023; Schokkenbroek and Snaphaan, 2025)
Shaming / derogation	Id; Beh	Shame, social evaluation, conformity, status loss, and shame-based behavioural pressure	**KEEP; BEHAVIORAL PRESSURE REQUIRED.** Negative evaluation of the recipient's character, identity, worth, status, social image, or anticipated noncompliance must function as leverage around conduct. Insult, ridicule, moral condemnation, or pejorative labeling without conduct-directed pressure is not sufficient.	(Scheff, 1988; Tangney et al., 2007; Grasmick et al., 1991)

Supplementary Table S37: Provenance of canonical Manipulation Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Pride / ego appeals	Bel; Att; Id; Beh	Self-image, ego involvement, identity-consistent responding, and positive self-concept leverage	**KEEP.** Acceptance of a proposition, evaluation, or course of conduct is linked to a positively valued self-concept, competence, trait, status, or social image so that the promoted response is represented as diagnostic of or consistent with that valued identity. Praise alone is insufficient unless self- or status-relevance supplies the influence warrant.	(Walton et al., 2008; Howard et al., 1995; Wheeler et al., 2005; Perkins and Forehand, 2012)
Dangling contingent rewards	Beh	Compliance through incentives and prospective conduct–benefit contingencies	**KEEP; PROSPECTIVE CONTINGENCY REQUIRED.** A reward, benefit, prize, privilege, discount, access condition, or comparable positive consequence is offered conditional on identifiable recipient conduct and functions as instrumental leverage for that conduct. Desirable consequences that arise from the behaviour itself are distinguished from source-offered contingent rewards.	(Kelman, 1958; Mathur et al., 2019; Luguri and Strahilevitz, 2021)
Affective associative cueing	P/A	Affective attention, motivational salience, evaluative priming, taboo effects, and socially or identity-relevant cueing	**KEEP; ATTENTIONAL OPERATION IS CONSTITUTIVE.** An affectively, morally, socially, or identity-relevant cue or association is activated or foregrounded so that its pre-existing motivational significance makes a focal representation disproportionately salient. No explicit evaluation, proposition, or behavioural request is required. The Tactic is not inferred from merely emotional or controversial subject matter.	(MacKay et al., 2004; Gable and Harmon-Jones, 2010; Brosch and Van Bavel, 2012; Kaiser et al., 2006; Herring et al., 2013; Brady et al., 2020)
Manipulative choice architecture	A/C; Beh	Dark patterns, defaults, obstruction, asymmetric choice architecture, and interface-based steering	**KEEP; STRUCTURAL ASYMMETRY REQUIRED.** Presentation, availability, sequence, effort, default status, access, or reversibility is structured so that one response is materially easier or more actionable while alternatives are burdened. The defining operation is the architecture of choice rather than the substantive description or merits of the alternatives.	(Mathur et al., 2019; Luguri and Strahilevitz, 2021)

Supplementary Table S37: Provenance of canonical Manipulation Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Gaslighting / epistemic destabilization	Bel; Id; A/C	Gaslighting, epistemic self-trust, memory destabilization, and epistemic agency	**KEEP; TRUTH-AGNOSTIC OPERATIONALIZATION.** An identifiable perception, memory, interpretation, or judgement is challenged while the recipient's own capacity to perceive, remember, reason, or judge reliably is itself discredited. Ordinary contradiction, denial, correction, or competing recollection is insufficient. Annotation does not require determining whose account is factually correct or whether deception was intended.	(Abramson, 2014; Spear, 2023; Darke et al., 2025; Proust, 2008)

Target abbreviations: P/A = Perception/Attention; Bel = Belief; Att = Attitude; Id = Identity; A/C = Agency/Choice; Beh = Behavior.

Manipulation decision ledger. Table S38 records the principal distinctions required to derive the final Manipulation inventory from neighbouring affective, relational, incentive-based, associative, and choice-structuring influence constructs.

Supplementary Table S38: Principal Manipulation inclusion, distinction, and decomposition decisions.

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Guilt, pity, shame, and hypocrisy as generic emotional pressure	SPLIT: by warrant	*Guilt trip*; *Appeal to pity / sympathy*; *Shaming / derogation*; *Induced hypocrisy / self-inconsistency pressure*	The final inventory does not classify these Tactics by the emotion a recipient might experience. It distinguishes their observable warrant structures. *Guilt trip* uses represented responsibility, obligation, indebtedness, implication, or transgression as pressure; *Appeal to pity / sympathy* uses another's suffering or vulnerability as a basis for judgement, leniency, or response; *Shaming / derogation* uses negative self-, identity-, status-, or social-image evaluation as pressure on conduct; and *Induced hypocrisy* foregrounds an endorsed-standard–own-conduct discrepancy as pressure for alignment. Possible guilt, sympathy, shame, embarrassment, or dissonance are mechanisms or effects rather than sufficient coding evidence (Tangney et al., 2007; Boster et al., 2016; Walton, 1997; Stone et al., 1994; Scheff, 1988).
Cognitive dissonance	BOUNDARY: mechanism, not Tactic	*Induced hypocrisy / self-inconsistency pressure* where its observable conditions are met	Cognitive dissonance is an internal psychological state or explanatory mechanism and is not directly inferred from message content. The annotatable operation is the recoverable juxtaposition of a recipient's endorsed standard with their inconsistent conduct and use of that discrepancy as pressure for subsequent alignment (Aronson et al., 1991; Stone et al., 1994; Stone et al., 1997).
Accusation of hypocrisy	BOUNDARY: distinguish by target of pressure	*Induced hypocrisy / self-inconsistency pressure*, or *Ad hominem / personal-credibility attack*	Calling another person inconsistent or hypocritical does not itself instantiate induced hypocrisy. The Manipulation Tactic requires the recipient's own endorsed standard and inconsistent conduct to be recoverable and used as pressure for behavioural alignment. Where another actor's inconsistency is instead used to discount that actor's claim, testimony, argument, criticism, or source standing, the applicable operation is *Ad hominem / personal-credibility attack* (Stone et al., 1994; Priolo et al., 2019).

Supplementary Table S38: Principal Manipulation decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Love-bombing versus ingratiation	BOUNDARY: distinguish relational intensity and dependency from ordinary liking appeals	*Love-bombing*; *Ingratiation*	Both may involve praise, agreement, favourable attention, similarity, or affiliative behaviour. *Ingratiation* requires cultivation of favourable evaluation or liking of the source. *Love-bombing* additionally requires unusually intense, accelerated, dense, or repeated affiliative behaviour functioning to establish or accelerate intimacy, trust, attachment, or relational commitment. Ordinary flattery or rapport building therefore does not automatically instantiate love-bombing (Jones, 1964; Gordon, 1996; Strutzenberg et al., 2017; Schokkenbroek and Snaphaan, 2025).
Gift or benefit as influence	SPLIT: by temporal / contingency relation	*Love-bombing*; *Reciprocity*; *Dangling contingent rewards*	A gift does not determine the Tactic by itself. It may participate in *Love-bombing* where unusually intense affiliative behaviour accelerates relational attachment; in *Reciprocity* where a benefit already provided functions as a warrant for a return act; or in *Dangling contingent rewards* where a prospective benefit is made conditional on recipient conduct. Co-code only where the corresponding relations are independently recoverable (Gouldner, 1960; Strutzenberg et al., 2017; Kelman, 1958).
Positive versus negative self-image leverage	SPLIT: by valence and operation	*Pride / ego appeals*; *Shaming / derogation*	Positive self-concept, competence, status, or social image used to make a promoted judgement, evaluation, or behaviour diagnostic of a valued identity is represented by *Pride / ego appeals*. Negative character, identity, worth, status, or social-image evaluation used as pressure on conduct is represented by *Shaming / derogation*. The distinction is not simply positive versus negative wording; the self- or identity-relevant evaluation must function as influence leverage (Howard et al., 1995; Wheeler et al., 2005; Scheff, 1988).

Supplementary Table S38: Principal Manipulation decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Praise / positive identity / role assignment	BOUNDARY: distinguish by warrant	*Ingratiation*; *Pride / ego appeals*; *Altercasting*	Praise may perform different operations. Where it primarily cultivates liking of the source, it supports *Ingratiation*. Where a positively valued self-concept, competence, status, or social image is tied to the response being promoted, it supports *Pride / ego appeals*. Where the recipient is assigned or invited into a social or relational role carrying a role-consistent behavioural expectation, it supports *Altercasting*. Surface praise alone does not license all three.
Prospective benefit versus behavioural consequence	BOUNDARY: distinguish	*Dangling contingent rewards*; *Gain/loss goal framing*	*Dangling contingent rewards* requires an offered benefit whose receipt or availability is contingent on identifiable conduct and functions as instrumental leverage. *Gain/loss goal framing* instead represents desirable or undesirable consequences associated with performing or failing to perform the focal behaviour. The distinction is therefore between an exchange-like conduct–reward contingency and valenced representation of the behaviour's consequences (Kelman, 1958; Levin et al., 1998; Rothman and Salovey, 1997).
Prospective reward versus prior benefit	BOUNDARY: distinguish	*Dangling contingent rewards*; *Reciprocity*	A prospective conduct–benefit contingency is represented as *Dangling contingent rewards*. A benefit, favour, gift, service, or concession already provided and invoked as a reason for a corresponding return act is represented as *Reciprocity*. The temporal and relational structure, rather than the mere presence of a benefit, determines the Tactic (Gouldner, 1960; Regan, 1971; Kelman, 1958).
Affective association versus associative transfer	BOUNDARY: distinguish by function	*Affective associative cueing*; *Associative transfer / guilt-or-honor by association*	Both rely on associations, but the constitutive operations differ. *Affective associative cueing* uses the pre-existing affective, motivational, moral, social, or identity significance of an association to alter relative salience or attention. *Associative transfer* uses an association as a warrant for attributing represented qualities, commitments, credibility, prestige, stigma, blame, approval, or evaluative standing from one entity to another. Neither automatically entails the other (Herring et al., 2013; Brosch and Van Bavel, 2012; Mohammed, 2023).

Supplementary Table S38: Principal Manipulation decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Emotionally charged cue versus explicit evaluation or threat	BOUNDARY: distinguish by operation	*Affective associative cueing*; *Loaded language*; *Appeal to fear*; *Threat framing*	An affectively significant association used to change relative salience is represented as *Affective associative cueing*. Explicitly evaluative lexical connotation is represented as *Loaded language*; a danger or harm relation organising interpretation is *Threat framing*; and threatened harm used as a reason for an identifiable protective response is *Appeal to fear*. Emotional or controversial subject matter by itself does not license any of these Tactics (Brosch and Van Bavel, 2012; Capraro and Vanzo, 2019; Eriksson, 2020; Witte and Allen, 2000).
Dark patterns	MERGE: decomposed by observable operation	*Manipulative choice architecture* plus other independently satisfied Tactics	The dark-pattern literature contains heterogeneous interface and influence practices rather than one uniform message-internal operation. *Manipulative choice architecture* represents cases in which asymmetric access, effort, default status, sequencing, obstruction, or reversibility privileges one response. A specific implementation may additionally satisfy Deception, Scarcity, Social proof, Shaming, Loaded language, contingent reward, or another Tactic, but membership in a broad dark-pattern category does not mechanically instantiate them (Mathur et al., 2019; Luguri and Strahilevitz, 2021).
Choice architecture versus framing	BOUNDARY: distinguish by decision environment	*Manipulative choice architecture*; applicable Framing Tactic	Choice architecture is licensed when alternatives are structurally unequal in accessibility, effort, default status, sequence, completion, or reversibility. Framing applies where alternatives remain practically available but their description, representation, consequences, attributes, or interpretive context are selectively structured. Unequal rhetoric or attractiveness alone does not establish manipulative choice architecture (Mathur et al., 2019; Luguri and Strahilevitz, 2021).

Supplementary Table S38: Principal Manipulation decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Gaslighting versus ordinary disagreement	BOUNDARY: distinguish	*Gaslighting / epistemic destabilization*	Contradiction, denial, factual correction, competing recollection, or criticism of a particular conclusion is insufficient. Gaslighting requires that the recipient's own epistemic competence or self-trust—their capacity to perceive, remember, reason, interpret, or judge reliably—be discredited as part of the influence operation (Abramson, 2014; Spear, 2023).
Gaslighting versus external doubt	BOUNDARY: distinguish by epistemic target	*Gaslighting / epistemic destabilization*; *Doubt-seeding*	*Doubt-seeding* directs uncertainty, evidential limitations, source-reliability concerns, disagreement, or similar qualifications toward an external claim, source, evidence base, consensus, conclusion, or inference. *Gaslighting / epistemic destabilization* instead attacks the recipient's confidence in their own perception, memory, reasoning, or judgement. Both may occur in the same message only where both epistemic targets are independently undermined (Spear, 2023; Biddle and Leuschner, 2015).
Gaslighting as deception	BOUNDARY: no inheritance	*Gaslighting / epistemic destabilization*; applicable Deception Tactic only if independently satisfied	The operational definition of gaslighting is truth-agnostic. Annotators do not need to establish that the source knowingly lies, that the competing account is false, or that an abusive relationship exists. A Deception Tactic is co-coded only where its own message-internal form-level conditions are independently satisfied (Abramson, 2014; Spear, 2023).

J.1.4 Coercion

Supplementary Table S39: Provenance of canonical Coercion Tactics.

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Explicit threats	A/C; Beh	Threats, coercive power, deterrence, compellence, and conditional conduct–consequence relations	**KEEP; EXPRESSION FORM SPECIFIED.** An identifiable prospective adverse consequence is directly made contingent on performance, nonperformance, continuation, or cessation of identifiable conduct and is communicatively used as leverage over that conduct. The source need not personally execute the threatened harm; what is required is appropriation of the adverse consequence as coercive leverage.	(Schelling, 1966; Raven, 1993; De Dreu, 1995; Evans and Twyman-Musgrove, 1998; Waxman, 2014; Reichberg and Syse, 2018)
Implied threats / coercive intimidation	A/C; Beh	Implicit threats, indirect speech, coercive signaling, and implied conduct–consequence relations	**KEEP; INDIRECT EXPRESSION FORM SPECIFIED.** The same basic conduct–consequence leverage relation required for coercion is recoverable, but the adverse consequence, contingency, source commitment, connection to the harm, or another component is communicated indirectly through implication, allusion, contextual inference, or communicative action. Ominous language or aggressive tone alone is insufficient.	(Schelling, 1966; Kellermann and Shea, 1996; Haigh et al., 2011; Reichberg and Syse, 2018; Jankovic and Roeben, 2025)
Blackmail / disclosure leverage	A/C; Beh	Blackmail, extortionate threats, disclosure leverage, and information-for-conduct contingencies	**KEEP; SPECIALIZED LEVERAGE TYPE.** Prospective disclosure, publication, reporting, exposure, transmission, or non-disclosure of information functions as contingent leverage over identifiable conduct. Possession, publication, or threatened publication of sensitive information alone is insufficient unless a conduct–disclosure contingency is recoverable. The information need not be reputational in nature.	(Shavell, 1993; Gordon, 1993; Lamond, 1996; Elhauge, 2016)

Supplementary Table S39: Provenance of canonical Coercion Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Coercive ultimatum	A/C; Beh	Ultimata, forceful persuasion, bounded compliance conditions, and sanction-backed demands	**KEEP; SPECIALIZED FINALITY STRUCTURE.** A prospective adverse consequence is used as leverage over identifiable conduct and the communication additionally represents compliance as final, non-negotiable, last-chance, deadline-based, or otherwise bounded. Urgency, emphasis, or a deadline alone does not establish the Tactic without coercive sanction leverage.	(George, 1991; Schelling, 1966; Reichberg and Syse, 2018)
Collective punishment threats	A/C; Beh	Collective sanctions, distributed responsibility, and group-based compliance pressure	**KEEP; SPECIALIZED DISTRIBUTED-LIABILITY STRUCTURE.** A prospective sanction extends beyond the immediately responsible actor to additional members of an associated group, family, organisation, community, population, or collective on account of conduct attributable to one or some members, and that distributed consequence functions as behavioural leverage. Harm to one associated third party alone is insufficient.	(Heckathorn, 1990; Whitmeyer, 2002)

Target abbreviations: A/C = Agency/Choice; Beh = Behavior. All canonical Coercion Tactics inherit the fixed TargetConstruct set {Agency/Choice, Behavior}.

Coercion decision ledger. Table S40 records the principal distinctions used to derive the final Coercion inventory. Coercive Tactics are identified by an observable conduct–consequence leverage relation rather than by frightening content, source hostility, legal status, or inferred recipient fear.

Supplementary Table S40: Principal Coercion inclusion, distinction, and structural decisions.

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Generic threat	SPLIT: by expression form	*Explicit threats*; *Implied threats / coercive intimidation*	Both Tactics require an identifiable conduct–consequence relation in which a prospective adverse consequence functions as leverage. They differ in how that relation is communicated. *Explicit threats* directly state the relevant adverse consequence and contingency; *Implied threats / coercive intimidation* leave one or more components indirect but recoverable through context, implication, allusion, or communicative action. Ordinary pragmatic implications carried by an otherwise explicit threat do not justify coding both forms (Schelling, 1966; Kellermann and Shea, 1996; Haigh et al., 2011; Reichberg and Syse, 2018).
Threat expression versus specialised coercive structure	BOUNDARY: allow co-coding	Threat expression form plus applicable *Blackmail / disclosure leverage*, *Coercive ultimatum*, or *Collective punishment threats*	Explicit/implied labels describe how the coercive relation is expressed. Blackmail, ultimatum, and collective-punishment labels instead identify additional structure within that same relation: informational disclosure as the sanctioning leverage, bounded or final compliance conditions, or distributed sanction liability. A single coercive episode may therefore receive both an expression-form label and a specialised structural label without requiring a second adverse consequence.
Threat versus warning or danger	BOUNDARY: distinguish by appropriation as leverage	Applicable Coercion Tactic; *Appeal to fear*; *Threat framing*	A harmful outcome does not become coercive merely because it may motivate behaviour. Coercion requires the communication to appropriate the prospective adverse consequence as leverage around identifiable target conduct. *Appeal to fear* instead uses danger as a reason for an identifiable protective response without appropriating the danger as a sanction, while *Threat framing* represents a phenomenon as dangerous or harmful without requiring either behavioural relation (Schelling, 1966; Rogers, 1975; Witte and Allen, 2000; Eriksson, 2020).

Supplementary Table S40: Principal Coercion decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Source control of threatened harm	BOUNDARY: require observable selection	✗ Applicable Coercion Tactic	The coercive operation does not require the communicator to claim that they will personally inflict or institutionally control the prospective harm. The consequence may be represented as imposed by an authority, associated actor, group, proxy, institution, or another process whose occurrence the source can facilitate, authorise, permit, or decline to prevent. What must be recoverable is that the communication uses that consequence as leverage over target conduct, rather than merely predicting independent danger (Schelling, 1966; Reichberg and Syse, 2018; Jankovic and Roeben, 2025).
Authority-backed direction versus coercion	BOUNDARY: distinguish by adverse-consequence leverage	✗ *Authority-based directive*; applicable Coercion Tactic	Represented institutional, hierarchical, legal, organisational, or professional authority may function as a warrant for compliance without coercion. Coercion requires an additional prospective adverse consequence communicatively used as leverage over conduct. The existence of authority, a command, obligation, or lawful rule therefore does not by itself license a Coercion Tactic.
Prospective disclosure or exposure	BOUNDARY: distinguish by contingency	✗ *Blackmail / disclosure leverage*; other applicable privacy, expression, or coercion categories	Possessing, discussing, publishing, exposing, or announcing an intention to publish sensitive or consequential information does not by itself constitute blackmail. The communication must make disclosure or non-disclosure contingent on an identifiable response and use that informational consequence as leverage. Reputational harm, privacy loss, embarrassment, or legal exposure may explain why disclosure matters but is not constitutive of the Tactic (Shavell, 1993; Gordon, 1993; Lamond, 1996).
Blackmail as legal category	RESCOPE: independent of legal judgement	*Blackmail / disclosure leverage*	The ontology records an observable disclosure-for-conduct leverage structure rather than determining whether the communication satisfies the criminal or civil-law elements of blackmail, extortion, coercion, or another offence in a particular jurisdiction. Truth of the information, legality of its acquisition, legitimacy of disclosure, and private intention are not annotation requirements (Gordon, 1993; Lamond, 1996; Elhauge, 2016).

Supplementary Table S40: Principal Coercion decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Deadline / "take it or leave it" language	BOUNDARY distinguish by sanction	*Coercive ultimatum*; *Scarcity appeal* where applicable	A deadline or final condition does not itself establish coercion. *Coercive ultimatum* requires a prospective adverse consequence used as leverage plus a final, non-negotiable, last-chance, or bounded compliance condition. *Scarcity appeal* instead concerns declining availability, time, supply, access, or opportunity without an adverse sanction. Thus "accept by Friday or the offer expires" and "sign by Friday or you are fired" instantiate different operations (George, 1991; Schelling, 1966).
Threat against an associated third party	BOUNDARY distinguish from collective liability	Applicable explicit/implied threat; *Collective punishment threats* only where distributed liability is present	A prospective harm directed at a recipient's relative, colleague, or other valued third party may constitute coercive leverage without constituting collective punishment. The collective-punishment Tactic requires a recoverable structure in which persons beyond the immediately responsible actor are prospectively sanctioned *on account of conduct attributable to one or some members* of the relevant group or collective (Heckathorn, 1990; Whitmeyer, 2002).
Group relation in collective punishment	BOUNDARY no inheritance	*Collective punishment threats*	Family, organisational, communal, national, ethnic, religious, or other group relations may determine who bears the threatened sanction or why that sanction has leverage. They do not imply that the message seeks to alter the recipient's self-concept or group identification. Identity is therefore not a fixed TargetConstruct of collective punishment threats (Heckathorn, 1990; Whitmeyer, 2002).
Fear, intimidation, credibility, legality, or proportionality	BOUNDARY not licensing conditions	Applicable Coercion Tactic	Annotation records the recoverable coercive operation rather than its psychological success or normative status. Annotators do not need to establish that the recipient experiences fear, that the threatened consequence is credible or likely, that the communicator privately intends to carry it out, or that the sanction is lawful, justified, proportionate, or morally legitimate (Evans and Twyman-Musgrove, 1998; Reichberg and Syse, 2018).

J.1.5 Deception

The label *Deception* denotes the ontology family to which the following message-level presentation operations are assigned; it does not require annotators to establish that a proposition is objectively false, that the communicator knowingly lied, or that deceptive intent was present. Canonical Tactics in this family are licensed by message-internal relations such as recoverable scalar distortion, material underdetermination, selective evidential presentation, distorted comparison, or differential perceptual salience. External fact-checking and private-intent inference are therefore not prerequisites for annotation.

Supplementary Table S41: Provenance of canonical Deception Tactics.

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Exaggeration / overstatement	Bel; Att	Exaggeration, overstatement, claim inflation, and distortion of evidential or scalar strength	**KEEP; RECOVERABLE BASELINE REQUIRED.** A quantity, frequency, likelihood, certainty, scope, effect, severity, quality, importance, evaluation, or evidential or causal strength is represented as more extreme than a comparison point, source characterisation, or baseline recoverable from the available communication. External facts are not used to determine whether the statement is exaggerated. Intensifiers, superlatives, and hyperbole alone are insufficient.	(Jowett and O'Donnell, 2018; Da San Martino et al., 2019; Sumner et al., 2014; Haber et al., 2018)
Minimization / understatement	Bel; Att	Minimization, understatement, mitigation, and downscaling of represented severity or evidential strength	**KEEP; RECOVERABLE BASELINE REQUIRED.** A represented quantity, severity, harm, risk, cost, responsibility, importance, effect, or evidential or causal strength is downscaled relative to a comparison point or baseline recoverable from the communication. Downtoners, euphemisms, or mitigating language alone are insufficient unless they participate in a recoverable reduction relation.	(Dimitrov et al., 2021; Kassin and McNall, 1991; Luke and Alceste, 2020)

Supplementary Table S41: Provenance of canonical Deception Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Strategic vagueness / equivocation	Bel	Strategic ambiguity, equivocation, vagueness, and noncommittal communication	**KEEP; CONTEXTUAL NEED FOR DETERMINACY REQUIRED.** A materially relevant proposition, quantity, scope, referent, agent, responsibility relation, stance, condition, or commitment is left underdetermined where the available communicative context makes greater determinacy relevant. Vagueness, hedging, passive voice, approximation, or ambiguity is not inherently deceptive and does not alone license the Tactic.	(Eisenberg, 1984; Bavelas et al., 1990; Solt, 2015)
Selective evidence presentation (cherry-picking)	Bel; Att	Selective reporting, spin, selective disclosure, and evidential cherry-picking	**KEEP; RECOVERABLE EVIDENTIAL SET REQUIRED.** A systematically supportive subset is selected from a broader evidence or consideration set that is itself recoverable from the available annotation context, while countervailing members are excluded, omitted, or denied comparable evidential use. One-sided advocacy or failure to mention externally imagined counterevidence is insufficient.	(Chiu et al., 2017; Farina et al., 2026)
False balance / distorted evidential parity	Bel	False balance, distorted consensus representation, and misleading evidential parity	**KEEP; EPISTEMIC PARITY DISTORTION REQUIRED.** Positions, claims, sources, expert judgements, or bodies of evidence whose support, prevalence, consensus, credibility, or evidential standing is materially asymmetric in the recoverable context are represented as more evenly balanced or comparably supported. Merely presenting opposing viewpoints or reporting disagreement is insufficient.	(Boykoff and Boykoff, 2004; Koehler, 2016; Dixon and Clarke, 2013)

Supplementary Table S41: Provenance of canonical Deception Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
False equivalence / distorted comparability	Bel; Att	False or moral equivalence, distorted analogy, and inappropriate parity across differentiated cases	**KEEP; NON-EVIDENTIAL COMPARABILITY DISTORTION REQUIRED.** Two or more cases, actors, actions, policies, events, outcomes, harms, or other comparands are represented as equivalent or comparably significant on an identifiable dimension despite a material difference on that same dimension that is recoverable from the annotation context. External knowledge is not used to judge whether the comparison is fair or appropriate.	(Piskorski et al., 2025; Walton and Hyra, 2018; Shorten, 2011)
Salience-based concealment	P/A; Bel	Dark patterns, hidden information, visual hierarchy, and differential perceptual accessibility of disclosures	**KEEP; INFORMATION MUST REMAIN PRESENT.** Materially qualifying, limiting, costly, adverse, or otherwise decision-relevant information remains within the communication but is made markedly less perceptually prominent or accessible than the focal representation it qualifies. The Tactic therefore concerns differential salience rather than complete omission. Low prominence alone is insufficient unless the suppressed information materially bears on the focal representation.	(Gray et al., 2018; Mathur et al., 2019; Luguri and Strahilevitz, 2021)

Target abbreviations: P/A = Perception/Attention; Bel = Belief; Att = Attitude.

Deception decision ledger. Table S42 records the principal distinctions used to derive the final Deception inventory. The distinctions concern observable relations within the available message or message context and do not require external adjudication of factual truth or private communicative intent.

Supplementary Table S42: Principal Deception inclusion, distinction, and boundary decisions.

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Deception as lying or factual falsity	RESCOPE: form-level operationalisation	Applicable canonical Deception Tactic	The ontology does not annotate a generic category of "lying." Establishing that a proposition is false or that the communicator knowingly stated a falsehood would generally require factual verification or private-intent inference beyond the message-internal annotation rule. Deception-family Tactics are therefore defined through observable presentation relations such as recoverable scalar distortion, underdetermination, selective evidence, distorted parity, or perceptual suppression.
Exaggeration versus minimization	SPLIT: by direction of scalar disortion	*Exaggeration / overstatement*; *Minimization / understatement*	Both require a recoverable comparison point, source characterisation, or evidential baseline. Exaggeration represents the relevant degree or strength as greater or more extreme than that baseline; minimization represents it as less extensive, serious, consequential, important, or strong. The labels therefore preserve opposite message-internal distortion relations rather than inferring objective truth or falsity (Sumner et al., 2014; Luke and Alceste, 2020).
Intensification / attenuation versus loaded language	BOUNDARY: distinguish by recoverable degree relation	*Exaggeration / overstatement*; *Minimization / understatement*; *Loaded language*	Evaluatively or affectively marked wording is represented as *Loaded language* where connotation itself provides the framing operation. Exaggeration or minimization requires an independently recoverable scalar, quantitative, modal, causal, evidential, or evaluative baseline relative to which the represented degree is increased or decreased. Intensifiers, euphemisms, superlatives, diminutives, or downtoners therefore do not mechanically instantiate a Deception Tactic.

Supplementary Table S42: Principal Deception decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Vagueness / ambiguity as lexical property	RESCOPE: contextualised	*Strategic vagueness / equivocation* only where material underdetermination is recoverable	Vague quantifiers, hedges, passive constructions, modal expressions, approximators, ambiguity, or indirect wording can serve ordinary and non-deceptive communicative functions. The canonical Tactic requires the available context to establish a materially relevant issue for which greater determinacy matters while the formulation leaves materially different interpretations, referents, quantities, responsibilities, positions, or commitments unresolved (Eisenberg, 1984; Bavelas et al., 1990; Solt, 2015).
Vagueness versus minimization	BOUNDARY: distinguish by representational relation	*Strategic vagueness / equivocation*; *Minimization / understatement*	Minimization represents an identifiable degree, quantity, severity, responsibility, or evidential strength as lower relative to a recoverable baseline. Strategic vagueness instead leaves the relevant representation materially indeterminate. "Less" and "unclear" are therefore not treated as equivalent operations.
Vagueness versus obfuscation	BOUNDARY: distinguish semantic determinacy from processing difficulty	*Strategic vagueness / equivocation*; *Obfuscation (jargon/complexity)*	Strategic vagueness leaves a materially relevant proposition, quantity, referent, scope, responsibility relation, condition, stance, or commitment underdetermined. Obfuscation leaves content available but makes it difficult to process or comprehend through jargon, complexity, or comparable presentation. Both may co-occur, but neither is inherited from the other.
Vagueness versus evasive agenda shifting	BOUNDARY: distinguish focal matter from focal determinacy	*Strategic vagueness / equivocation*; *Agenda shifting / evasive refocusing*	Strategic vagueness retains the focal matter but leaves a materially relevant element unresolved. Agenda shifting displaces an established question, issue, criterion, comparison, actor, time frame, or other focus with a different agenda. Co-code only where the message independently both changes the focal agenda and leaves a relevant element underdetermined.

Supplementary Table S42: Principal Deception decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Cherry-picking versus one-sided advocacy	BOUNDARY require observable selection	*Selective evidence presentation (cherry-picking)*	Advocacy, presenting reasons for one side, or omitting unspecified counterarguments is insufficient. The annotation context must make a broader evidential or consideration set recoverable and show systematic selection of supportive members while countervailing members are excluded, omitted, or denied comparable evidential use. Annotators do not search externally for missing evidence or infer what evidence the source knew (Chiu et al., 2017; Farina et al., 2026).
Cherry-picking versus salience concealment	BOUNDARY distinguish selection from perceptual suppression	*Selective evidence presentation (cherry-picking)*; *Salience-based concealment*	Cherry-picking concerns which members of a broader recoverable evidential or consideration set are selected for evidential use. Salience concealment concerns materially qualifying information that remains present but is made substantially less perceptually accessible than the focal representation. Selection does not require perceptual suppression, and perceptual suppression does not require evidential subset selection.
Cherry-picking versus false balance	BOUNDARY distinguish subset selection from parity distortion	*Selective evidence presentation (cherry-picking)*; *False balance / distorted evidential parity*	Cherry-picking uses a systematically selected subset from a broader recoverable evidential set. False balance instead takes a recoverably asymmetric distribution of evidence, expert endorsement, prevalence, credibility, or consensus and communicates greater parity than that baseline supports. Both may co-occur where subset selection and parity distortion are independently present.
Balanced presentation / multiple viewpoints	BOUNDARY no inheritance	*False balance / distorted evidential parity* only where asymmetry and subsequent parity are both recoverable	Presenting two views, quoting one source from each side, reporting disagreement, or using symmetric formatting does not establish false balance. The available context must independently establish a material asymmetry in support, prevalence, consensus, credibility, or evidential standing and a representation that makes the positions appear more nearly balanced (Boykoff and Boykoff, 2004; Koehler, 2016; Dixon and Clarke, 2013).

Supplementary Table S42: Principal Deception decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
False balance versus false equivalence	SPLIT: by comparison dimension	*False balance / distorted evidential parity*; *False equivalence / distorted comparability*	False balance is restricted to epistemic distributions: evidence, support, expert endorsement, prevalence, consensus, credibility, or comparable standing. False equivalence concerns parity or equivalence across cases, actors, actions, policies, harms, responsibility, seriousness, merit, legitimacy, or other non-evidential dimensions despite a recoverable material difference. The shared language of "balance" or "equivalence" does not collapse these distinct relations.
False balance versus doubt-seeding	BOUNDARY: distinguish distortion from persuasive use of uncertainty	*False balance / distorted evidential parity*; *Doubt-seeding*	A false-balance representation distorts the relative distribution of represented support or evidential standing. *Doubt-seeding* instead uses uncertainty, disagreement, evidential limitations, possible error, source-reliability concerns, or procedural limitations as a reason to reduce confidence in or suspend judgement toward an epistemic target. False balance does not automatically instantiate doubt-seeding merely because uncertainty could result downstream.
False equivalence versus ordinary analogy	BOUNDARY: distinguish mapping from distorted parity	*False equivalence / distorted comparability*; *Metaphor / analogy*	An analogy maps relational structure from one domain or case onto another in order to organise interpretation or inference. It does not become false equivalence merely because the compared items differ. False equivalence requires the available context itself to establish a material difference on the relevant comparison dimension that is nevertheless overridden or collapsed by the communicated parity claim (Walton and Hyra, 2018).
False equivalence versus whataboutism	BOUNDARY: distinguish comparison from deflection	*False equivalence / distorted comparability*; *Whataboutism*	A whataboutist comparison redirects scrutiny from an established criticism, question, or accusation through a countercase. It need not claim that the two cases are equivalent. False equivalence instead requires recoverable material difference on the comparison dimension plus communicated equivalence or comparable standing. Both may be coded where the countercomparison independently performs both operations.

Supplementary Table S42: Principal Deception decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Omitted information versus salience concealment	BOUNDARY distinguish absence from differential accessibility	*Salience-based concealment* only where the qualification remains present	Salience concealment requires materially relevant qualifying information to remain within the communication while presentation features make it markedly less perceptually prominent or accessible than the focal representation. Generic omission of information that is absent from the available communication cannot satisfy this relation because the required focal-information/qualification salience comparison is not observable (Gray et al., 2018; Mathur et al., 2019; Luguri and Strahilevitz, 2021).
Salience concealment versus choice architecture	BOUNDARY distinguish information access from response access	*Salience-based concealment*; *Manipulative choice architecture*	Salience concealment asymmetrically structures the perceptual accessibility of information. Manipulative choice architecture asymmetrically structures the practical accessibility, effort, default status, sequencing, completion, or reversibility of response alternatives. A dark-pattern implementation may instantiate both where information and action paths are independently asymmetric.
Salience concealment versus obfuscation	BOUNDARY distinguish noticeability from comprehensibility	*Salience-based concealment*; *Obfuscation (jargon/complexity)*	Salience concealment makes materially relevant information disproportionately difficult to notice or attend to. Obfuscation leaves information available to attention but makes it difficult to understand or process. Both may co-occur where different aspects of the communication independently satisfy both relations.

J.1.6 Distraction

Supplementary Table S43: Provenance of canonical Distraction Tactics.

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Whataboutism	P/A; Bel	Whataboutism, comparative counteraccusation, tu quoque reasoning, and meta-argumentative deflection	**KEEP; COMPARATIVE-DEFLECTION STRUCTURE REQUIRED.** An established criticism, accusation, question, or issue under scrutiny is answered with another case, wrongdoing, inconsistency, actor, or neglected concern that both redirects scrutiny and functions as a reason to qualify, suspend, weaken, or reconsider the focal judgement. The expression "what about" or a comparison alone is insufficient.	(Bowell, 2023; Aikin and Casey, 2026; Phi et al., 2024)
Red herring	P/A	Relevance fallacies, diversionary argument, and argumentative distraction	**KEEP; FOCAL RELEVANCE RELATION REQUIRED.** A consideration, issue, claim, event, or line of argument is introduced so as to displace engagement with an identifiable focal matter without substantively addressing that matter in the role for which the new consideration is introduced. Semantic unrelatedness or abrupt topic change is not required.	(Walton, 2004; Walton, 2005)
Straw man	Bel; P/A	Straw-man argumentation, position substitution, and distorted opponent representation	**KEEP; RECOVERABLE SOURCE POSITION REQUIRED.** An identifiable opponent's position, argument, proposal, or commitment is represented as a materially different substitute position, and criticism or rebuttal is then directed toward that substitute. Both the relevant source position and its material alteration must be recoverable from the available annotation context; external reconstruction of what the actor "really meant" is not permitted.	(Walton, 1996; Aikin and Casey, 2018; Oswald and Lewiński, 2014)

Supplementary Table S43: Provenance of canonical Distraction Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Obfuscation (jargon/complexity)	P/A; A/C	Obfuscation, linguistic complexity, jargon, processing difficulty, and epistemic accessibility	**KEEP; PROCESSING-BURDEN OPERATION REQUIRED.** Focal content is presented through jargon, abstraction, technical density, syntactic complexity, nominalisation, or comparable linguistic opacity that materially increases the effort required to recover, comprehend, evaluate, or scrutinise the content in the available communicative context. Technical language or intrinsic subject complexity alone is insufficient.	(Supran and Oreskes, 2017; Brunton and Nissenbaum, 2015; Markowitz and Hancock, 2016; Bullock et al., 2019; Shulman et al., 2020)
Agenda shifting / evasive refocusing	P/A	Agenda control, question evasion, conversational refocusing, and issue substitution	**KEEP; AGENDA SUBSTITUTION REQUIRED.** An identifiable question, criticism, issue, criterion, comparison, actor, time frame, level of analysis, or other focal matter is replaced or refocused so that discourse proceeds on a different agenda rather than resolving the one already established. Ordinary topic development or relevant contextualisation is insufficient.	(Greatbatch, 1986; Rasiah, 2010; Gabrielsen et al., 2020)
Information flooding / overload	P/A; A/C	Strategic distraction through information volume, flooding, information overload, and competing-content saturation	**KEEP; CUMULATIVE VOLUME IS CONSTITUTIVE.** A message or recoverable message stream contains a sufficiently dense accumulation of secondary, tangential, repetitive, weakly prioritised, or competing content that the accumulation itself materially increases the processing and selection burden required to isolate or scrutinise focal information. Length, detail, repetition, or posting frequency alone is insufficient.	(King et al., 2017; Cirone and Hobbs, 2023; Stukal et al., 2022; Persson, 2018)

Supplementary Table S43: Provenance of canonical Distraction Tactics (continued)

Canonical Tactic	Fixed targets	Literature basis	Ontology interpretation / decision	Key sources
Humor as deflection	P/A; Att	Political humour, irony, joking, non-serious response, and humour-mediated evasion	**KEEP; DEFLECTION REQUIRED.** Humor, irony, sarcasm, playful absurdity, or another recognizably non-serious mode is used in place of or materially displaces substantive engagement with an identifiable question, criticism, accusation, challenge, or issue while also repositioning its evaluative stance. Humor itself is not sufficient.	(Carpenter et al., 2019; Nuolijärvi and Tiittula, 2011; Chernobrov, 2024)
Ad hominem / personal-credibility attack	P/A; Bel; Att	Ad hominem argument, personal attack, source-credibility undercutting, and political personal attack	**KEEP; PERSON-TO-CONTRIBUTION RELATION REQUIRED.** A person is negatively characterised through their character, motives, competence, trustworthiness, affiliations, circumstances, conduct, or consistency, and that characterisation is used to discount an identifiable contribution, diminish the person's standing as a source or critic, or substitute person-directed scrutiny for engagement with the substantive issue. Personal criticism or insult alone is insufficient.	(Walton, 1998; Walton, 2008; Aikin, 2025)

Target abbreviations: P/A = Perception/Attention; Bel = Belief; Att = Attitude; A/C = Agency/Choice.

Distraction decision ledger. Table S44 records the principal distinctions used to derive the final Distraction inventory. The family is defined by observable displacement, redirection, substitution, accessibility burden, or contribution-undercutting relations rather than by inferred confusion, evasive intent, audience distraction, or argumentative fallaciousness.

Supplementary Table S44: Principal Distraction inclusion, distinction, and boundary decisions.

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Generic diversion / distraction	SPLIT: by displacement relation	*Whataboutism*; *Red herring*; *Straw man*; *Agenda shifting / evasive refocusing*; *Information flooding / overload*; *Humor as deflection*; applicable *Ad hominem / personal-credibility attack*	A generic "distraction" label would collapse several observably different relations. The final inventory distinguishes comparative countercases, tangential diversion, substitution of an opponent's position, replacement of the focal agenda, cumulative overload, humorous displacement, and person-directed credibility undercutting. Each therefore requires its own minimum evidence rather than inferred recipient distraction.
Whataboutism versus ordinary comparison	BOUNDARY: distinguish by deflective function	*Whataboutism*	Comparison, identification of inconsistent treatment, or reference to another case can be substantively relevant. Whataboutism requires both redirection from an established point of scrutiny and use of the comparative or counteraccusatory case as a reason to qualify, weaken, suspend, or reconsider the focal judgement. The words "what about" carry no independent diagnostic status (Bowell, 2023; Aikin and Casey, 2026).
Tu quoque / appeal to hypocrisy	MERGE: decomposed by function	*Whataboutism*; *Ad hominem / personal-credibility attack*; *Induced hypocrisy / self-inconsistency pressure*	An allegation of hypocrisy does not define one unique influence operation. Where the critic's inconsistency functions as a comparative countercase that redirects scrutiny, it may instantiate *Whataboutism*. Where another actor's inconsistency is used to discount that person's credibility or contribution, it may instantiate an *Ad hominem / personal-credibility attack*. Where the recipient's own endorsed standard is juxtaposed with their inconsistent conduct and used as pressure for subsequent alignment, it instantiates *Induced hypocrisy / self-inconsistency pressure*. Co-coding requires each respective relation to be independently recoverable.
Whataboutism versus false equivalence	BOUNDARY: distinguish deflection from comparability distortion	*Whataboutism*; *False equivalence / distorted comparability*	Whataboutism requires a comparative countercase to redirect scrutiny and bear on the focal judgement; it does not require the two cases to be represented as equivalent. False equivalence requires communicated parity or equivalence on an identifiable dimension despite a recoverable material difference on that same dimension. A whataboutist comparison may satisfy both only when both relations are independently present.

Supplementary Table S44: Principal Distraction decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Red herring versus whataboutism	BOUNDARY distinguish generic diversion from comparative counter-case	*Red herring*; *Whataboutism*	A Red herring introduces a particular consideration that displaces engagement with an established focal issue without substantively resolving it. Whataboutism is the more specific structure in which the diversion takes the form of another case, wrongdoing, inconsistency, actor, or neglected concern that additionally functions as a comparative reason bearing on the focal judgement (Walton, 2004; Bowell, 2023).
Red herring versus agenda shifting	BOUNDARY distinguish diversionary item from agenda substitution	*Red herring*; *Agenda shifting / evasive refocusing*	A Red herring principally introduces a diversionary consideration as putative engagement with the established issue. Agenda shifting instead replaces or refocuses the issue, question, dimension, comparison, criterion, actor, time frame, or other agenda that discourse is to resolve. A Red herring's success in redirecting attention does not automatically establish agenda shifting.
Straw man versus unfavourable interpretation	BOUNDARY require material position substitution	*Straw man*	Disagreement, criticism, simplification, paraphrase, summary, unfavourable interpretation, or the phrase "so you are saying" is insufficient. The available annotation context must establish an identifiable source position, a materially different representation attributed to the same actor, and criticism or rebuttal directed at the substitute position (Walton, 1996; Aikin and Casey, 2018; Oswald and Lewiński, 2014).
Straw man versus external reconstruction	BOUNDARY message-internal constraint	*Straw man* only where source position and distortion are recoverable	Annotators do not reconstruct an opponent's "real" position from external knowledge. Both the relevant position and its materially altered substitute must be recoverable from the available message, message sequence, or explicitly represented source material. This preserves the message-internal annotation rule while retaining position distortion as an observable Tactic.

Supplementary Table S44: Principal Distraction decisions (continued)

Candidate construct / overlap	**Status / qualifier**	**Canonical representation**	**Decision rationale**
Obfuscation versus strategic vagueness	BOUNDARY distinguish indeterminacy from processing difficulty	*Obfuscation (jargon/complexity)*; *Strategic vagueness / equivocation*	Obfuscation presents focal content in a form that materially increases the effort required to comprehend, recover, evaluate, or scrutinise it. Strategic vagueness instead leaves a materially relevant proposition, quantity, referent, responsibility relation, stance, condition, or commitment semantically or pragmatically underdetermined. Determinate but difficult content and indeterminate content are therefore separate operations.
Jargon / technical language	BOUNDARY no inheritance	*Obfuscation (jargon/complexity)* only where processing burden is recoverable	Specialist vocabulary, abbreviations, formal language, abstraction, technical detail, complex syntax, or intrinsic subject difficulty can be appropriate and informative. Obfuscation requires the linguistic packaging itself to materially burden access to or scrutiny of focal content in the available communicative context; jargon is therefore evidence only when it performs that operation (Bullock et al., 2019; Shulman et al., 2020).
Obfuscation versus salience concealment	BOUNDARY distinguish comprehensibility from noticeability	*Obfuscation (jargon/complexity)*; *Salience-based concealment*	Obfuscation leaves focal content perceptually available but makes it materially difficult to understand or scrutinise. Salience-based concealment instead makes materially qualifying information disproportionately difficult to notice relative to the focal representation. The distinction is therefore processing accessibility versus perceptual accessibility.
Obfuscation versus information flooding	BOUNDARY distinguish linguistic opacity from cumulative volume	*Obfuscation (jargon/complexity)*; *Information flooding / overload*	Obfuscation burdens processing primarily through the linguistic packaging of focal content. Information flooding burdens processing and selection through the volume or accumulation of multiple competing informational items. Individual items in a flooding episode may be perfectly clear; individual obfuscated passages need not occur in a high-volume stream.

Supplementary Table S44: Principal Distraction decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Agenda shifting versus ordinary topic development	BOUNDARY: require displaced focal agenda	✗ *Agenda shifting / evasive refocusing*	Natural progression to a related topic, additional context, a relevant change in level of analysis, or partial response to a compound question does not instantiate agenda shifting. The available context must establish an issue, question, criterion, comparison, actor, time frame, or other focal dimension whose resolution is displaced by a substitute agenda (Greatbatch, 1986; Rasiah, 2010; Gabrielsen et al., 2020).
Information volume / long messages	BOUNDARY: no inheritance	✗ *Information flooding / overload* only where accumulation performs the distracting operation	Length, detail, repetition, numerous statistics, a large evidential record, or high posting frequency is insufficient. The available message or message stream must show that accumulation of competing material itself substantially increases the burden required to isolate, prioritise, or scrutinise identifiable focal information (King et al., 2017; Cirone and Hobbs, 2023; Persson, 2018).
Information overload as audience state	BOUNDARY: mechanism/effect, not licensing condition	✗ *Information flooding / overload*	The annotatable operation is the recoverable high-volume or dense competing-content structure. Annotators do not need to infer that a recipient actually experienced confusion, fatigue, overload, indecision, or reduced comprehension. Those are possible psychological consequences of the observable operation rather than required evidence for the Tactic.
Humor / irony / sarcasm	BOUNDARY: no inheritance	✗ *Humor as deflection* only where focal engagement is displaced	Humorous, ironic, sarcastic, playful, parodic, or absurd communication may directly express an argument, evaluation, criticism, identity claim, or political message without distracting from anything. The Distraction Tactic requires an identifiable issue or point of scrutiny whose substantive engagement is materially displaced through the humorous turn (Carpenter et al., 2019; Nuolijärvi and Tiittula, 2011; Chernobrov, 2024).

Supplementary Table S44: Principal Distraction decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Humor as deflection versus ridicule	BOUNDARY: distinguish by deflective function	*Humor as deflection*; *Ridicule / mockery*	Humor as deflection requires humorous or non-serious response to displace substantive engagement with an identifiable focal issue. Ridicule requires a target-directed operation that renders an identifiable person, group, institution, idea, policy, action, or other object laughable, foolish, absurd, contemptible, or otherwise diminished. Either may occur without the other; co-code only where both functions are independently recoverable.
Personal criticism / insult	BOUNDARY: no inheritance	*Ad hominem / personal-credibility attack* only where the person-to-contribution relation is present	Negative characterisation, insult, mockery, labeling, allegations of bias, affiliation, misconduct, or hypocrisy do not themselves constitute an ad hominem Tactic. The personal characterisation must function to discount an identifiable claim, argument, testimony, criticism, question, or other contribution, diminish the person's standing as a source or critic, or substitute person-directed scrutiny for engagement with the substantive issue (Walton, 2008; Aikin, 2025).
Ad hominem versus name-calling	BOUNDARY: distinguish characterisation from contribution undercutting	*Ad hominem / personal-credibility attack*; *Name-calling / labeling*	Name-calling requires a condensed evaluative label or characterisation of a target. Ad hominem requires a person-directed characterisation to perform the additional argumentative function of undermining that person's contribution or source standing. A pejorative label may satisfy both only where those two operations are independently recoverable.
Ad hominem versus induced hypocrisy	BOUNDARY: distinguish other-directed credibility attack from self-directed alignment pressure	*Ad hominem / personal-credibility attack*; *Induced hypocrisy / self-inconsistency pressure*	The same inconsistency can enter different influence structures. An ad hominem Tactic uses another actor's inconsistency or hypocrisy as grounds for discounting that actor's claim, argument, testimony, criticism, or standing. Induced hypocrisy confronts the recipient with their own endorsed standard and inconsistent conduct and uses the discrepancy as pressure for subsequent behavioural alignment.

Supplementary Table S44: Principal Distraction decisions (continued)

Candidate construct / overlap	Status / qualifier	Canonical representation	Decision rationale
Logical fallacy status	BOUNDARY not required	*Whataboutism*; *Red herring*; *Straw man*; *Ad hominem / personal-credibility attack* as applicable	Annotation records the communicated influence structure rather than adjudicating formal or informal fallaciousness. A source-characteristic argument may sometimes be relevant to credibility; a comparison may sometimes be germane; and contextual information may legitimately redirect discussion. The applicable Tactic is licensed by its operational message relation, not by an annotator's judgement that the argument is logically invalid or normatively improper.

J.1.7 Cross-family candidate disposition

The canonical inventory is not intended to preserve every terminology variant, surface form, psychological mechanism, or campaign-level construct encountered in the literature and framework-coverage review as a separate Tactic. Table S45 records selected candidate constructs whose disposition is particularly informative for understanding the boundaries and coverage of the final Method/Tactic vocabulary. The table is a construct-selection record rather than a chronology of ontology revisions: it states how each candidate is represented under the final inclusion, observability, and layer-assignment rules.

Supplementary Table S45: Disposition of selected non-canonical Method/Tactic candidates.

Candidate construct / form	Status / qualifier	Canonical representation	Rationale
Plain-folks / common-person appeal	MERGE	*Ingratiation*; or *Unity (shared identity)* where independently satisfied	Common-person presentation does not define a unique influence operation. Ordinariness, similarity, humility, relatability, or informality is represented as *Ingratiation* where it functions to cultivate favourable evaluation of the source. Where shared membership in an identifiable collective itself provides the warrant, *Unity* applies. Surface markers such as ordinary clothing, informal language, or familiar settings are insufficient on their own (Institute for Propaganda Analysis, 1939; Jones, 1964; Gordon, 1996; Cialdini, 2016).

Supplementary Table S45: Disposition of selected non-canonical Method/Tactic candidates (continued)

Candidate construct / form	**Status / qualifier**	**Canonical representation**	**Rationale**
Glittering generalities / generic virtue words	REJECT: decomposed across existing categories	*Loaded language*; *Moral framing*; or another specific warrant where independently satisfied	Generic positive-value terminology does not identify one stable influence relation. Positive evaluative connotation may instantiate *Loaded language*; an identifiable moral value or principle may instantiate *Moral framing*; and authority, tradition, novelty, shared identity, or other specific warrants are represented through their own canonical Tactics. The broad historical form is therefore covered without a residual generic virtue-word label (Institute for Propaganda Analysis, 1939; Da San Martino et al., 2020; Feinberg and Willer, 2015).
Bandwagon	MERGE: canonical alias	*Social proof (bandwagon)*	The historical bandwagon category is represented by the more theoretically specific social-proof operation: represented beliefs, approval, choices, or conduct of others function as social evidence for corresponding judgement or behaviour. A mere reference to popularity or a crowd is insufficient unless that plurality performs the warranting operation (Institute for Propaganda Analysis, 1939; Deutsch and Gerard, 1955; Cialdini, 2007).
Testimonial	SPLIT: by function	*Appeal to authority*; *Source / channel credibility*; *Social proof*; *Associative transfer / guilt-or-honor by association*, as independently satisfied	A testimonial is a communicative form rather than one unique influence operation. Expertise or recognised authority may warrant acceptance of a claim; source or outlet provenance may provide a credibility warrant; another person's endorsement may function as social evidence; and represented prestige or stigma may be transferred through association. The canonical Tactic is therefore determined by the function performed by the testimonial, not by testimonial form alone (Institute for Propaganda Analysis, 1939; Cialdini, 2007; Hovland and Weiss, 1951; Deutsch and Gerard, 1955).

Supplementary Table S45: Disposition of selected non-canonical Method/Tactic candidates (continued)

Candidate construct / form	**Status / qualifier**	**Canonical representation**	**Rationale**
Transfer	MERGE: decomposed across existing categories by function	*Associative transfer / guilt-or-honor by association*; *Unity (shared identity)*; or other independently satisfied operation	Where an association is used to transfer represented qualities, commitments, credibility, prestige, stigma, blame, approval, or evaluative standing to a focal target, it is represented by *Associative transfer*. Where a symbol or association instead makes shared group membership itself the warrant, *Unity* may apply. Symbols and juxtaposition do not receive a generic transfer label without a recoverable influence relation (Institute for Propaganda Analysis, 1939; Mohammed, 2023; Cialdini, 2016).
Card stacking	MERGE: conditional	*Selective evidence presentation (cherry-picking)* where its minimum evidence is satisfied	Selective presentation is represented as cherry-picking only when the available annotation context makes a broader evidential or consideration set recoverable and shows systematic selection of supportive members while countervailing members are excluded or denied comparable evidential use. One-sided advocacy or failure to mention counterevidence that would have to be supplied externally is insufficient (Institute for Propaganda Analysis, 1939; Chiu et al., 2017; Farina et al., 2026).
Loaded / complex question	MERGE: as form	*Presuppositional embedding* where independently satisfied	Interrogative form is not a standalone influence Tactic. A question instantiates *Presuppositional embedding* where it communicates an unestablished determinate proposition as taken for granted rather than as the focal assertion. Other leading, difficult, rhetorical, or adversarial questions receive no special Tactic unless another canonical operation is independently satisfied (Thoma et al., 2023; Macagno, 2024).

Supplementary Table S45: Disposition of selected non-canonical Method/Tactic candidates (continued)

Candidate construct / form	**Status / qualifier**	**Canonical representation**	**Rationale**
Slogans	REJECT: standalone Tactic	Narrative *Symbolic condensation*; plus any independently satisfied Method Tactic	Slogan form does not specify one influence relation. A slogan may compress collective narratives or symbols, repeat a proposition, use loaded language, activate shared identity, make a claim, or perform several other operations. Its slogan status is therefore representational form rather than sufficient evidence for a Method Tactic. Method labels are added only where their own minimum conditions are satisfied (Edelman, 1988; Hariman and Lucaites, 2007).
Thought-terminating clichés	REJECT: decomposed across existing categories	Applicable *Loaded language*; *Obfuscation (jargon/complexity)*; Narrative *Symbolic condensation*; or another independently recoverable operation	Formulaicity, familiarity, or cliché status does not define a unique message-internal influence relation. Where the expression performs evaluative lexical framing, obstructs substantive processing, compresses a narrative or symbolic structure, or independently instantiates another canonical Tactic, that operation is coded directly. The phrase's conventional or repetitive form alone is insufficient.
Satire, parody, sarcasm, irony, or caricature	REJECT: standalone Tactic	*Ridicule / mockery*; *Humor as deflection*; or other independently satisfied operation	These are expressive modes rather than uniquely determined influence relations. They instantiate *Ridicule / mockery* where they render an identifiable target laughable, foolish, absurd, contemptible, or otherwise diminished; they instantiate *Humor as deflection* where humorous or non-serious presentation displaces substantive engagement with an established point of scrutiny. They may also directly communicate arguments or evaluations without satisfying either Tactic (D'Errico and Poggi, 2016; Zappavigna, 2022; Carpenter et al., 2019).

Supplementary Table S45: Disposition of selected non-canonical Method/Tactic candidates (continued)

Candidate construct / form	**Status / qualifier**	**Canonical representation**	**Rationale**
Cognitive dissonance	REJECT: standalone Tactic	Possible mechanism of *Induced hypocrisy / self-inconsistency pressure* and other influence operations	Cognitive dissonance is an internal psychological state or explanatory mechanism rather than a directly observable message operation. The ontology therefore annotates recoverable communicative structures that may elicit such a state—for example, juxtaposition of a recipient's endorsed standard with their inconsistent conduct—without inferring that dissonance was actually experienced (Aronson et al., 1991; Stone et al., 1994; Stone et al., 1997).
That's-not-all technique	REJECT: no separate canonical Tactic	Applicable *Reciprocity*, reward, scarcity, commitment, or other canonical operation where independently satisfied	The literature describes a sequential compliance procedure in which an offer is improved before the recipient responds. The procedure is theoretically associated with mechanisms including reciprocity and altered evaluation of the deal. For the present message-internal ontology, an improving-offer sequence does not by itself add a distinct influence warrant beyond the independently recoverable canonical operations instantiated by the communication. Consequently, the relevant canonical Tactic is determined by the observable warrant or leverage relation rather than by the procedural label alone (Burger, 1986).
Paltering	REJECT: no separate canonical Tactic	Applicable *Strategic vagueness / equivocation*, *Selective evidence presentation*, *Salience-based concealment*, or another independently satisfied form-level Tactic	Paltering is defined in the literature through the use of truthful statements to produce a misleading impression. Determining that relation ordinarily requires adjudicating the truth of the component statements and the misleadingness of the resulting impression, which conflicts with the ontology's truth- and intent-agnostic message-internal annotation rule. Observable forms through which such an impression is constructed remain codable when they independently satisfy a canonical Tactic (Rogers et al., 2017).

Supplementary Table S45: Disposition of selected non-canonical Method/Tactic candidates (continued)

Candidate construct / form	**Status / qualifier**	**Canonical representation**	**Rationale**
Generic lying, misinformation, or disinformation	REJECT: generic Method Tactic	Applicable canonical Deception or other Tactic where its message-internal conditions are independently satisfied	Truth or falsity of a proposition and the communicator's knowledge or intent are not generic Method-level annotation criteria. A message may therefore contain misinformation or intentional deception in an external factual sense without receiving a generic "lying" or "disinformation" Tactic. Conversely, a Deception-family Tactic can be coded where its recoverable form-level relation is present without establishing objective falsity or deceptive intent.
Information overload, confusion, indecision, or cognitive fatigue	REJECT: audience-state labels	*Information flooding / overload*; *Obfuscation*; or another observable operation where independently satisfied	Internal processing states and downstream audience effects are not themselves Tactics. *Information flooding / overload* requires a recoverable high-volume or dense competing-content structure; *Obfuscation* requires linguistic opacity or complexity that materially burdens access to focal content. Annotators do not infer actual confusion, fatigue, indecision, or overload.
Campaign-level repetition, amplification, boosting, or seeding	RESCOPE: message-level portion retained	*Repetition* only where recurrent propositional content is recoverable within the available message or message sequence; otherwise outside Method scope	Scheduling, reach, repeated distribution, amplification, boosting, and campaign-level exposure are delivery properties rather than message-internal influence operations. Repetition remains canonical only where the same or substantially equivalent proposition is recoverably repeated and recurrence itself functions as reinforcement.
Sockpuppets, coordinated accounts, asset creation, operational infrastructure, or actor coordination	REJECT: outside unit of analysis	Outside Method/Tactic annotation	Account identity, asset acquisition, coordination, infrastructure, operational evasion, and comparable campaign behaviours concern actors, networks, platform operations, or campaign organisation rather than the form and function of the individual communication. They therefore remain outside the message-internal Method/Tactic vocabulary.

J.2 TargetConstruct provenance and mapping verification

TargetConstruct is not independently annotated. Observable message-internal evidence supports assignment of a canonical Tactic; the Method family and the fixed TargetConstruct set are then inherited deterministically from that Tactic. The same Tactic therefore yields the same TargetConstruct set wherever it is annotated. TargetConstruct does not record an empirically observed audience effect and is not inferred separately from the surrounding message.

The six-category vocabulary is an operational partition of audience-state dimensions that are constitutively implicated by the canonical Tactics. It is not intended as an exhaustive taxonomy of psychological processes or as a temporal sequence of influence effects. A TargetConstruct is included in a Tactic's fixed mapping only where that dimension is part of the Tactic's defining influence operation. Psychological mechanisms, rhetorical leverage, likely reactions, and downstream consequences do not independently add TargetConstruct values.

Supplementary Table S46: Conceptual provenance and inclusion rules for TargetConstructs.

TargetConstruct	Conceptual basis	Ontology inclusion rule	Key sources
Perception / Attention	Attention, selective processing, limited processing capacity, perceptual accessibility, and relative salience	Mapped only where changing attentional access, allocation, selection, processing priority, or relative salience is constitutive of the Tactic. Recoverable operations include foregrounding or backgrounding information, redirecting attention between focal objects, perceptually suppressing available information, or imposing competing processing demands. Mere vividness, emotionality, emphasis, importance, or evaluative force does not yield the mapping.	(Posner, 1992; Lang, 2000; Entman, 1993; Mack and Rock, 1998; Simons and Chabris, 1999)
Belief	Propositional representation, belief formation and revision, causal and relational representation, and acceptance of claims	Mapped where the defining operation bears on what is represented or accepted as being the case, including establishment, support, dispute, qualification, reframing, distortion, or undermining of a proposition or relation. *Belief* concerns propositional acceptance and is distinguished from evaluative orientation.	(Fishbein and Ajzen, 1975; Ajzen, 1991; Kunda, 1990)

Supplementary Table S46: Conceptual provenance and inclusion rules for TargetConstructs (continued)

TargetConstruc	Conceptual basis	Ontology inclusion rule	Key sources
Attitude	Favourable–unfavourable evaluation, persuasive evaluation, attraction and aversion, approval and disapproval	Mapped where alteration or positioning of evaluative stance toward an object, actor, issue, policy, behaviour, or outcome is intrinsic to the Tactic. Evaluation that may arise merely as a downstream consequence of changing a belief, identity, or behaviour does not independently license the construct.	(Ajzen, 1991; Petty and Cacioppo, 1986)
Identity	Personal self-concept, social identification, group membership, and identity-based social influence	Mapped where activation, affirmation, threat, strengthening, weakening, or redefinition of who the recipient is, aspires to be, or belongs with is part of the defining influence operation. Mere reference to a social group, identity-relevant topic, or group relationship is insufficient where identity itself is not operated on.	(Tajfel and Turner, 1979; Bem, 1972; Cialdini and Goldstein, 2004)
Agency / Choice	Personal agency, epistemic agency, perceived control, freedom of choice, capability, and practical availability of alternatives	Mapped where the Tactic constitutively alters, constrains, scaffolds, or undermines the recipient's capacity or perceived capacity for self-directed judgement or action. This includes epistemic self-direction and practical availability, feasibility, controllability, or viability of alternatives. Providing reasons for preferring one otherwise available option, or merely changing its expected attractiveness, does not itself yield the mapping.	(Bandura, 2001; Proust, 2008; Brehm, 1966; Ajzen, 1991; Bandura, 1997)
Behavior	Observable conduct, compliance, conformity, action, inhibition, and behavioural performance	Mapped where eliciting, inhibiting, maintaining, selecting, or altering observable conduct is intrinsic to the Tactic's defining operation. Behavior does not follow automatically because changes in belief, evaluation, attention, identity, or agency might eventually affect action. Nor does the mapping imply that the recipient actually performs the targeted behaviour.	(Ajzen, 1991; Cialdini and Goldstein, 2004; Freedman and Fraser, 1966; Cialdini et al., 1975)

TargetConstruct boundary decisions. The fixed mappings apply several cross-cutting distinctions that prevent psychological mechanisms and

plausible downstream consequences from expanding a Tactic's TargetConstruct profile.

Supplementary Table S47: Principal TargetConstruct boundary decisions.

Boundary	**Decision rule**	**Rationale**
Mechanism versus TargetConstruct	Map only constitutive target dimensions	A psychological state or process through which a Tactic may exert influence does not become a TargetConstruct merely because it is theoretically involved. Fear, anger, guilt, sympathy, dissonance, processing fluency, cognitive load, or similar mechanisms contribute a mapping only where the corresponding TargetConstruct operation is independently constitutive of the Tactic.
Downstream effect versus TargetConstruct	Do not inherit plausible consequences	A Tactic aimed at belief does not automatically inherit Attitude or Behavior because changed beliefs may alter evaluation or conduct. Likewise, Attitude does not automatically imply Behavior, Identity does not automatically imply Attitude, and attentional capture does not automatically imply Belief. Fixed mappings represent the Tactic's defining target profile, not its complete possible causal pathway.
Perception / Attention versus salience as ordinary communication	Require differential attentional operation	P/A is not assigned merely because information is vivid, emotional, important, repeated, evaluative, or foregrounded in the ordinary sense of being communicated. The Tactic must constitutively change relative accessibility, allocation, processing priority, selection, or salience in a recoverable way.
Belief versus Attitude	Separate proposition from evaluation	*Belief* concerns what is represented or accepted as being the case; *Attitude* concerns favourable or unfavourable evaluation. A Tactic maps to both only where propositional and evaluative operations are each constitutive rather than because one could causally influence the other.
Identity versus identity relevance	Require self-concept or identification operation	Identity-related content, an ingroup or outgroup referent, group membership, family relation, social status, or collective consequence does not by itself yield *Identity*. The Tactic must constitutively activate, position, affirm, threaten, or redefine self-concept or social identification.

Supplementary Table S47: Principal TargetConstruct boundary decisions (continued)

Boundary	Decision rule	Rationale
Agency / Choice versus Behavior	Separate conditions of choosing from targeted conduct	*Behavior* concerns the act toward which the Tactic is directed. *Agency/Choice* concerns the recipient's capacity or perceived capacity for independent judgement or action and the practical availability, feasibility, controllability, or viability of alternatives. A Tactic may map to one or both depending on its constitutive operation.
Stable class mapping versus instance-specific inference	Use one fixed set per canonical Tactic	Where a Tactic has multiple TargetConstructs, the set represents its stable theoretically supported target profile. Annotators do not add or remove TargetConstruct values according to what a particular message instance seems likely to affect or according to an observed audience reaction.

Deterministic mapping audit. Because TargetConstruct is derived rather than independently annotated, its reproducibility can be checked mechanically against the canonical Tactic–TargetConstruct relation. Table S48 summarises the final mapping inventory. Counts are non-exclusive because a canonical Tactic may map to more than one TargetConstruct.

Supplementary Table S48: Deterministic TargetConstruct mapping audit.

Method family	Tactics	P/A	Bel	Att	Id	A/C	Beh
Persuasion	25	0	15	9	2	3	12
Framing	16	1	10	9	1	2	2
Manipulation	10	2	5	4	3	2	8
Coercion	5	0	0	0	0	5	5
Deception	7	1	7	4	0	0	0
Distraction	8	8	3	2	0	2	0
Total	**71**	**12**	**40**	**28**	**6**	**14**	**27**

P/A = Perception/Attention; Bel = Belief; Att = Attitude; Id = Identity; A/C = Agency/Choice; Beh = Behavior. Column totals exceed 71 because mappings are multi-valued.

Verification conditions. The TargetConstruct mapping is internally consistent if and only if all of the following conditions hold:

1. every canonical Tactic belongs to exactly one Method family;

2. every canonical Tactic has exactly one fixed, non-empty TargetConstruct set;
3. the Tactic–TargetConstruct relation in the machine-readable specification reproduces the mappings reported in the operational Tactic catalogue;
4. the reverse index in Table S1 is exactly the inverse of that relation;
5. the union of the six TargetConstruct reverse-index sets contains all 71 canonical Tactics and no non-canonical Tactic; and
6. annotation software or downstream analysis derives TargetConstruct exclusively from the assigned Tactic and does not permit independent instance-level TargetConstruct assignment.

These conditions concern specification integrity. They do not require a separate inter-annotator reliability estimate for TargetConstruct, because annotators do not independently assign that variable. Reliability of TargetConstruct assignment is inherited from Tactic assignment together with the deterministic integrity of the mapping.

J.3 Narrative provenance and derivation

The Narrative dimension is specified independently of the Method/Tactic dimension. Narrative coding records message-internal story organisation, actor roles, temporal and causal structuring, symbolic and metaphorical organisation, moral positioning, exemplification, and represented stakes. The presence of a NarrativeDevice does not automatically imply a corresponding Method Tactic, TargetConstruct, Polarization motif, or AboutObject.

Within the Narrative dimension, the ontology distinguishes two levels. *NarrativeDevices* and their applicable attributes are identified from message-internal evidence. *NarrativeArchetypes* are macro-story packages derived from diagnostic configurations of those coded devices and attributes. The minimal signatures reported in Table S9 provide the formal basis for deterministic archetype assignment; descriptive cues in Table S8 aid recognition but are not additional conjunctive activation requirements.

Accordingly, provenance is reported separately for the device vocabulary and for the archetype signatures. The former records the theoretical traditions from which the narrative mechanisms are operationalised; the latter records the literature-supported story packages represented through formally defined configurations of those mechanisms.

J.3.1 NarrativeDevice provenance

Supplementary Table S49: Conceptual provenance of canonical NarrativeDevices.

NarrativeDevice	Literature basis	Ontology interpretation	Key sources
Role-casting (actants)	Narrative actants, character functions, and strategic narrative actor roles	Records assignment of actors to story-functional positions such as hero/protector, villain/opponent, victim, traitor, or savior/strong leader. The device concerns the actor's role within the represented story structure; it does not by itself imply that the recipient is being assigned a role or behavioural expectation.	(Propp, 2010; Greimas, 1966; Roselle et al., 2014)
Conflict–resolution arc	Narrative grammar, narrative coherence, conflict, complication, and resolution	Records a story structure in which a situation or problem develops through conflict, struggle, or complication toward an actual, promised, proposed, or anticipated resolution. The device describes narrative organisation rather than independently establishing a behavioural request or persuasive effect.	(Bruner, 1990; Fisher, 1984; Green and Brock, 2000)
Temporal sequencing & contrast	Narrative temporality, sequencing, temporal comparison, and emplotment	Records meaningful organisation of events or conditions through temporal relations such as before/after, decline/restoration, crisis/solution, tipping-point, continuity, or progress structures. Chronological reference alone is insufficient where temporal organisation performs no narrative structuring function.	(Ricœur, 1984; Linde, 1993)
Causal emplotment	Narrative causality, emplotment, attribution, and explanatory story structure	Records represented causal relations linking actors, events, conditions, or outcomes into an explanatory narrative—for example, who or what produced a problem and through what relation. Ordinary causal representation does not automatically imply oversimplification, scapegoating, conspiracy, or another Method operation.	(White, 1973; Ricœur, 1984; Entman, 1993)
Metaphorical mapping	Conceptual metaphor and metaphorical organisation of problem construal	Records recurring metaphorical source structures—such as war, invasion, flood, disease, decay, or machine/control imagery—used to organise a narrative representation. The NarrativeDevice records the mapping as a story-structuring mechanism and does not automatically instantiate the Method Tactic *Metaphor / analogy*.	(Lakoff, 2004; Thibodeau and Boroditsky, 2011)

Supplementary Table S49: Conceptual provenance of canonical NarrativeDevices (continued)

NarrativeDevice	Literature basis	Ontology interpretation	Key sources
Exemplars & vignettes	Episodic exemplification, exemplars, vivid cases, and concrete narrative illustration	Records concrete persons, cases, episodes, or short illustrative accounts used within a broader story or interpretive package. An exemplar need not function as statistical or representative evidence, and its presence does not by itself establish selective-evidence presentation.	(Iyengar, 1991; Zillmann, 1999)
Moral valuation	Narrative moral evaluation, virtue and vice, and moralised interpretation	Records moral positioning of actors, actions, relations, or developments within a narrative through such dimensions as care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, purity/degradation, or liberty/oppression. Moral content is not sufficient to instantiate the Method Tactic *Moral framing* unless the moral criterion independently performs that Tactic's interpretive or evaluative influence operation.	(Entman, 1993; Haidt, 2012)
Symbolic condensation	Political symbolism, icons, slogans, emblems, and compressed narrative representation	Records symbols or compact representational forms that condense broader collective narratives, identities, histories, conflicts, or values. Possession of slogan, emblem, icon, flag, or symbolic form does not by itself create a Method Tactic.	(Edelman, 1988; Hariman and Lucaites, 2007)
Stakes & urgency cues	Crisis rhetoric, threat salience, urgency, and representation of consequential stakes	Records narrative positioning of a situation as carrying low/moderate, high, or existential/systemic stakes within an identifiable domain. Urgency, danger, crisis, or last-chance language does not automatically instantiate *Appeal to fear*, *Threat framing*, or Coercion; those Method Tactics require their own independently recoverable influence relations.	(Tannenbaum et al., 2015; Moffitt, 2016; Brady et al., 2017)

Narrative cross-layer boundary decisions. NarrativeDevices may share content, lexical material, or evidence spans with other ontology dimensions. Such overlap is expected and does not imply identity between constructs. Table S50 records the principal boundaries used to prevent automatic cross-layer inheritance.

Supplementary Table S50: Principal Narrative cross-layer boundary decisions.

Narrative construct / overlap	**Independent construct**	**Boundary rule**
Narrative structure generally	*Storytelling / narrative packaging* / Persuasion	The Narrative dimension records story organisation whenever the corresponding devices are evidenced. The Persuasion Tactic *Storytelling / narrative packaging* additionally requires an observable narrative sequence to function as a persuasive vehicle through which a claim or evaluation is advanced. Narrative organisation therefore does not automatically instantiate the Method Tactic.
Role-casting (actants)	*Altercasting* / Persuasion	Narrative Role-casting assigns an actor a function within the represented story—for example hero, villain, victim, traitor, or savior. *Altercasting* instead requires the recipient to be assigned, projected into, or invited to assume a social or relational role carrying an associated behavioural expectation. A message may do both, but neither is inherited from the other.
Causal emplotment	*Causal oversimplification* / Framing	Narrative Causal emplotment requires a represented causal relation organising events or actors into an explanatory story. Causal oversimplification requires the stronger relation that one factor is represented as sufficient, overriding, or exhaustive of the relevant causal explanation. A causal story may therefore be complex and multi-causal without instantiating causal oversimplification.
Causal emplotment	*Scapegoating* / Framing; Polarization scapegoating where applicable	Attribution of causation within a narrative does not itself constitute scapegoating. Scapegoating requires concentrated blame or responsibility for a broader negative condition to be placed on an identifiable person or group; the Polarization motif additionally requires the relevant group-positioning relation. Co-code only where those extra conditions are independently satisfied.
Causal emplotment: conspiracy / hidden hand	*Conspiracy / hidden hand* NarrativeArchetype; Polarization conspiratorial attribution; AboutObject conspiratorial-worldview outcome	The narrative causal pattern records covert coordinated causation when that form is represented. The corresponding NarrativeArchetype is derived when its signature is met. Polarization requires an additional adversarial intergroup positioning relation, while an AboutObject concerning conspiratorial worldview requires a recoverable broader substantive outcome toward increased or decreased conspiratorial worldview. None is inherited merely from causal emplotment.

Supplementary Table S50: Principal Narrative cross-layer boundary decisions (continued)

Narrative construct / overlap	**Independent construct**	**Boundary rule**
Moral valuation	*Moral framing* / Framing	Narrative Moral valuation records the moral positioning of actors, events, or relations within a story. The Method Tactic *Moral framing* requires an identifiable moral value, principle, right, obligation, virtue, vice, or prohibition to function as the interpretive or evaluative frame of an influence operation. Moral content in a narrative does not automatically satisfy that Method criterion.
Metaphorical mapping	*Metaphor / analogy* / Framing	Narrative Metaphorical mapping records a metaphorical structure organising the story world or narrative representation. The Method Tactic *Metaphor / analogy* requires relational structure from a source or comparison domain to be used to organise interpretation or inference as an influence operation. A metaphor can therefore be narratively descriptive without independently satisfying the Method Tactic.
Metaphorical mapping: disease / parasite / contamination	*Dehumanizing representation* / Framing; Polarization Dehumanization	Disease or contamination imagery does not itself establish dehumanization. The Framing Tactic requires an identifiable human target to be materially represented as less than fully human or through an animalistic, object-like, mechanistic, contaminating, parasitic, biological, or comparable nonhuman category. Polarization additionally requires the corresponding group-relational operation. Literal disease claims or metaphorical disease imagery may therefore occur without either dehumanization label.
Stakes & urgency cues	*Threat framing* / Framing	High or existential stakes describe the consequential structure of the narrative. *Threat framing* additionally requires an identifiable phenomenon to be represented as standing in a relation of danger, risk, hazard, or potential harm to a valued target. Urgency or crisis language alone does not automatically establish that framing operation.
Stakes & urgency cues	*Appeal to fear* / Persuasion	Narrative stakes may represent serious or existential consequences without using those consequences as persuasive leverage for behaviour. *Appeal to fear* requires threatened harm to function as a reason for adopting, maintaining, or performing an identifiable protective response. Possible recipient fear is not itself sufficient.

Supplementary Table S50: Principal Narrative cross-layer boundary decisions (continued)

Narrative construct / overlap	Independent construct	Boundary rule
Stakes & urgency cues	Applicable Coercion Tactic	Narrative urgency or threatened consequences are not coercive merely because they are severe. Coercion requires a prospective adverse consequence to be communicatively appropriated as contingent leverage over identifiable target conduct.
Exemplars & vignettes	*Statistical evidence* / Persuasion; *Selective evidence presentation* / Deception	A concrete example or vignette does not become statistical evidence merely because it supports a claim. Nor does selection of one example establish cherry-picking unless a broader evidential or consideration set and the selective relation are recoverable from the annotation context. Each Method operation therefore requires independent evidence beyond narrative exemplification.
Symbolic condensation	*Unity (shared identity)*, *Repetition*, *Loaded language*, or another Method Tactic	Flags, emblems, slogans, icons, chants, and other condensed symbols may carry narrative meaning without constituting a particular Method operation. *Unity* requires shared collective membership itself to function as the warrant; *Repetition* requires recurrent propositional content; *Loaded language* requires evaluative lexical positioning. Symbolic form alone licenses none of these.

J.3.2 NarrativeArchetype provenance and deterministic derivation

NarrativeArchetypes are not treated as mutually exclusive categories. Multiple archetypes may be derived from the same message where their respective minimal signatures are independently satisfied. Archetype overlap does not imply a parent–child relation among archetypes.

Supplementary Table S51: Provenance and deterministic signatures of NarrativeArchetypes.

Archetype	Literature basis	Minimal operational signature	Sources
Threat / Enemy at the gates	External-enemy, securitisation, and strategic-threat narratives	External RoleCasting as villain/opponent toward a populated collective referent, together with existential/systemic Stakes.	(Buzan et al., 1998; Keen, 1986; Miskimmon et al., 2013)
Enemy within (fifth column)	Internal-enemy, fifth-column, subversion, and collusion narratives	Internal RoleCasting as villain/opponent or traitor, together with CausalEmplotment pattern `subversion/collusion`.	(Wodak, 2015; Chernobrov, 2019; Radnitz and Mylonas, 2022)

Supplementary Table S51: Provenance and deterministic signatures of NarrativeArchetypes (continued)

Archetype	Literature basis	Minimal operational signature	Sources
Crisis & collapse	Political crisis, breakdown, and systemic-failure narratives	TemporalSequencing direction `decline` or `tipping_point`, existential/systemic Stakes, and a populated Stakes domain.	(Moffitt, 2016; Boin, 2005)
Corruption / rot at the top	Elite corruption, self-enrichment, and anti-establishment narratives	Elite or institutional RoleCasting as villain/opponent together with CausalEmplotment pattern `corruption/self_enrichment`.	(Edelman, 1988; Mudde and Kaltwasser, 2017)
Betrayal	Betrayal, violated loyalty, and internal-traitor narratives	Internal RoleCasting with Function `traitor`.	(Haidt, 2012; Chernobrov, 2019)
Moral decline / decadence	Moral panic, degradation, and cultural-decline narratives	TemporalSequencing direction `decline`, Stakes domain `cultural/moral`, and MoralValuation axis `purity/degradation`.	(Cohen, 2011; Haidt, 2012)
Golden age / restoration	Golden-age, nostalgia, and restorative strategic narratives	TemporalSequencing direction `restoration` and reference point `golden_past`.	(Boym, 2001; Roselle et al., 2014)
Savior / strong leader	Charismatic, heroic-leadership, and leader-as-rescuer narratives	RoleCasting with Function `savior/strong_leader`.	(Weber, 1947; Eatwell, 2006)
Collective victimhood	Collective victimhood, shared suffering, and group grievance narratives	RoleCasting with Function `victim` and ActorGroup `people_in_group` or `minority_in_group`.	(Noor et al., 2017)
Contamination / disease	Disease, parasite, contamination, and social-body metaphor narratives	Metaphor Type `disease/parasite/contamination`.	(Sontag, 1978; Musolff, 2015; Thibodeau and Boroditsky, 2011)
Conspiracy / hidden hand	Conspiracy, secret-plot, covert-coordination, and hidden-control narratives	CausalEmplotment Pattern `conspiracy/hidden_hand`.	(Hofstadter, 1965; Sunstein and Vermeule, 2009; Douglas et al., 2019)
Order vs. chaos (law-and-order)	Law-and-order, social control, disorder, and restoration-of-authority narratives	Stakes Domain `institutional/order` together with MoralValuation Axis `authority/subversion`.	(Hall et al., 1978; Wodak, 2015)

Narrative verification conditions. The Narrative specification is internally reproducible if all of the following conditions hold:

1. the canonical NarrativeDevice vocabulary contains exactly the devices specified in Table S4;
2. only attributes licensed for a present device are populated;
3. coded attribute values are restricted to the controlled vocabularies specified in Table S5;
4. each NarrativeArchetype is derived from the minimal signature in Table S9;
5. archetypes are non-exclusive and every satisfied signature may be returned independently;
6. no Method Tactic, TargetConstruct, Polarization motif, or AboutObject is inferred solely from a NarrativeDevice or archetype; and
7. where the machine-readable specification implements deterministic archetype derivation, its output exactly reproduces the signatures in Table S9.

Because NarrativeDevice and applicable attribute assignment require message-level interpretation, their empirical annotatability is not established by the present specification analysis. Future work should evaluate independent annotation agreement for device presence and applicable attribute values, together with disagreement patterns at difficult construct boundaries. Where NarrativeArchetypes are generated deterministically from those coded attributes, archetype reproducibility is principally an integrity property of the derivation rules: disagreement in derived archetypes should trace to device or attribute disagreement or to an implementation error rather than to a separate archetype-annotation judgement.

J.4 Polarization provenance and deterministic axis derivation

The Polarization dimension records message-internal group-positioning operations rather than population-level polarization or empirically observed changes in intergroup attitudes, identities, networks, or political outcomes. Groups and boundaries are coded as represented or constructed within the communication and need not correspond to externally verified or pre-existing social categories.

The dimension distinguishes three non-exclusive relational directions: Divide, Cohere, and Bridge. These labels are operational syntheses for message-level annotation rather than claims that the source literature contains a single standardised three-direction taxonomy. Each canonical motif belongs to one direction and maps deterministically to one or more polarization axes. The six axis labels likewise synthesise distinct strands of polarization research; axis assignment records the group-relational dimension

directly instantiated by a motif rather than every downstream outcome with which that motif may be empirically associated.

The final specification contains 35 canonical motifs: 14 DIVIDE, 12 COHERE, and 9 BRIDGE. Directions are non-exclusive at message level: a communication may instantiate motifs from more than one direction whenever their respective group-relational operations are independently recoverable.

J.4.1 Polarization direction provenance

Supplementary Table S52: Provenance of Polarization directions.

Direction	Operational relation	Literature basis and ontology interpretation	Key sources
DIVIDE	Constructs, sharpens, or reinforces a socially or politically salient boundary between groups, including differentiation of an outgroup from an ingroup or partitioning of a previously inclusive collective into opposing camps.	Operational synthesis of social categorisation, intergroup affect, partisan and ideological polarization, and group-boundary research. The direction records a boundary-producing or boundary-strengthening move in the message rather than inferring that population-level polarization actually increases.	(Tajfel and Turner, 1979; Iyengar et al., 2012; Mason, 2015; McCoy et al., 2018)
COHERE	Constructs, extends, or reinforces a shared collective "we" through common identity, belonging, solidarity, valued characteristics, commitments, continuity, or collective agency.	Operational synthesis of social-identity, self-categorisation, collective identification, group norms, solidarity, and collective-action literatures. The direction does not require an already salient outgroup or the crossing of an existing intergroup boundary.	(Tajfel and Turner, 1979; Hogg and Reid, 2006; Turner et al., 1989; Bandura, 2000)
BRIDGE	Crosses, softens, complicates, or supersedes a salient boundary between differentiated groups through common identity, contact, perspective, cross-cutting membership, shared goals, fair standing, common epistemic ground, or successful cross-group relations.	Operational synthesis of common-ingroup identity, intergroup contact, cross-categorisation, perspective-taking, cooperation, and procedural- inclusion research. Bridge requires a recoverable differentiated-group boundary that the message relationally crosses or softens; generic unity is therefore insufficient.	(Gaertner et al., 1993; Pettigrew and Tropp, 2006; Crisp and Hewstone, 1999; Sherif et al., 1961; Tyler, 2000)

J.4.2 Polarization-axis provenance

Supplementary Table S53: Conceptual provenance of Polarization axes.

Axis	Conceptual basis	Ontology interpretation	Key sources
Affective / Identity	Social identification, partisan attachment, affective polarization, group evaluation, hostility, favouritism, and social distance	Records group-relational positioning through collective belonging, identification, positive ingroup regard, dislike, distrust, contempt, hostility, or social distance. Purely substantive issue disagreement without group-based identification or evaluation is excluded.	(Iyengar et al., 2012; Iyengar and Westwood, 2015; Mason, 2015)
Issue / Ideological	Political attitudes, policy preferences, substantive issue positions, ideological commitments, values, and interests	Records convergence, divergence, opposition, or interdependence in substantive positions or goals across salient groups. Group dislike or categorical distance without issue or ideological positioning does not independently license the axis.	(Abramowitz and Saunders, 2008)
Sorting / Alignment	Partisan sorting, aligned identities, ideological sorting, cleavage reinforcement, and cross-cutting categorisation	Records alignment of identities, cleavages, issue positions, values, or norms into common camps, or the introduction of cross-cutting memberships that complicate such alignment. Solidarity or disagreement within a single already defined cleavage does not itself constitute sorting.	(Levendusky, 2009; Mason, 2015; Mason, 2016)
Perception / Distance	Perceived polarization, stereotyped group composition, extremity, homogeneity, consensus, and represented intergroup distance	Records how the message represents groups as internally homogeneous, extreme, consensual, compositionally similar or different, or more or less distant. The axis does not require that the represented group properties are factually accurate.	(Westfall et al., 2015; Levendusky and Malhotra, 2016; Ahler and Sood, 2018)
Network / Segregation	Homophily, social separation, selective exposure, intergroup contact, association, and relational integration	Records message-level orientation toward social contact, association, exposure, cooperation, avoidance, isolation, or separation across group boundaries. It does not infer the actual structure of a social network.	(Boutyline and Willer, 2017; Mepham et al., 2025)

Legitimacy / Tolerance	Political tolerance, moral inclusion, legitimate standing, procedural rights, participation, and entitlement to ordinary protections	Records expansion or contraction of who is represented as legitimately entitled to political, civic, procedural, or moral standing and tolerated participation. Mere dislike, hostility, or disagreement without normative standing is insufficient.	(Opotow, 1990; Svolik, 2019; Tilley et al., 2026)

J.4.3 Polarization-motif provenance

Supplementary Table S54: Conceptual provenance and deterministic axis assignment of canonical Polarization motifs.

Motif	Direction	Construct basis	Operational interpretation	Axis
Dehumanization	Divide	Dehumanization and denial of full humanness (Haslam, 2006; Kteily et al., 2015; Esses et al., 2013)	Human group members are represented as less than fully human through animalistic, objectifying, mechanistic, biological, disease, vermin, or contamination categories.	Affective/Identity
Existential threat	Divide	Intergroup threat and collective existential threat (Hirschberger et al., 2016; Stephan et al., 2009)	An outgroup is represented as threatening the continued physical existence or defining identity and continuity of an ingroup or valued collective.	Affective/Identity
Zero-sum	Divide	Zero-sum beliefs and intergroup resource competition (Esses et al., 1998; Różycka-Tran et al., 2015; Roberts and Davidai, 2022)	Group interests, resources, rights, opportunities, status, or power are represented as inversely related such that one group's gain entails another's loss.	Issue/Ideological
Scapegoating / concentrated blame	Divide	Scapegoating and displaced collective blame (Rothschild et al., 2012)	Responsibility for a broader negative collective condition is concentrated on a salient social or political group.	Affective/Identity
Political/procedural delegitimization	Divide	Political legitimacy, exclusion, and denial of valid political standing (Rivlin-Angert and Mor-Lan, 2025; Daur, 2025)	A group or its representatives are denied legitimate standing, authority, or entitlement to participate as valid actors in the relevant civic or political order.	Legitimacy/Tolerance
Shunning / segregation	Divide	Political avoidance, social separation, selective association, and relational segregation (Carlson and Settle, 2022; Peterson and Kagalwala, 2021)	Reduced contact, association, shared space, interaction, or exposure is advocated or normalised across a salient group boundary.	Network/Segregation
Extremity portrayal	Divide	Perceived partisan extremity and distorted group-composition perceptions (Westfall et al., 2015; Levendusky and Malhotra, 2016; Ahler and Sood, 2018)	Extreme or radical positions are represented as prevalent, typical, or defining of the salient outgroup.	Perception/Distance
Moral condemnation	Divide	Group morality judgements and moral-character evaluation (Leach et al., 2007; Brambilla et al., 2013)	A salient outgroup is represented as characteristically immoral, depraved, corrupt, degraded, or violative of fundamental moral standards.	Affective/Identity

Supplementary Table S54: Conceptual provenance and deterministic axis assignment of canonical Polarization motifs (continued)

Motif	Direction	Construct basis	Operational interpretation	Axis
Collective punishment	Divide	Collective responsibility and group-based punishment (Lickel et al., 2006; Pereira et al., 2015)	Punitive or adverse treatment is extended beyond individually responsible actors to wider group members on the basis of shared membership or attributed collective responsibility.	Legitimacy/Tolerance
Homogenization	Divide	Outgroup homogeneity and perceived within-group uniformity (Rubin and Badea, 2007; Rubin and Badea, 2012)	A salient outgroup is represented as a monolithic bloc whose members uniformly or near-uniformly share relevant beliefs, traits, motives, identities, or recurrent behaviour.	Perception/Distance
Criminalization / public-order threat	Divide	Criminalization, public-order threat, and group-based dangerousness (Brouwer et al., 2017; Gruber, 2023)	A salient outgroup is represented as characteristically criminal, lawless, violent, or dangerous to public safety or social order.	Affective/Identity
Conspiratorial attribution	Divide	Intergroup conspiracy attribution and covert hostile coalition (Prooijen and Douglas, 2018; Douglas et al., 2019)	A consequential event, condition, or harm is attributed to covert, coordinated, intentional action by a salient outgroup or coalition acting adversarially against a valued collective.	Affective/Identity
Identity-based ridicule / disparagement	Divide	Disparagement humour, group-directed ridicule, and identity-based contempt (Ford and Ferguson, 2004; Abrams and Bippus, 2011)	A salient outgroup or identity is mocked or derisively devalued because of group membership or characteristics represented as emblematic of that identity.	Affective/Identity
Membership denial	Divide	Identity denial, group prototypicality, authentic membership, and belonging (Cheryan and Monin, 2005; Kteily et al., 2014; Howard et al., 2023)	Authentic membership or belonging of a person, subgroup, or constituency in a salient collective "we" is denied, revoked, or challenged.	Affective/Identity
Symbolic / ritual affirmation	Cohere	Collective symbolism, ritual, identity affirmation, and group cohesion (Schatz and Lavine, 2007; Whitehouse and Lanman, 2014; Callahan and Ledgerwood, 2016)	Shared symbols or ritual practices are affirmatively invoked or enacted as expressions of common ingroup membership, belonging, or unity.	Affective/Identity
Collective nostalgia	Cohere	Collective nostalgia, identity continuity, and shared positive group memory (Wildschut et al., 2014; Smeekes et al., 2021)	A positively remembered ingroup past is used to affirm common identity, belonging, or continuity across time.	Affective/Identity
Achievement / collective pride	Cohere	Group-based pride and collective accomplishment (Smith and Mackie, 2016; Mackie and Smith, 2018)	An accomplishment or valued collective performance is affirmatively represented as grounds for common positive regard, celebration, or pride.	Affective/Identity
Service / sacrifice	Cohere	Self-sacrifice, helping, social identity, and contribution to collective good (Swann et al., 2010; Levine et al., 2005)	Contribution of effort, resources, comfort, interests, or safety toward a shared collective good is presented as service, commitment, sacrifice, or a marker of belonging.	Affective/Identity

Supplementary Table S54: Conceptual provenance and deterministic axis assignment of canonical Polarization motifs (continued)

Motif	Direction	Construct basis	Operational interpretation	Axis
Mutual support / solidarity	Cohere	Collective solidarity and mutual support under shared identity (Drury et al., 2009; Drury et al., 2016)	A shared collective "we" is affirmed through mutual care, assistance, or support among its represented or invited members.	Affective/Identity
Collective efficacy	Cohere	Collective efficacy and group-based capacity for joint action (Bandura, 2000; Zomeren et al., 2010)	The ingroup is represented as capable, through joint or coordinated action, of achieving a shared goal or collective effect.	Affective/Identity
Character valorization	Cohere	Social identity, collective self-esteem, and valued group characteristics (Tajfel and Turner, 1979; Luhtanen and Crocker, 1992)	Valued traits or dispositions are affirmatively attributed to the ingroup as sources of collective worth or identity.	Affective/Identity
Tradition / cultural continuity	Cohere	Collective continuity, inherited tradition, and temporal group identity (Sani et al., 2007; Smeekes and Verkuyten, 2013)	Inherited customs, practices, norms, or ceremonies are represented as linking the current ingroup with earlier generations or periods of the same group.	Affective/Identity
Consensus / normative alignment	Cohere	Self-categorisation, group norms, identity-based consensus, and normative alignment (Turner et al., 1989; Hogg and Reid, 2006; Reicher and Hopkins, 1996)	Shared agreement around a belief, value, norm, preference, goal, or position is used to construct or affirm collective identity or as a basis for belonging.	Affective/Identity; Sorting/Alignment; Perception/Distance
Norms / commitment pledges	Cohere	Group norms, commitment, identity enactment, and normative belonging (Hogg and Reid, 2006; Feldman, 1984; Ellemers et al., 1999)	Norms, rules, principles, oaths, or pledges are represented as standards or commitments constitutive of the shared collective "we".	Affective/Identity; Sorting/Alignment
Collective mourning / remembrance	Cohere	Collective grief, remembrance, shared loss, and identity affirmation (Wagoner and Brescó de Luna, 2021; Gur et al., 2024)	Shared mourning or remembrance of collective losses is used to construct, extend, or affirm common group identity.	Affective/Identity
Self-directed humor	Cohere	Affiliative humour, insider humour, collective familiarity, and shared identity (Flesher Fominaya, 2007; Moody, 2019)	Playful self-reference, self-mockery, teasing, or insider humour about a collective "we" is used to affirm common identity, familiarity, or inclusion.	Affective/Identity
Humanization	Bridge	Humanization, outgroup personhood, individuation, and attribution of human mental capacities (Haslam, 2006; Prati et al., 2016; Capozza et al., 2017)	A salient outgroup is represented through personhood, individuality, agency, emotion, mentality, or other properties that foreground members as fully human persons.	Affective/Identity
Superordinate / common identity	Bridge	Common-ingroup identity and recategorisation (Gaertner et al., 1993; Gaertner et al., 1996; Dovidio et al., 2000)	Members of differentiated groups are recast as belonging to a broader shared category or collective "we", while subgroup identities may remain salient.	Affective/Identity
Contact invitation	Bridge	Intergroup contact and active engagement across group boundaries (Pettigrew and Tropp, 2006; Esses and Dovidio, 2002; Meleady, 2021)	Constructive interpersonal contact, dialogue, listening, or interaction is invited or affirmatively endorsed across a salient differentiated-group boundary.	Network/Segregation

Supplementary Table S54: Conceptual provenance and deterministic axis assignment of canonical Polarization motifs (continued)

Motif	Direction	Construct basis	Operational interpretation	Axis
Perspective-taking	Bridge	Intergroup perspective-taking and reduction of intergroup bias (Galinsky and Moskowitz, 2000; Vescio et al., 2003; Todd et al., 2011)	Recipients are invited to consider an experience, disagreement, or situation from the subjective standpoint of members of a salient differentiated outgroup.	Affective/Identity
Cross-cutting identity	Bridge	Cross-categorisation, multiple identities, and cross-cutting group membership (Brown and Turner, 1979; Crisp and Hewstone, 1999; Crisp et al., 2001)	An identity dimension is made salient that intersects rather than reproduces the focal division, thereby complicating the categorical us–them structure.	Sorting/Alignment
Superordinate goal	Bridge	Superordinate goals, cooperative interdependence, and joint problem solving (Sherif et al., 1961; Gaertner et al., 1999)	Differentiated groups are represented as sharing a valued objective whose attainment requires joint or coordinated contribution.	Issue/Ideological; Network/Segregation
Procedural fairness	Bridge	Procedural justice, impartial standing, reciprocal treatment, and legitimate participation (Tyler, 1988; Tyler, 1994; Tyler, 2000)	Differentiated groups are represented as entitled to impartial, reciprocal, or consistently applied procedures, voice, consideration, participation, or standing.	Legitimacy/Tolerance
Epistemic common ground	Bridge	Shared reality, common beliefs, and cross-group epistemic agreement (Echterhoff et al., 2009; Stern and Ondish, 2018; Tessler et al., 2024)	Differentiated groups are represented as sharing or accepting a factual claim, evidence base, premise, observation, or other epistemic reference point despite broader disagreement.	Perception/Distance
Joint success exemplars	Bridge	Extended/vicarious contact, positive intergroup exemplars, and successful cross-group interaction (Wright et al., 1997; Mazziotta et al., 2011; Vezzali et al., 2019)	A successful instance of cross-group contact, friendship, cooperation, problem-solving, or joint action is represented as producing or sustaining a valued outcome.	Network/Segregation

Polarization cross-layer boundary decisions. Polarization motifs may share actors, claims, evaluations, metaphors, or other surface evidence with Method Tactics, Narrative structures, TopicDomains, and AboutObjects. Such overlap does not create automatic cross-layer inheritance.

Supplementary Table S55: Principal Polarization cross-layer boundary decisions.

Polarization construct / overlap	Independent construct	Boundary rule
Polarization generally	Population-level polarization	The dimension records group-relational positioning performed within the message. It does not infer that social, political, affective, ideological, or network polarization in a population actually increases or decreases.

Supplementary Table S55: Principal Polarization cross-layer boundary decisions (continued)

Polarization construct / overlap	Independent construct	Boundary rule
Divide	AboutObject *Polarization-Increase*	A Divide motif records a message-internal boundary-strengthening operation. The AboutObject instead records a broader substantive outcome toward increased polarization where such an outcome is independently recoverable from the message. A single Divide cue does not automatically license that AboutObject.
Cohere	AboutObject *SocietalCohesion-Increase* or *CollectiveIdentity-Strengthening*	Cohere records construction or reinforcement of a shared collective "we". A broader cohesion or collective-identity outcome is coded only where the message independently bears toward that substantive endpoint. The relational operation and the AboutObject may overlap semantically without being the same construct.
Bridge	AboutObject *Polarization-Decrease*, *SocietalCohesion-Increase*, or related intergroup outcome	Bridge records a message-internal operation crossing or softening a salient differentiated-group boundary. It does not automatically imply a broader outcome toward depolarization, societal cohesion, inclusion, or reduced hostility; any such AboutObject must independently satisfy its own recoverability rule.
Cohere	Bridge	Constructing or strengthening a collective "we" does not require a prior differentiated-group boundary and therefore licenses Cohere without Bridge. Bridge specifically requires crossing, softening, complicating, or superseding such a boundary. Both may co-occur where a broader shared identity is simultaneously constructed and an existing subgroup division is reduced.
Divide	Cohere	Strengthening solidarity within an ingroup and sharpening distinction from an outgroup are separate relational operations. A message may perform either or both; ingroup affirmation does not mechanically imply outgroup division.
Dehumanization	*Dehumanizing representation* / Framing	The Polarization motif requires dehumanizing representation to function within a salient social or political intergroup relation. The Method Tactic records the representational operation itself and does not require a polarization direction. Neither automatically licenses the other.

Supplementary Table S55: Principal Polarization cross-layer boundary decisions (continued)

Polarization construct / overlap	**Independent construct**	**Boundary rule**
Scapegoating / concentrated blame	*Scapegoating* / Framing; Narrative Causal emplotment	The Polarization motif requires concentrated blame directed at a salient group within an intergroup boundary. The Framing Tactic requires concentrated blame on an identifiable person or group as the framing operation, while Narrative Causal emplotment merely records represented causation. Co-code only where the corresponding conditions are independently satisfied.
Existential threat	*Threat framing* / Framing; *Appeal to fear* / Persuasion	The Polarization motif requires an outgroup to threaten the continued physical existence or defining identity of an ingroup or collective. Threat framing requires a recoverable danger relation but not an intergroup existential relation. Appeal to fear additionally requires threatened harm to function as a reason for a protective response.
Identity-based ridicule / disparagement	*Ridicule / mockery* / Framing	The Polarization motif requires a salient outgroup identity itself to be derisively devalued. Ridicule / mockery can target an individual, institution, policy, idea, or action without any intergroup positioning.
Moral condemnation	*Moral framing* / Framing; Narrative Moral valuation	Polarization Moral condemnation requires a salient outgroup to be characterised in moral terms as part of a group-relational positioning move. Moral framing and Narrative moral valuation have their own independent interpretive and story-structuring criteria.
Conspiratorial attribution	Narrative *Conspiracy / hidden hand*; AboutObject *ConspiratorialWorldview-Increase/Decrease*	The Polarization motif requires covert coordinated causation by a salient outgroup or coalition positioned adversarially toward a collective. A conspiracy NarrativeArchetype requires its narrative causal signature but not necessarily intergroup polarization. A conspiratorial-worldview AboutObject records a broader substantive outcome and is not inherited from either representation.
Membership denial	Narrative Betrayal / internal-traitor role	Narrative betrayal or traitor casting does not automatically deny authentic membership. Polarization Membership denial specifically revokes, challenges, or rejects the target's genuine belonging within the salient collective "we".

Supplementary Table S55: Principal Polarization cross-layer boundary decisions (continued)

Polarization construct / overlap	Independent construct	Boundary rule
Symbolic / ritual affirmation	*Unity (shared identity)* / Persuasion; Narrative Symbolic condensation	A symbol or ritual may condense narrative meaning, affirm group membership, or function as an identity-based persuasion warrant. The Cohere motif requires affirmative use of the symbol or ritual to construct or reinforce the shared collective "we"; the other layers retain their independent minimum conditions.
Collective nostalgia	*Collective nostalgic framing* / Framing; *Appeal to tradition* / Persuasion	Cohere Collective nostalgia requires a positively remembered ingroup past to affirm common identity or continuity. Collective nostalgic framing uses the shared past as a temporal interpretive frame, while Appeal to tradition uses longstanding or customary status as a persuasive warrant. The three may co-occur but are not interchangeable.
Consensus / normative alignment	*Social proof* / Persuasion; *Unity* / Persuasion	The Cohere motif requires represented agreement to function as a marker or basis of collective identity. Social proof uses others' beliefs, approval, choices, or behaviour as social evidence, while Unity uses shared collective membership as the influence warrant. Agreement itself does not mechanically instantiate all three.
Contact invitation	Behavioural request or generic interaction	Bridge Contact invitation requires constructive interaction specifically across a salient differentiated-group boundary. Generic invitations to talk, meet, or engage without such an intergroup relation do not instantiate the motif.
Superordinate / common identity	Unity / Cohere	A broad shared "we" can instantiate Cohere without Bridge where no differentiated groups are being recategorised. Bridge Superordinate/common identity requires previously differentiated or simultaneously salient groups to be represented as members of a broader encompassing collective.
Epistemic common ground	Consensus / normative alignment; Social proof	Bridge Epistemic common ground requires differentiated groups to share a factual proposition, evidence base, premise, observation, or other epistemic reference point across their boundary. Ingroup consensus or popularity does not satisfy that cross-group epistemic relation.

Polarization verification conditions. The Polarization specification is internally reproducible if all of the following conditions hold:

1. the canonical motif inventory contains exactly 35 motifs: 14 Divide, 12 Cohere, and 9 Bridge;

2. every motif belongs to exactly one of the three relational directions;
3. every motif has a fixed, non-empty mapping to one or more of the six Polarization axes;
4. the machine-readable motif–direction–axis mapping exactly reproduces Table S13;
5. multiple directions and motifs may be returned for the same message whenever their respective operations are independently evidenced;
6. group categories are interpreted relative to their representation in the message rather than assumed from external demographic or political classifications;
7. no Method Tactic, NarrativeDevice, TopicDomain, or AboutObject is inferred solely from the presence of a Polarization motif, and vice versa; and
8. no population-level polarization effect is inferred from the message-level annotation.

Polarization motif assignment requires message-level interpretation and its empirical annotatability is not established by the present specification analysis. Future validation work should therefore examine independent agreement for motif assignment and difficult direction or motif boundaries. Because axis values are inherited deterministically from motif assignment, separate human annotation of the axis is unnecessary if the implementation correctly applies the fixed lookup relation.

J.5 TopicDomain provenance

TopicDomain records the substantive issue area addressed by a communication. It is independent of Method/Tactic, TargetConstruct, Narrative, Polarization, and AboutObject. A TopicDomain identifies what issue the message concerns; it does not specify how influence is attempted, what audience-state dimension is targeted, how groups are positioned, how the message is narratively organised, or what broader substantive outcome the message bears toward.

The canonical TopicDomain inventory is adapted from established general- purpose policy and media classification systems rather than derived from a single source taxonomy. Its purpose is operational issue classification for the kinds of civic, political, social, security, health, technological, and commercial communications covered by the ontology. The resulting categories therefore preserve broad comparability with established taxonomies while allowing issue areas that are analytically salient for contemporary communication research to remain separately identifiable.

Supplementary Table S56: Source-taxonomy provenance of TopicDomain.

Source / design input	Role in ontology construction	Ontology treatment
EuroVoc	General cross-domain policy vocabulary	Provides broad substantive coverage across governmental, social, economic, international, environmental, scientific, and related policy areas. The ontology uses EuroVoc as one coverage reference rather than reproducing its hierarchy or vocabulary verbatim (Publications Office of the European Union, n.d.).
Comparative Agendas Project	Comparative policy-agenda major-topic classification	Provides a cross-national policy-agenda reference for broad issue-area partitioning. The ontology draws on its major-topic logic while retaining a smaller message-oriented controlled vocabulary suitable for annotation (Baumgartner and Jones, 2009).
IPTC Media Topics	General-purpose news and media subject classification	Provides a media-oriented reference for classifying the substantive subjects of communications. Its role is complementary to policy taxonomies because the ontology is applied to messages rather than only to formal policy agendas (Kuzman and Ljubešić, 2025).
Contemporary communication-research salience	Operational adaptation criterion	Issue areas that are frequently analytically distinct in contemporary political and influence communication are retained as separate TopicDomains where collapsing them into broader source-taxonomy categories would obscure substantively important distinctions. The final inventory therefore represents an adaptation across source taxonomies rather than a direct union or one-to-one mapping.

TopicDomain adaptation decisions.

Supplementary Table S57: Principal TopicDomain adaptation decisions.

Candidate issue area	Treatment	Rationale
Race / ethnicity	KEEP: distinct domain	Race- and ethnicity-related communication is kept independently identifiable rather than folded into a broader civil-rights or social-affairs category because discrimination, racialised intergroup positioning, policing, reparations, and related debates form a recurrent substantive issue area in the intended communication domain.
Abortion / reproductive rights	KEEP: distinct domain	Reproductive-rights communication is kept distinct rather than absorbed into general health, family, or social-policy categories so that abortion, contraception, IVF, and related reproductive-policy debates remain independently classifiable.

Supplementary Table S57: Principal TopicDomain adaptation decisions (continued)

Candidate issue area	Treatment	Rationale
Public-health mandates	SPLIT FROM GENERAL HEALTH	Mandate-oriented communication concerning vaccination, masking, lockdowns, and comparable collective public-health requirements is separated from general health and healthcare communication because the former often concerns public regulation and population-level protective measures rather than individual healthcare or wellness alone.
Gender & sexuality	KEEP: distinct domain	Gender identity, LGBTQ rights, sex education, and related policy debates are kept independently identifiable rather than subsumed under broader social, civil-rights, family, or health categories.
Culturally or jurisdictionally variable issue salience	BOUNDARY: application-level adaptation permitted	Which topics become politically salient varies across jurisdictions and historical periods. The controlled vocabulary provides stable broad issue areas, while application to a particular corpus may require culturally appropriate examples, coder guidance, or subtopic exemplars. Such adaptation does not alter the meaning of TopicDomain as the message's substantive issue area (McCombs and Shaw, 1972; Baumgartner and Jones, 2009).

Canonical inventory audit. The final TopicDomain vocabulary contains 23 canonical categories:

1. Elections & Governance
2. Immigration
3. Race/Ethnicity
4. Religion
5. Abortion/Reproductive Rights
6. Guns/Crime/Policing
7. Environment & Climate
8. Public-Health Mandates
9. Gender & Sexuality
10. National Security/War
11. Family & Children
12. Culture & Media
13. Health (general)
14. Economy & Taxation
15. Education
16. Energy & Utilities
17. Housing & Urban Affairs
18. Labor & Employment
19. Science & Technology
20. Foreign Affairs (non-war)
21. Consumer/Marketplace
22. Transportation
23. Disasters & Emergencies

The authoritative definitions and illustrative subtopics remain those in Table S14; this inventory is included only to make vocabulary closure and machine-readable specification checks explicit.

TopicDomain cross-layer boundaries.

Supplementary Table S58: Principal TopicDomain cross-layer boundary decisions.

Boundary	Independent construct	Decision rule
Subject matter versus influence operation	Method / Tactic	TopicDomain records what substantive issue is being discussed. It does not determine how the message attempts influence. The same Tactic can occur in many TopicDomains, and the same TopicDomain can contain any independently evidenced Tactic.
Subject matter versus audience-state target	TargetConstruct	A health, election, security, commercial, or other topic does not determine whether a Tactic constitutively bears on Belief, Attitude, Identity, Agency/Choice, Behavior, or Perception/Attention. TargetConstruct is inherited from Tactic, not from TopicDomain.
Subject matter versus narrative organisation	Narrative	A message about war, elections, health, immigration, commerce, or another domain may or may not instantiate narrative devices or archetypes. Topic content does not determine story structure.
Subject matter versus group positioning	Polarization	Topics frequently associated with political or intergroup conflict do not automatically instantiate Divide, Cohere, or Bridge. Polarization requires a recoverable group-relational operation independent of the topic label.
Subject matter versus substantive outcome	AboutObject	TopicDomain answers *what issue is the message about*; AboutObject answers *what broader substantive outcome the message bears toward*. The relation is many-to-many. For example, climate-related content may bear toward climate acceptance, climate-policy support, political-actor support, institutional trust, purchase behaviour, or another independently recoverable AboutObject. TopicDomain therefore never determines AboutObject by lookup.
Subtopic versus additional TopicDomain	TopicDomain itself	An illustrative subtopic does not create a second domain merely because it can occur within another policy context. Code additional TopicDomains only where the message substantively addresses those issue areas rather than because of a plausible association or downstream consequence.

TopicDomain verification conditions. The TopicDomain specification is internally reproducible if:

1. the canonical vocabulary contains exactly the 23 domains listed in Table S14;
2. machine-readable labels reproduce those canonical names exactly;
3. illustrative subtopics are treated as coder guidance rather than additional canonical labels;
4. TopicDomain assignment does not mechanically determine a Method, TargetConstruct, Narrative, Polarization, or AboutObject label; and
5. culturally or jurisdictionally adapted examples preserve the substantive meaning and boundaries of the canonical domain.

TopicDomain requires message-level classification and its empirical annotatability is not established by the present specification analysis. Future validation work may therefore assess independent agreement for TopicDomain assignment, particularly for messages spanning multiple issue areas and for boundaries between neighbouring domains. Such empirical testing is not part of the present study.

J.6 AboutObject provenance and vocabulary derivation

`AboutObject` records the broader substantive civic or commercial outcome toward which a communication observably bears. It is not an empirical audience effect, a statement of the communicator's private intention, a TopicDomain, a TargetConstruct, or a synonym for the communicative operation represented by method/Tactic.

The vocabulary was constructed by identifying substantively distinct outcome objects supported by the literature and framework-coverage review, then applying a common set of inclusion and consolidation rules. A candidate was retained where it represented a durable civic or commercial behaviour, stance, relation, condition, institutional state, epistemic state, or resilience capacity that could itself constitute a recoverable logical outcome of a message. Candidates were consolidated where they represented the same logical endpoint, excluded where they represented only mechanisms or intermediate processing states, and separated where superficially related outcomes could vary independently.

The resulting specification is hierarchical. Each canonical L3 `aboutObjectLabel` belongs to exactly one L2 outcome domain. L1 sphere tags are derived deterministically from those L2 assignments: domains A–K instantiate `CivicOutcome`, domain L instantiates `CommercialOutcome`, and messages containing independently satisfied labels from both sets instantiate both sphere tags. `Unclear` carries no sphere assignment.

J.6.1 AboutObject construct provenance

Supplementary Table S59: Principal construct-design decisions underlying AboutObject.

Design construct	Role in the ontology	Inclusion / boundary rule	Sources
Logical outcome rather than empirical effect	Defines the level of abstraction represented by AboutObject	Code an outcome that the communication logically advances, opposes, strengthens, weakens, maintains, prevents, or otherwise bears toward from message-internal evidence. Do not require observed audience change or infer a private campaign objective.	(Kavanagh and Rich, 2018; Marwick and Lewis, 2017)
Independence from Method and TargetConstruct	Separates outcome from influence mechanism and audience-state target	Method specifies how influence is performed; TargetConstruct records the audience-state dimension constitutively associated with a Tactic; AboutObject records the broader substantive endpoint. Semantic overlap between these dimensions does not make them redundant.	(Kavanagh and Rich, 2018; Marwick and Lewis, 2017)
Independence from TopicDomain	Separates subject matter from influence object	TopicDomain answers what issue the message addresses; AboutObject answers what broader civic or commercial outcome it bears toward. The relation is many-to-many and no TopicDomain mechanically determines an AboutObject.	(Kavanagh and Rich, 2018; Marwick and Lewis, 2017)
Multiple-outcome rule	Permits genuine multilabel outcome annotation	Multiple AboutObjects may be assigned where the message independently bears toward multiple substantive outcomes. The fact that one outcome serves as a premise, means, or intermediate condition for another does not prevent both from being coded where both are independently recoverable.	(Kavanagh and Rich, 2018; Marwick and Lewis, 2017)
Same-endpoint / specificity rule	Prevents hierarchical double counting	Where a generic and a more specific candidate describe the same logical endpoint, instantiate only the most specific satisfied label. A broader label co-occurs only where an additional broader outcome is independently advanced.	—

Supplementary Table S59: Principal AboutObject construct-design decisions (continued)

Design construct	Role in the ontology	Inclusion / boundary rule	Sources
Causal-chain rule	Prevents inference of unstated downstream outcomes	Do not infer an AboutObject merely because it is a plausible consequence of another coded outcome. Trust loss does not automatically imply information avoidance; policy support does not automatically imply compliance; confidence loss does not automatically imply run participation.	—
Substantive-object criterion	Defines what kinds of states may legitimately be AboutObjects	Behaviors, participation patterns, policy stances, social relations, trust relations, legitimacy relations, epistemic conditions, institutional conditions, and resilience capacities may be retained where alteration of the condition itself is the recoverable broader outcome.	—
Processing-state exclusion	Separates outcomes from transient mechanisms or intermediate states	Generic confusion, overload, indecision, paralysis, and comparable processing states are not standalone AboutObjects merely because they may mediate influence or contribute causally to a later substantive outcome.	—
Directional neutrality	Prevents normative interpretation of label polarity	Increase and decrease labels indicate outcome direction only. Increased trust, compliance, cohesion, support, or engagement is not inherently benign, and decrease labels are not inherently malign. Method and normative evaluation remain independent.	—
Resilience as substantive capacity	Distinguishes transferable capacity from a single successful response	Information, climate, and infrastructure resilience are retained where the communication bears toward broader capacity to recognise, withstand, adapt to, resist, recover from, or continue functioning under relevant challenges. A single defensive act or successful response is insufficient.	(Roozenbeek et al., 2022a; OECD, 2019)

Supplementary Table S59: Principal AboutObject construct-design decisions (continued)

Design construct	Role in the ontology	Inclusion / boundary rule	Sources
Evidence templates as indicators	Separates observable evidence from the outcome labels themselves	L4 templates are families of message-internal indicators that can justify an L3 assignment. They are not additional outcomes and are not lexical trigger lists. The directional L3 relation must still be independently evidenced.	—

J.6.2 Outcome-domain provenance and inventory closure

Supplementary Table S60: AboutObject domain provenance and canonical L3 inventory counts.

L2	Outcome domain	L3	Construct coverage	Key sources
A	Politics	18	Voter registration and turnout; political-actor support; generic policy stance and compliance; protest and broader civic engagement; election and political-system legitimacy.	(Kim et al., 2018; Brennen et al., 2020; Kavanagh and Rich, 2018; Anderson and Just, 2013; Dowling and Legrand, 2023)
B	Identity, Intergroup Relations & Social Cohesion	13	Collective-identity strengthening, weakening, and reconfiguration; polarization; outgroup hostility; societal cohesion; group inclusion; and group humanization.	(Kavanagh and Rich, 2018; EUvsDisinfo, n.d.; United Nations, 2019)
C	Security & Violence	10	Political-violence support and participation; terrorism support; violent-extremism support; and violent-extremist behavioural engagement.	(Fortify Rights, 2018; United Nations, 2019; McCauley and Moskalenko, 2017)
D	Epistemic Trust, Legitimacy & Attribution	11	Media and science trust; institutional trust and legitimacy; generalized epistemic trust; and responsibility reassignment.	(Ognyanova et al., 2020; Lazer et al., 2018; Kavanagh and Rich, 2018; Diel et al., 2024)
E	Ideological Worldview & Community Engagement	6	Generalized conspiratorial and anti-pluralist worldviews and behavioural engagement with communities organised around such worldviews.	(EUvsDisinfo, n.d.; Malcolm, 2023; Medzihorsky and Lindberg, 2024)

Supplementary Table S60: AboutObject domain provenance and inventory counts (continued)

L2	Outcome domain	L3	Construct coverage	Key sources
F	Public Health & Civil Protection	10	Health-guidance adherence; health-care engagement; substitute-remedy adoption; crisis protective action; and emergency-service capacity.	(Nascimento et al., 2022; Hilberts et al., 2025)
G	Environment, Climate Action & Resilience	12	Environmental-protection behaviour; climate-change acceptance; climate-policy stance; mitigation and adaptation action; and climate resilience.	(Lewandowsky, 2021; Chatterjee et al., 2024)
H	Economic, Financial & Infrastructure Stability	12	Economic and financial-system confidence; essential-resource stockpiling; deposit-run participation; and critical-infrastructure confidence and resilience.	(Ali et al., 2023; Sandri et al., 2023; OECD, 2019)
I	Information & Influence Resilience	4	Transferable information-resilience capacity, increased information avoidance, and increased information seeking.	(Roozenbeek et al., 2022a; Hicks et al., 2025)
J	Geopolitics & Conflict Alignment	14	International military-force support; military-service participation; foreign assistance; sanctions; conflict-party alignment; alliance support; and confidence in alliance reliability.	(Krishnarajan and Tolstrup, 2023; Geissler et al., 2023; Tomz and Weeks, 2021; Pretorius, 2019)
K	Personal Security, Privacy & Public Expression	6	Personal privacy, personal security, and participation in or suppression of public expression.	(Chen et al., 2019; Eckert and Metzger-Riftkin, 2020; Björkenfeldt, 2026)
L	Commercial Acquisition, Relationship & Advocacy	24	Commercial inquiry and purchase; subscription, application, and trial enrollment; customer retention and reactivation; brand trust and loyalty; recommendation, reviews, and public brand advocacy.	(Lahtinen et al., 2020; Gordon, 2012; Buttle, 1998; Friedman, 1991; Chaudhuri and Holbrook, 2001)
	Domain-assigned L3 total	**140**	**116 CivicOutcome labels (A–K); 24 CommercialOutcome labels (L).**	

The closed L3 vocabulary additionally contains the fallback `Unclear`. It is not assigned to an L2 domain or L1 sphere and is used only where a broader civic or commercial outcome orientation is recoverable but the available evidence does not support selection of a specific canonical L3 label. Where no AboutObject is recoverable, the layer remains empty.

AboutObject decision ledger.

Supplementary Table S61: Principal AboutObject inclusion, consolidation, and boundary decisions.

Candidateconstruct/overlap	Status/qualifier	Canonicalrepresentation	Rationale
Generic call to action / compliance outcome	SPLIT: by substantive endpoint	Applicable domain-specific behavioural AboutObject	Behavioural influence is not represented through one generic call-to-action or compliance outcome. Voting, protest, policy compliance, health adherence, climate action, violent participation, military service, purchasing, and other behaviours remain distinct because they constitute different substantive civic or commercial endpoints.
Topic-specific issue material	BOUNDARY: do not map by topic	Applicable independently evidenced AboutObject	TopicDomain and AboutObject are many-to-many. Immigration, health, climate, security, or commercial content can be used in pursuit of electoral, intergroup, trust, policy, expression, purchase, or other outcomes. The topic therefore never supplies the outcome label by itself.
Generic versus specific policy stance	BOUNDARY: specificity rule	Domain-specific policy label where available; otherwise `Policy-Support` or `Policy-Opposition`	Climate policy, foreign aid, sanctions, and other domain-specific policy objects are not duplicated under the generic Domain A policy label where both names describe the same endpoint. The generic label remains available for policy outcomes without a more specific canonical object.
Generic policy compliance versus specific behavioural outcome	BOUNDARY: specificity rule	Most-specific behavioural AboutObject	Generic policy compliance is not duplicated where the same behaviour is more precisely represented by health-guidance adherence, crisis protective action, climate action, military-service participation, or another specific AboutObject.
Specific civic act versus broad civic engagement	BOUNDARY: specificity rule	Voter registration, turnout, protest participation, or `CivicEngagement-*` as independently warranted	One specific act of voting, registration, or protest does not automatically instantiate generic civic engagement. The broad label is used only where participation or withdrawal across civic channels is itself the outcome.

Supplementary Table S61: Principal AboutObject decisions (continued)

Candidateconstruct/overlap	Status/qualifier	Canonicalrepresentation	Rationale
Political-violence support versus terrorism / violent-extremism support	BOUNDARY: specificity rule	Most-specific support object	The generic political-violence object is not duplicated where the identical support endpoint is already represented as terrorism or violent-extremism support. Broader political-violence support may additionally be coded only when independently advanced.
Attitudinal support versus behavioural participation	SPLIT	Support and participation / engagement labels	Approval, justification, legitimacy, or endorsement of violence or violent extremism is distinct from joining, performing, contributing to, or withdrawing from the corresponding activity. Neither is inherited from the other.
Trust versus legitimacy	SPLIT	`InstitutionalTrust-*`; `InstitutionalLegitimacy-*`	Trust concerns reliability, honesty, competence, or procedural dependability. Legitimacy concerns rightful authority, standing, or entitlement to exercise an institutional role. Both are coded only where both relations are independently recoverable.
Source-specific trust versus generalised epistemic trust	SPLIT	`MediaTrust-*`; `ScienceTrust-*`; `InstitutionalTrust-*`; `GeneralizedEpistemicTrust-*`	Distrust of one source or institution does not automatically imply that reliable knowledge or dependable methods of distinguishing warranted from unwarranted information are generally unavailable. The generalised object requires broader epistemic collapse or restoration.
Specific conspiracy narrative versus conspiratorial worldview	SPLIT: across layers / SCALES	Narrative *Conspiracy / hidden hand*; `ConspiratorialWorldview-*`	A single hidden-plot explanation is a narrative structure. An AboutObject is added only where strengthening or weakening a generalised conspiratorial interpretation of consequential social or political reality is independently recoverable.

Supplementary Table S61: Principal AboutObject decisions (continued)

Candidateconstruct/overlap	Status/qualifier	Canonicalrepresentation	Rationale
Polarization move versus polarization outcome	SPLIT: across layers	Polarization motif; Polarization-Increase/Decrease	A Divide or Bridge annotation records the group-relational move enacted by the message. The AboutObject requires increased or decreased wider intergroup division itself to be an independently recoverable broader outcome.
Ingroup cohesion versus societal cohesion	SPLIT: by social scope	Cohere / collective-identity objects; SocietalCohesion-*	Stronger solidarity inside one ingroup does not establish greater cohesion of the broader society. Ingroup strengthening may coexist with deterioration of societal cohesion where other constituent groups are excluded or antagonised.
Climate-change acceptance versus science trust	SPLIT	ClimateChangeAcceptance-*; ScienceTrust-*	Acceptance of represented climate propositions and trust in scientific actors or institutions can vary independently and are therefore separate outcome objects.
Climate action versus climate resilience	SPLIT: by endpoint level	ClimateMitigationAction-*; ClimateAdaptationAction-*; ClimateResilience-*	A specific mitigation or adaptation behaviour is not equivalent to the broader capacity of a social, institutional, economic, or ecological system to cope, recover, adapt, learn, or transform under climate disturbance.
Infrastructure confidence versus resilience	SPLIT	CriticalInfrastructureConfidence-*; CriticalInfrastructureResilience-*	Confidence concerns represented dependability or continuity. Resilience concerns underlying system capacity to withstand, absorb, adapt to, recover from, or maintain essential functioning during disruption.
Information seeking versus information avoidance	BOUNDARY: do not force binary opposites	InformationSeeking-Increase; InformationAvoidance-Increase	Seeking and avoidance can coexist across different information sources or classes and therefore are not represented as a single bipolar continuum. The vocabulary deliberately does not infer decreased avoidance from seeking or decreased seeking from avoidance.

Supplementary Table S61: Principal AboutObject decisions (continued)

Candidateconstruct/overlap	Status/qualifier	Canonicalrepresentation	Rationale
Information practice versus information resilience	SPLIT	InformationSeeking-Increase; InformationAvoidance-Increase; InformationResilience-*	One verification, seeking, or avoidance act does not establish transferable capacity to recognise, evaluate, resist, or recover from subsequent influence. Resilience requires that broader capacity to be independently targeted.
Threat, blackmail, doxxing, intimidation, or coercion	BOUNDARY: Method/evidence, not generic outcome	Applicable substantive AboutObject	These operations may be means by which privacy, security, expression, voting, commercial behaviour, policy stance, or another outcome is pursued. Their presence does not itself supply an AboutObject.
Disclosure threat versus privacy loss	BOUNDARY: distinguish	PersonalPrivacy-Decrease only where privacy loss itself is an independent outcome	A threatened disclosure used solely as leverage for another response does not automatically mean that reducing privacy is itself the broader outcome.
Commercial boycott	MERGE: decomposed by market endpoint	PurchaseBehavior-Decrease, CustomerRetention-Decrease, or another applicable commercial label	Boycott is a mobilising form rather than a unique commercial outcome. The ontology records the substantive market relationship being reduced.
Brand trust, loyalty, and retention	SPLIT	BrandTrust-*; BrandLoyalty-*; CustomerRetention-*	Trust concerns reliability and promise fulfilment; loyalty concerns durable preferential commitment; retention concerns continuation of an existing customer relationship. These outcomes may causally influence one another but are not interchangeable (Chaudhuri and Holbrook, 2001).
Recommendation, review, and advocacy	SPLIT	BrandRecommendation-*; ReviewGeneration-*; BrandAdvocacy-*	Interpersonal recommendation, production of public ratings or reviews, and broader sustained public defence or promotion are distinct communication outcomes and are retained separately.

Confusion, overload, indecision, or paralysis	REJECT: standalone AboutObjects	Applicable substantive outcome, if any	These are generic processing states or intermediate conditions. They are not retained merely because they could mediate another influence outcome. A substantive civic or commercial object must itself be recoverable.

AboutObject cross-layer boundaries.

Supplementary Table S62: Global AboutObject cross-layer boundary rules.

Other dimension	What it records	Why it does not determine AboutObject
Method / Tactic	Observable communicative influence operation	The same Tactic can pursue different substantive outcomes, and the same AboutObject can be pursued through different Methods.
TargetConstruct	Audience-state dimension constitutively associated with a Tactic	Belief, Attitude, Identity, Agency/Choice, Behavior, or P/A does not specify which civic or commercial object the communication bears toward.
Narrative	Message-level organisation of actors, events, causes, symbols, values, and stakes	A narrative may provide evidence for an outcome but its story structure does not determine that outcome.
Polarization	Message-internal group-positioning operation	Divide, Cohere, and Bridge describe relational moves, not broader societal outcomes. Intergroup AboutObjects require their wider condition to be independently recoverable.
TopicDomain	Substantive issue area addressed by the communication	Topic and outcome are many-to-many; neither determines the other.
L4 evidence template	Observable cue family supporting L3 selection	An evidence template is an indicator, not an outcome label. Only the L3 directional relation independently evidenced by the message is instantiated.

AboutObject verification conditions. The AboutObject specification is internally reproducible if all of the following conditions hold:

1. the canonical L2 vocabulary contains exactly 12 domains, A–L;
2. the canonical L3 vocabulary contains exactly 140 domain-assigned labels plus the non-domain fallback `Unclear`;
3. every specific L3 label belongs to exactly one L2 domain;
4. L1 sphere tags are derived deterministically from instantiated L3 domains, with A–K yielding `CivicOutcome` and L yielding `CommercialOutcome`;
5. the same-endpoint rule prevents simultaneous coding of generic and specific labels where they describe the identical logical endpoint;

6. multiple L3 labels remain permissible where independently recoverable distinct outcomes are present;
7. no L3 label is inferred solely from a Method Tactic, TargetConstruct, Narrative label, Polarization motif, TopicDomain, or plausible causal consequence;
8. each canonical L3 label has at least one corresponding L4 evidence template or otherwise explicit message-internal selection guidance;
9. `Unclear` is used only where an AboutObject orientation is recoverable but specific classification is not; absence of an AboutObject leaves the layer empty; and
10. the machine-readable specification reproduces the canonical L1–L2–L3 assignments, names, directional distinctions, and selection constraints specified in Appendix F.

AboutObject assignment requires message-level interpretation. Its empirical annotatability is not established by the present specification analysis. Future validation work should therefore evaluate independent annotation agreement, with particular attention to multilabel cases, same-endpoint specificity decisions, and boundaries between related substantive outcomes. Such empirical testing is outside the scope of the present study.

References

Abramowitz AI, Saunders KL (2008) Is polarization a myth? *The Journal of Politics* 70(2): 542–555 https://doi.org/10.1017/S0022381608080493

Abrams JR, Bippus A (2011) An Intergroup Investigation of Disparaging Humor. *Journal of Language and Social Psychology* 30(2): 193–201 https://doi.org/10.1177/0261927X10397162

Abramson K (2014) Turning Up the Lights on Gaslighting. *Philosophical Perspectives* 28(1): 1–30 https://doi.org/10.1111/phpe.12046

Ahler DJ, Sood G (2018) The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics* 80(3): 964–981 https://doi.org/10.1086/697253

Aikin S, Casey J (2018) Straw Man. In: Arp R, Barbone S, Bruce M (eds) *Bad Arguments: 100 of the Most Important Fallacies in Western Philosophy*. Wiley, pp 223–226. https://doi.org/10.1002/9781119165811.ch47

Aikin SF (2025) Three Puzzles with Ad Hominem Arguments. *Journal of the American Philosophical Association* 11(3): 622–639 https://doi.org/10.1017/apa.2025.3

Aikin SF, Casey J (2026) What about Whataboutism? *Social Epistemology* 40(3): 329–338 https://doi.org/10.1080/02691728.2024.2343729

Ajzen I (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2): 179–211 https://doi.org/10.1016/0749-5978(91)90020-T

Ali M, Gomes LM, Azab N et al (2023) Panic buying and fake news in urban vs. rural England: A case study of Twitter during COVID-19. *Technological Forecasting and Social Change* 193: 122598 https://doi.org/10.1016/j.techfore.2023.122598

Allan K, Burridge K (1991) *Euphemism & Dysphemism: Language Used as Shield and Weapon*. Oxford University Press, New York

Allen M (1991) Meta-Analysis Comparing the Persuasiveness of One-Sided and Two-Sided Messages. *Western Journal of Speech Communication* 55(4): 390–404 https://doi.org/10.1080/10570319109374395

Allen M, Preiss RW (1997) Comparing the Persuasiveness of Narrative and Statistical Evidence Using Meta-Analysis. *Communication Research Reports* 14(2): 125–131 https://doi.org/10.1080/08824099709388654

Anderson CJ, Just A (2013) Legitimacy from Above: The Partisan Foundations of Support for the Political System in Democracies. *European Political Science Review* 5(3): 335–362 https://doi.org/10.1017/S1755773912000082

Arabi S (2023) Narcissistic and Psychopathic Traits in Romantic Partners Predict Post-Traumatic Stress Disorder Symptomology: Evidence for Unique Impact in a Large Sample. *Personality and Individual Differences* 201: 111942 https://doi.org/10.1016/j.paid.2022.111942

Aronson E, Fried C, Stone J (1991) Overcoming Denial and Increasing the Intention to Use Condoms through the Induction of Hypocrisy. *American Journal of Public Health* 81(12): 1636–1638 https://doi.org/10.2105/AJPH.81.12.1636

Banas JA, Rains SA (2010) A Meta-Analysis of Research on Inoculation Theory. *Communication Monographs* 77(3): 281–311 https://doi.org/10.1080/03637751003758193

Bandura A (1997) *Self-Efficacy: The Exercise of Control*. W. H. Freeman and Company, New York

Bandura A (2000) Exercise of Human Agency Through Collective Efficacy. *Current Directions in Psychological Science* 9(3): 75–78 https://doi.org/10.1111/1467-8721.00064

Bandura A (2001) Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology* 52: 1–26 https://doi.org/10.1146/annurev.psych.52.1.1

Barton B, Zlatevska N, Oppewal H (2022) Scarcity tactics in marketing: A meta-analysis of product scarcity effects on consumer purchase intentions. *Journal of Retailing* 98(4): 741–758 https://doi.org/10.1016/j.jretai.2022.06.003

Baumeister RF, Stillwell AM, Heatherton TF (1994) Guilt: An Interpersonal Approach. *Psychological Bulletin* 115(2): 243–267 https://doi.org/10.1037/0033-2909.115.2.243

Baumgartner FR, Jones BD (2009) *Agendas and Instability in American Politics*. 2nd edn. University of Chicago Press, Chicago https://doi.org/10.7208/chicago/9780226039534.001.0001

Bavelas JB, Black A, Chovil N et al (1990) Truths, Lies, and Equivocations: The Effects of Conflicting Goals on Discourse. *Journal of Language and Social Psychology* 9(1–2): 135–161 https://doi.org/10.1177/0261927X9091008

Beller S, Bender A, Kuhnmünch G (2005) Understanding Conditional Promises and Threats. *Thinking & Reasoning* 11(3): 209–238 https://doi.org/10.1080/13546780442000141

Bem DJ (1972) Self-Perception Theory. In: Berkowitz L (ed) *Advances in Experimental Social Psychology*, vol 6. Academic Press, pp 1–62. https://doi.org/10.1016/S0065-2601(08)60024-6

Benford RD, Snow DA (2000) Framing Processes and Social Movements: An Overview and Assessment. *Annual Review of Sociology* 26: 611–639 https://doi.org/10.1146/annurev.soc.26.1.611

Bianchi M, Carnaghi A, Piccoli V et al (2019) On the Descriptive and Expressive Function of Derogatory Group Labels: An Experimental Test. *Journal of Language and Social Psychology* 38(5–6): 756–772 https://doi.org/10.1177/0261927X19867739

Biddle JB, Leuschner A (2015) Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental? *European Journal for Philosophy of Science* 5(3): 261–278 https://doi.org/10.1007/s13194-014-0101-x

Biržietienė S, Gabrėnaitė E (2014) Eristic Argumentation in Advertising. *Respectus Philologicus* (26(31)): 190–202 https://doi.org/10.15388/RESPECTUS.2014.26.31.15

Björkenfeldt O (2026) Self-Censorship in Journalism: The Role of Emotional, Professional, and Institutional Factors. *Digital Journalism* 14(3): 493–510 https://doi.org/10.1080/21670811.2025.2601961

Boin A (2005) Disaster research and future crises: Broadening the research agenda. *International Journal of Mass Emergencies & Disasters* 23(3): 199–214 https://doi.org/10.1177/028072700502300310

Boster FJ, Cruz S, Manata B et al (2016) A Meta-Analytic Review of the Effect of Guilt on Compliance. *Social Influence* 11(1): 54–67 https://doi.org/10.1080/15534510.2016.1142892

Boutyline A, Willer R (2017) The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks. *Political Psychology* 38(3): 551–569 https://doi.org/10.1111/pops.12337

Bowdle BF, Gentner D (2005) The Career of Metaphor. *Psychological Review* 112(1): 193–216 https://doi.org/10.1037/0033-295X.112.1.193

Bowell T (2023) Whataboutisms: The Good, the Bad and the Ugly. *Informal Logic* 43(1): 91–112 https://doi.org/10.22329/il.v43i1.7304

Boykoff MT, Boykoff JM (2004) Balance as Bias: Global Warming and the US Prestige Press. *Global Environmental Change* 14(2): 125–136 https://doi.org/10.1016/j.gloenvcha.2003.10.001

Boym S (2001) *The Future of Nostalgia*. Basic Books, New York

Braddock K, Dillard JP (2016) Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs* 83(4): 446–467 https://doi.org/10.1080/03637751.2015.1128555

Brady WJ, Gantman AP, Van Bavel JJ (2020) Attentional Capture Helps Explain Why Moral and Emotional Content Go Viral. *Journal of Experimental Psychology: General* 149(4): 746–756 https://doi.org/10.1037/xge0000673

Brady WJ, Wills JA, Jost JT et al (2017) Emotion Shapes the Diffusion of Moralized Content in Social Networks. *Proceedings of the National Academy of Sciences* 114(28): 7313–7318 https://doi.org/10.1073/pnas.1618923114

Brambilla M, Sacchi S, Pagliaro S et al (2013) Morality and Intergroup Relations: Threats to Safety and Group Image Predict the Desire to Interact with Outgroup and Ingroup Members. *Journal of Experimental Social Psychology* 49(5): 811–821 https://doi.org/10.1016/j.jesp.2013.04.005

Brehm JW (1966) *A Theory of Psychological Reactance*. Academic Press, New York

Brennen JS, Simon F, Howard PN et al (2020) Types, Sources, and Claims of COVID-19 Misinformation. Factsheet, Reuters Institute for the Study of Journalism. https://doi.org/10.60625/risj-awvq-sr55

Brosch T, Van Bavel JJ (2012) The Flexibility of Emotional Attention: Accessible Social Identities Guide Rapid Attentional Orienting. *Cognition* 125(2): 309–316 https://doi.org/10.1016/j.cognition.2012.07.007

Brouwer J, Woude M van der, Leun J van der (2017) Framing Migration and the Process of Crimmigration: A Systematic Analysis of the Media Representation of Unauthorized Immigrants in the Netherlands. *European Journal of Criminology* 14(1): 100–119 https://doi.org/10.1177/1477370816640136

Brown RJ, Turner JC (1979) The Criss-Cross Categorization Effect in Intergroup Discrimination. *British Journal of Social and Clinical Psychology* 18(4): 371–383 https://doi.org/10.1111/j.2044-8260.1979.tb00907.x

Bruner J (1990) *Acts of Meaning: Four Lectures on Mind and Culture*. Harvard University Press, Cambridge, MA

Brunton F, Nissenbaum H (2015) *Obfuscation: A user's guide for privacy and protest*. MIT Press, Cambridge, MA https://doi.org/10.7551/mitpress/9780262029735.001.0001

Bullock OM, Colón Amill D, Shulman HC et al (2019) Jargon as a Barrier to Effective Science Communication: Evidence from Metacognition. *Public Understanding of Science* 28(7): 845–853 https://doi.org/10.1177/0963662519865687

Burger JM (1986) Increasing Compliance by Improving the Deal: The That's-Not-All Technique. *Journal of Personality and Social Psychology* 51(2): 277–283 https://doi.org/10.1037/0022-3514.51.2.277

Burger JM (1999) The Foot-in-the-Door Compliance Procedure: A Multiple-Process Analysis and Review. *Personality and Social Psychology Review* 3(4): 303–325 https://doi.org/10.1207/S15327957PSPR0304_2

Burger JM, Caputo D (2015) The Low-Ball Compliance Procedure: A Meta-Analysis. *Social Influence* 10(4): 214–220 https://doi.org/10.1080/15534510.2015.1049203

Burger JM, Petty RE (1981) The Low-Ball Compliance Technique: Task or Person Commitment? *Journal of Personality and Social Psychology* 40(3): 492–500 https://doi.org/10.1037/0022-3514.40.3.492

Buttle FA (1998) Word of mouth: understanding and managing referral marketing. *Journal of Strategic Marketing* 6(3): 241–254 https://doi.org/10.1080/096525498346658

Buzan B, Wæver O, Wilde J de (1998) *Security: A New Framework for Analysis*. Lynne Rienner Publishers, Boulder, CO https://doi.org/10.1515/9781685853808

Calanchini J, Moons WG, Mackie DM (2016) Angry Expressions Induce Extensive Processing of Persuasive Appeals. *Journal of Experimental Social Psychology* 64: 88–98 https://doi.org/10.1016/j.jesp.2016.02.004

Callahan SP, Ledgerwood A (2016) On the Psychological Function of Flags and Logos: Group Identity Symbols Increase Perceived Entitativity. *Journal of Personality and Social Psychology* 110(4): 528–550 https://doi.org/10.1037/pspi0000047

Capozza D, Di Bernardo GA, Falvo R (2017) Intergroup Contact and Outgroup Humanization: Is the Causal Relationship Uni- or Bidirectional? *PLOS ONE* 12(1): e0170554 https://doi.org/10.1371/journal.pone.0170554

Capraro V, Vanzo A (2019) The Power of Moral Words: Loaded Language Generates Framing Effects in the Extreme Dictator Game. *Judgment and Decision Making* 14(3): 309–317 https://doi.org/10.1017/S1930297500004356

Carlson TN, Settle JE (2022) *What Goes Without Saying: Navigating Political Discussion in America*. Cambridge University Press, New York https://doi.org/10.1017/9781108912495

Carnaghi A, Maass A, Gresta S et al (2008) Nomina Sunt Omina: On the Inductive Potential of Nouns and Adjectives in Person Perception. *Journal of Personality and Social Psychology* 94(5): 839–859 https://doi.org/10.1037/0022-3514.94.5.839

Carpenter Dick M. I, Webster MJ, Bowman CK (2019) White House Wit: How Presidents Use Humor as a Leadership Tool. *Presidential Studies Quarterly* 49(1): 23–55 https://doi.org/10.1111/psq.12492

Cayuela Tudela AD, Pastor-Galindo J, Nespoli P et al (2025) *The Influence Operation Ontology (IOO)*. https://doi.org/10.48550/arXiv.2503.07304

Chaiken S (1980) Heuristic versus Systematic Information Processing and the Use of Source versus Message Cues in Persuasion. *Journal of Personality and Social Psychology* 39(5): 752–766 https://doi.org/10.1037/0022-3514.39.5.752

Chatterjee JS, Thongprasert S, Some S (2024) Communication, Climate Mitigation, and Behavior Change Interventions: Understanding Message Design and Digital Media Technologies. *Annual Review of Environment and Resources* 49: 655–672 https://doi.org/10.1146/annurev-environ-121322-070317

Chaudhuri A, Holbrook MB (2001) The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The Role of Brand Loyalty. *Journal of Marketing* 65(2): 81–93 https://doi.org/10.1509/jmkg.65.2.81.18255

Chen M, Cheung ASY, Chan KL (2019) Doxing: What Adolescents Look for and Their Intentions. *International Journal of Environmental Research and Public Health* 16(2): 218 https://doi.org/10.3390/ijerph16020218

Chernobrov D (2019) Who Is the Modern 'Traitor'? 'Fifth Column' Accusations in US and UK Politics and Media. *Politics* 39(3): 347–362 https://doi.org/10.1177/0263395718776215

Chernobrov D (2024) Humor and Foreign Policy Narration: The Persuasive Power and Limitations of Russia's Foreign Policy Pranks. *Global Studies Quarterly* 4(1): ksae007 https://doi.org/10.1093/isagsq/ksae007

Cheryan S, Monin B (2005) Where Are You Really From? Asian Americans and Identity Denial. *Journal of Personality and Social Psychology* 89(5): 717–730 https://doi.org/10.1037/0022-3514.89.5.717

Chisholm RM, Feehan TD (1977) The Intent to Deceive. *The Journal of Philosophy* 74(3): 143–159 https://doi.org/10.2307/2025605

Chiu K, Grundy Q, Bero L (2017) Spin in Published Biomedical Literature: A Methodological Systematic Review. *PLOS Biology* 15(9): e2002173 https://doi.org/10.1371/journal.pbio.2002173

Chong D, Druckman JN (2007) Framing Theory. *Annual Review of Political Science* 10: 103–126 https://doi.org/10.1146/annurev.polisci.10.072805.103054

Cialdini RB, Reno RR, Kallgren CA (1990) A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58(6): 1015–1026 https://doi.org/10.1037/0022-3514.58.6.1015

Cialdini RB (2007) *Influence: The Psychology of Persuasion*. Revised0th edn. Harper Business, New York

Cialdini RB (2016) *Pre-Suasion: A Revolutionary Way to Influence and Persuade*. Simon & Schuster, New York

Cialdini RB, Cacioppo JT, Bassett R et al (1978) Low-Ball Procedure for Producing Compliance: Commitment Then Cost. *Journal of Personality and Social Psychology* 36(5): 463–476 https://doi.org/10.1037/0022-3514.36.5.463

Cialdini RB, Goldstein NJ (2004) Social Influence: Compliance and Conformity. *Annual Review of Psychology* 55: 591–621 https://doi.org/10.1146/annurev.psych.55.090902.142015

Cialdini RB, Trost MR, Newsom JT (1995) Preference for Consistency: The Development of a Valid Measure and the Discovery of Surprising Behavioral Implications. *Journal of Personality and Social Psychology* 69(2): 318–328 https://doi.org/10.1037/0022-3514.69.2.318

Cialdini RB, Vincent JE, Lewis SK et al (1975) Reciprocal Concessions Procedure for Inducing Compliance: The Door-in-the-Face Technique. *Journal of Personality and Social Psychology* 31(2): 206–215 https://doi.org/10.1037/h0076284

Cirone A, Hobbs W (2023) Asymmetric Flooding as a Tool for Foreign Influence on Social Media. *Political Science Research and Methods* 11(1): 160–171 https://doi.org/10.1017/psrm.2022.9

Cohen S (2011) *Folk devils and moral panics*. Routledge, Abingdon, Oxon https://doi.org/10.4324/9780203828250

Compton J, Linden S van der, Cook J et al (2021) Inoculation Theory in the Post-Truth Era: Extant Findings and New Frontiers for Contested Science, Misinformation, and Conspiracy Theories. *Social and Personality Psychology Compass* 15(6): e12602 https://doi.org/10.1111/spc3.12602

Condon P, DeSteno D (2011) Compassion for One Reduces Punishment for Another. *Journal of Experimental Social Psychology* 47(3): 698–701 https://doi.org/10.1016/j.jesp.2010.11.016

Coons C, Weber M, eds (2014) *Manipulation: Theory and Practice*. Oxford University Press, New York https://doi.org/10.1093/acprof:oso/9780199338207.001.0001

Crisp RJ, Hewstone M (1999) Differential Evaluation of Crossed Category Groups: Patterns, Processes, and Reducing Intergroup Bias. *Group Processes & Intergroup Relations* 2(4): 307–333 https://doi.org/10.1177/1368430299024001

Crisp RJ, Hewstone M, Rubin M (2001) Does Multiple Categorization Reduce Intergroup Bias? *Personality and Social Psychology Bulletin* 27(1): 76–89 https://doi.org/10.1177/0146167201271007

D'Errico F, Poggi I (2016) "The Bitter Laughter": When Parody Is a Moral and Affective Priming in Political Persuasion. *Frontiers in Psychology* 7: 1144 https://doi.org/10.3389/fpsyg.2016.01144

Da San Martino G, Barrón-Cedeño A, Wachsmuth H et al (2020) SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles. In: Herbelot A, Zhu X, Palmer A et al (eds) *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. International Committee for Computational Linguistics, Barcelona (online), pp 1377–1414. https://doi.org/10.18653/v1/2020.semeval-1.186

Da San Martino G, Yu S, Barrón-Cedeño A et al (2019) Fine-Grained Analysis of Propaganda in News Articles. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, pp 5636–5646. https://doi.org/10.18653/v1/D19-1565

Darke L, Paterson H, Golde C van (2025) Gaslighting and Memory: The Effects of Partner-Led Challenges on Recall and Self-Perception. *Memory* 33(7): 828–844 https://doi.org/10.1080/09658211.2025.2533253

Daur V (2025) Legitimize or Delegitimize? Mainstream Party Strategy toward (Former) Pariah Parties and How Voters Respond. *American Political Science Review* 119(3): 1514–1529 https://doi.org/10.1017/S0003055424000996

De Dreu CKW (1995) Coercive Power and Concession Making in Bilateral Negotiation. *Journal of Conflict Resolution* 39(4): 646–670 https://doi.org/10.1177/0022002795039004003

Demetriades SZ, Kalny CS, Turner MM et al (2024) Is All Anger Created Equal? A Meta-Analytic Assessment of Anger Elicitation in Persuasion Research. *Emotion* 24(6): 1428–1441 https://doi.org/10.1037/emo0001360

Deutsch M, Gerard HB (1955) A Study of Normative and Informational Social Influences upon Individual Judgement. *Journal of Abnormal and Social Psychology* 51(3): 629–636 https://doi.org/10.1037/h0046408

Diel A, Lalgi T, Schröter IC et al (2024) Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. *Computers in Human Behavior Reports* 16: 100538 https://doi.org/10.1016/j.chbr.2024.100538

Dimitrov D, Alam F, Hasanain M et al (2024) SemEval-2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes. In: *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Association for Computational Linguistics, pp 2009–2026. https://doi.org/10.18653/v1/2024.semeval-1.275

Dimitrov D, Bin Ali B, Shaar S et al (2021) SemEval-2021 Task 6: Detection of Persuasion Techniques in Texts and Images. In: *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*. Association for Computational Linguistics, Online, pp 70–98. https://doi.org/10.18653/v1/2021.semeval-1.7

DISARM Foundation (n.d.) *DISARM Frameworks*. https://github.com/DISARMFoundation/DISARMframeworks Accessed 28 Aug 2026

Dixon GN, Clarke CE (2013) Heightening Uncertainty Around Certain Science: Media Coverage, False Balance, and the Autism–Vaccine Controversy. *Science Communication* 35(3): 358–382 https://doi.org/10.1177/1075547012458290

Douglas KM, Uscinski JE, Sutton RM et al (2019) Understanding Conspiracy Theories. *Political Psychology* 40(S1): 3–35 https://doi.org/10.1111/pops.12568

Dovidio JF, Gaertner SL, Kafati G (2000) Group Identity and Intergroup Relations: The Common In-Group Identity Model. In: Thye SR, Lawler EJ, Macy MW et al (eds) *Advances in Group Processes*, vol 17. JAI Press, Stamford, CT, pp 1–34. https://doi.org/10.1016/S0882-6145(00)17002-X

Dowling M-E, Legrand T (2023) "I do not consent": Political Legitimacy, Misinformation, and the Compliance Challenge in Australia's Covid-19 Policy Response. *Policy and Society* 42(3): 319–333 https://doi.org/10.1093/polsoc/puad018

Drury J, Brown R, González R et al (2016) Emergent Social Identity and Observing Social Support Predict Social Support Provided by Survivors in a Disaster: Solidarity in the 2010 Chile Earthquake. *European Journal of Social Psychology* 46(2): 209–223 https://doi.org/10.1002/ejsp.2146

Drury J, Cocking C, Reicher S (2009) Everyone for Themselves? A Comparative Study of Crowd Solidarity among Emergency Survivors. *British Journal of Social Psychology* 48(3): 487–506 https://doi.org/10.1348/014466608X357893

Eatwell R (2006) Explaining Fascism and Ethnic Cleansing: The Three Dimensions of Charisma and the Four Dark Sides of Nationalism. *Political Studies Review* 4(3): 263–278 https://doi.org/10.1111/j.1478-9299.2006.00109.x

Echterhoff G, Higgins ET, Levine JM (2009) Shared Reality: Experiencing Commonality with Others' Inner States about the World. *Perspectives on Psychological Science* 4(5): 496–521 https://doi.org/10.1111/j.1745-6924.2009.01161.x

Eckert S, Metzger-Riftkin J (2020) Doxxing, privacy and gendered harassment. The shock and normalization of veillance cultures. *M&K Medien & Kommunikationswissenschaft* 68(3): 273–287 https://doi.org/10.5771/1615-634X-2020-3-273

Edelman M (1988) *Constructing the Political Spectacle*. University of Chicago Press, Chicago

Eisenberg EM (1984) Ambiguity as Strategy in Organizational Communication. *Communication Monographs* 51(3): 227–242 https://doi.org/10.1080/03637758409390197

Elhauge E (2016) Contrived Threats versus Uncontrived Warnings: A General Solution to the Puzzles of Contractual Duress, Unconstitutional Conditions, and Blackmail. *University of Chicago Law Review* 83(2): 503–585

Ellemers N, Kortekaas P, Ouwerkerk JW (1999) Self-Categorisation, Commitment to the Group and Group Self-Esteem as Related but Distinct Aspects of Social Identity. *European Journal of Social Psychology* 29(2–3): 371–389 https://doi.org/10.1002/(SICI)1099-0992(199903/05)29:2/3<371::AID-EJSP932>3.0.CO;2-U

Enock FE, Over H (2023) Animalistic Slurs Increase Harm by Changing Perceptions of Social Desirability. *Royal Society Open Science* 10(7), 230203: https://doi.org/10.1098/rsos.230203

Entman RM (1993) Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication* 43(4): 51–58 https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Eriksson J (2020) Threat Framing. In: *Oxford Research Encyclopedia of Politics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.1647

Esses VM, Dovidio JF (2002) The Role of Emotions in Determining Willingness to Engage in Intergroup Contact. *Personality and Social Psychology Bulletin* 28(9): 1202–1214 https://doi.org/10.1177/01461672022812006

Esses VM, Jackson LM, Armstrong TL (1998) Intergroup Competition and Attitudes Toward Immigrants and Immigration: An Instrumental Model of Group Conflict. *Journal of Social Issues* 54(4): 699–724 https://doi.org/10.1111/j.1540-4560.1998.tb01244.x

Esses VM, Medianu S, Lawson AS (2013) Uncertainty, Threat, and the Role of the Media in Promoting the Dehumanization of Immigrants and Refugees. *Journal of Social Issues* 69(3): 518–536 https://doi.org/10.1111/josi.12027

European Union Agency for Cybersecurity (ENISA), European External Action Service (EEAS) (2022) Foreign Information Manipulation and Interference (FIMI) and Cybersecurity: Threat Landscape. Technical report, European Union Agency for Cybersecurity. https://doi.org/10.2824/7501

EUvsDisinfo (n.d.) *EUvsDisinfo Database*. https://euvsdisinfo.eu/disinformation-cases/ Accessed 28 Aug 2026

Evans JSBT, Twyman-Musgrove J (1998) Conditional Reasoning with Inducements and Advice. *Cognition* 69(1): B11–B16 https://doi.org/10.1016/S0010-0277(98)00062-6

Farina A, Fréchette GR, Ispano A et al (2026) The Selective Disclosure of Evidence: An Experiment. *The Review of Economic Studies*: https://doi.org/10.1093/restud/rdag076

Fazio LK, Brashier NM, Payne BK et al (2015) Knowledge Does Not Protect Against Illusory Truth. *Journal of Experimental Psychology: General* 144(5): 993–1002 https://doi.org/10.1037/xge0000098

Feinberg M, Willer R (2015) From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence? *Personality and Social Psychology Bulletin* 41(12): 1665–1681 https://doi.org/10.1177/0146167215607842

Feldman DC (1984) The Development and Enforcement of Group Norms. *Academy of Management Review* 9(1): 47–53 https://doi.org/10.5465/amr.1984.4277934

Fishbein M, Ajzen I (1975) *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA

Fisher WR (1984) Narration as a Human Communication Paradigm: The Case of Public Moral Argument. *Communication Monographs* 51(1): 1–22 https://doi.org/10.1080/03637758409390180

Flesher Fominaya C (2007) The Role of Humour in the Process of Collective Identity Formation in Autonomous Social Movement Groups in Contemporary Madrid. *International Review of Social History* 52(S15): 243–258 https://doi.org/10.1017/S0020859007003227

Ford TE, Ferguson MA (2004) Social Consequences of Disparagement Humor: A Prejudiced Norm Theory. *Personality and Social Psychology Review* 8(1): 79–94 https://doi.org/10.1207/S15327957PSPR0801_4

Fortify Rights (2018) "They Gave Them Long Swords": Preparations for Genocide and Crimes Against Humanity Against Rohingya Muslims in Rakhine State, Myanmar. Technical report, Fortify Rights. https://www.fortifyrights.org/mya-inv-2018-07-19/

François C (2019) Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry and Regulatory Responses. Technical report, Transatlantic Working Group on Content Moderation Online and Freedom of Expression. https://docs.house.gov/meetings/SY/SY21/20190926/109980/HHRG-116-SY21-Wstate-FrancoisC-20190926-SD001.pdf

Freedman JL, Fraser SC (1966) Compliance Without Pressure: The Foot-in-the-Door Technique. *Journal of Personality and Social Psychology* 4(2): 195–202 https://doi.org/10.1037/h0023552

Friedman M (1991) Consumer Boycotts: A Conceptual Framework and Research Agenda. *Journal of Social Issues* 47(1): 149–168 https://doi.org/10.1111/j.1540-4560.1991.tb01817.x

Gable P, Harmon-Jones E (2010) The Blues Broaden, but the Nasty Narrows: Attentional Consequences of Negative Affects Low and High in

Motivational Intensity. *Psychological Science* 21(2): 211–215 https://doi.org/10.1177/0956797609359622

Gabrielsen J, Jønch-Clausen H, Pontoppidan C (2020) Answering without Answering: Shifting as an Evasive Rhetorical Strategy. *Journalism* 21(9): 1355–1370 https://doi.org/10.1177/1464884917738412

Gaertner SL, Dovidio JF, Anastasio PA et al (1993) The Common Ingroup Identity Model: Recategorization and the Reduction of Intergroup Bias. *European Review of Social Psychology* 4(1): 1–26 https://doi.org/10.1080/14792779343000004

Gaertner SL, Dovidio JF, Bachman BA (1996) Revisiting the Contact Hypothesis: The Induction of a Common Ingroup Identity. *International Journal of Intercultural Relations* 20(3–4): 271–290 https://doi.org/10.1016/0147-1767(96)00019-3

Gaertner SL, Dovidio JF, Rust MC et al (1999) Reducing Intergroup Bias: Elements of Intergroup Cooperation. *Journal of Personality and Social Psychology* 76(3): 388–402 https://doi.org/10.1037/0022-3514.76.3.388

Galinsky AD, Moskowitz GB (2000) Perspective-Taking: Decreasing Stereotype Expression, Stereotype Accessibility, and In-Group Favoritism. *Journal of Personality and Social Psychology* 78(4): 708–724 https://doi.org/10.1037/0022-3514.78.4.708

Geissler D, Bär D, Pröllochs N et al (2023) Russian propaganda on social media during the 2022 invasion of Ukraine. *EPJ Data Science* 12(1): 35 https://doi.org/10.1140/epjds/s13688-023-00414-5

Gentner D (1983) Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7(2): 155–170 https://doi.org/10.1207/s15516709cog0702_3

George AL (1991) *Forceful Persuasion: Coercive Diplomacy as an Alternative to War*. United States Institute of Peace Press, Washington, DC

Glick P (2002) Sacrificial Lambs Dressed in Wolves' Clothing: Envious Prejudice, Ideology, and the Scapegoating of Jews. In: Newman LS, Erber R (eds) *Understanding Genocide: The Social Psychology of the Holocaust*. Oxford University Press, pp 112–142. https://doi.org/10.1093/acprof:oso/9780195133622.003.0006

Glimm B, Horrocks I, Motik B et al (2014) HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning* 53(3): 245–269 https://doi.org/10.1007/s10817-014-9305-1

Goldstein NJ, Cialdini RB, Griskevicius V (2008) A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research* 35(3): 472–482 https://doi.org/10.1086/586910

Gordon RA (1996) Impact of Ingratiation on Judgments and Evaluations: A Meta-Analytic Investigation. *Journal of Personality and Social Psychology* 71(1): 54–70 https://doi.org/10.1037/0022-3514.71.1.54

Gordon R (2012) Re-thinking and re-tooling the social marketing mix. *Australasian Marketing Journal* 20(2): 122–126 https://doi.org/10.1016/j.ausmj.2011.10.005

Gordon WJ (1993) Truth and Consequences: The Force of Blackmail's Central Case. *University of Pennsylvania Law Review* 141(5): 1741–1785 https://doi.org/10.2307/3312573

Gouldner AW (1960) The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review* 25(2): 161–178 https://doi.org/10.2307/2092623

Grasmick HG, Bursik Jr. RJ, Kinsey KA (1991) Shame and Embarrassment as Deterrents to Noncompliance with the Law: The Case of an Antilittering Campaign. *Environment and Behavior* 23(2): 233–251 https://doi.org/10.1177/0013916591232006

Gray CM, Kou Y, Battles B et al (2018) The Dark (Patterns) Side of UX Design. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, pp 1–14. https://doi.org/10.1145/3173574.3174108

Greatbatch D (1986) Aspects of Topical Organization in News Interviews: The Use of Agenda-Shifting Procedures by Interviewees. *Media, Culture & Society* 8(4): 441–455 https://doi.org/10.1177/0163443786008004005

Green MC, Appel M (2024) Narrative Transportation: How Stories Shape How We See Ourselves and the World. In: Gawronski B (ed) *Advances in Experimental Social Psychology*, vol 70. Academic Press, pp 1–82. https://doi.org/10.1016/bs.aesp.2024.03.002

Green MC, Brock TC (2000) The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology* 79(5): 701–721 https://doi.org/10.1037/0022-3514.79.5.701

Greimas AJ (1966) *Sémantique structurale*. Larousse, Paris

Gruber JB (2023) Troublemakers in the Streets? A Framing Analysis of Newspaper Coverage of Protests in the UK 1992–2017. *The International Journal of Press/Politics* 28(2): 414–433 https://doi.org/10.1177/19401612221102058

Gur T, Ayal S, Wagner M et al (2024) A Group That Grieves Together Stays Together: Examining the Impact of Holocaust Memorial Day in Israel on Affective Polarization. *Political Psychology* 45(6): 1051–1068 https://doi.org/10.1111/pops.12956

Gustafson A, Rice RE (2020) A Review of the Effects of Uncertainty in Public Science Communication. *Public Understanding of Science* 29(6): 614–633 https://doi.org/10.1177/0963662520942122

Haber N, Smith ER, Moscoe E et al (2018) Causal Language and Strength of Inference in Academic and Media Articles Shared in Social Media (CLAIMS): A Systematic Review. *PLOS ONE* 13(5): e0196346 https://doi.org/10.1371/journal.pone.0196346

Haidt J (2012) *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon, New York

Haigh M, Stewart AJ, Wood JS et al (2011) Conditional Advice and Inducements: Are Readers Sensitive to Implicit Speech Acts during Comprehension? *Acta Psychologica* 136(3): 419–424 https://doi.org/10.1016/j.actpsy.2011.01.009

Haigh M, Wood JS, Stewart AJ (2016) Slippery Slope Arguments Imply Opposition to Change. *Memory & Cognition* 44(5): 819–836 https://doi.org/10.3758/s13421-016-0596-9

Hall S, Critcher C, Jefferson T et al (1978) *Policing the Crisis: Mugging, the State and Law and Order*. Macmillan Press, London

Hariman R, Lucaites JL (2007) *No Caption Needed: Iconic Photographs, Public Culture, and Liberal Democracy*. University of Chicago Press, Chicago

Hasher L, Goldstein D, Toppino T (1977) Frequency and the Conference of Referential Validity. *Journal of Verbal Learning and Verbal Behavior* 16(1): 107–112 https://doi.org/10.1016/S0022-5371(77)80012-1

Haslam N (2006) Dehumanization: An Integrative Review. *Personality and Social Psychology Review* 10(3): 252–264 https://doi.org/10.1207/s15327957pspr1003_4

Heckathorn DD (1990) Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control. *American Sociological Review* 55(3): 366–384 https://doi.org/10.2307/2095762

Herring DR, White KR, Jabeen LN et al (2013) On the Automatic Activation of Attitudes: A Quarter Century of Evaluative Priming Research. *Psychological Bulletin* 139(5): 1062–1089 https://doi.org/10.1037/a0031309

Hicks A, McKenzie P, Bronstein J et al (2025) Information Avoidance: A Critical Conceptual Review. An Annual Review of Information Science and Technology (ARIST) Paper. *Journal of the Association for Information Science and Technology* 76(1): 326–346 https://doi.org/10.1002/asi.24968

Hilberts S, Govers M, Petelos E et al (2025) The Impact of Misinformation on Social Media in the Context of Natural Disasters: Narrative Review. *JMIR Infodemiology* 5: e70413 https://doi.org/10.2196/70413

Hirschberger G, Ein-Dor T, Leidner B et al (2016) How Is Existential Threat Related to Intergroup Conflict? Introducing the Multidimensional Existential Threat (MET) Model. *Frontiers in Psychology* 7: 1877 https://doi.org/10.3389/fpsyg.2016.01877

Hoeken H, Hustinx L (2009) When Is Statistical Evidence Superior to Anecdotal Evidence in Supporting Probability Claims? The Role of Argument Type. *Human Communication Research* 35(4): 491–510 https://doi.org/10.1111/j.1468-2958.2009.01360.x

Hofstadter R (1965) *The Paranoid Style in American Politics and Other Essays*. Knopf, New York

Hogg MA, Reid SA (2006) Social Identity, Self-Categorization, and the Communication of Group Norms. *Communication Theory* 16(1): 7–30 https://doi.org/10.1111/j.1468-2885.2006.00003.x

Hornikx J (2005) A Review of Experimental Research on the Relative Persuasiveness of Anecdotal, Statistical, Causal, and Expert Evidence. *Studies in Communication Sciences* 5(1): 205–216

Hovland CI, Weiss W (1951) The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly* 15(4): 635–650 https://doi.org/10.1086/266350

Howard DJ, Gengler C, Jain A (1995) What's in a Name? A Complimentary Means of Persuasion. *Journal of Consumer Research* 22(2): 200–211 https://doi.org/10.1086/209445

Howard S, Kennedy KC, Vine KT (2023) You Don't Believe in God? You Ain't Black: Identifying as Atheist Elicits Identity Denial From Black Ingroup Members. *Cultural Diversity and Ethnic Minority Psychology* 29(2): 202–207 https://doi.org/10.1037/cdp0000397

Institute for Propaganda Analysis (1939) *The Fine Art of Propaganda: A Study of Father Coughlin's Speeches*. Harcourt, Brace and Company, New York

Iyengar S (1991) *Is Anyone Responsible? How Television Frames Political Issues*. University of Chicago Press, Chicago https://doi.org/10.7208/chicago/9780226388533.001.0001

Iyengar S, Sood G, Lelkes Y (2012) Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly* 76(3): 405–431 https://doi.org/10.1093/poq/nfs038

Iyengar S, Westwood SJ (2015) Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science* 59(3): 690–707 https://doi.org/10.1111/ajps.12152

Jankovic S, Roeben V (2025) Mind the Gap: The Determination, Legality and Consequences of Implicit Threats of Force. *Leiden Journal of International Law* 38(4): 812–838 https://doi.org/10.1017/S0922156525100290

Johnson N, Turnbull B, Maher T et al (2021) Semantically Modeling Cyber Influence Campaigns (CICs): Ontology Model and Case Studies. *IEEE Access* 9: 9365–9382 https://doi.org/10.1109/ACCESS.2020.3048269

Jones EE (1964) *Ingratiation: A Social Psychological Analysis*. Appleton-Century-Crofts, New York

Jowett GS, O'Donnell V (2018) *Propaganda & Persuasion*. 7th edn. SAGE Publications, Thousand Oaks, CA

Kaiser CR, Vick SB, Major B (2006) Prejudice Expectations Moderate Preconscious Attention to Cues That Are Threatening to Social Identity. *Psychological Science* 17(4): 332–338 https://doi.org/10.1111/j.1467-9280.2006.01707.x

Kassin SM, McNall K (1991) Police Interrogations and Confessions: Communicating Promises and Threats by Pragmatic Implication. *Law and Human Behavior* 15(3): 233–251 https://doi.org/10.1007/BF01061711

Kavanagh J, Rich MD (2018) Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life. Research Report RR-2314-RC, RAND Corporation, Santa Monica, CA. https://doi.org/10.7249/RR2314

Keen S (1986) *Faces of the Enemy: Reflections of the Hostile Imagination*. Harper & Row, San Francisco

Kellermann K, Shea BC (1996) Threats, Suggestions, Hints, and Promises: Gaining Compliance Efficiently and Politely. *Communication Quarterly* 44(2): 145–165 https://doi.org/10.1080/01463379609370007

Kelman HC (1958) Compliance, Identification, and Internalization: Three Processes of Attitude Change. *Journal of Conflict Resolution* 2(1): 51–60 https://doi.org/10.1177/002200275800200106

Kim YM, Hsu J, Neiman D et al (2018) The stealth media? Groups and targets behind divisive issue campaigns on Facebook. *Political Communication* 35(4): 515–541 https://doi.org/10.1080/10584609.2018.1476425

King G, Pan J, Roberts ME (2017) How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review* 111(3): 484–501 https://doi.org/10.1017/S0003055417000144

Kiousis S (2001) Public Trust or Mistrust? Perceptions of Media Credibility in the Information Age. *Mass Communication and Society* 4(4): 381–403 https://doi.org/10.1207/S15327825MCS0404_4

Knublauch H, Kontokostas D (2017) Shapes Constraint Language (SHACL). W3C Recommendation, World Wide Web Consortium (W3C). https://www.w3.org/TR/shacl/

Kodapanakkal RI, Brandt MJ, Kogler C et al (2022) Moral Frames Are Persuasive and Moralize Attitudes; Nonmoral Frames Are Persuasive and De-Moralize Attitudes. *Psychological Science* 33(3): 433–449 https://doi.org/10.1177/09567976211040803

Koehler DJ (2016) Can Journalistic "False Balance" Distort Public Perception of Consensus in Expert Opinion? *Journal of Experimental Psychology: Applied* 22(1): 24–38 https://doi.org/10.1037/xap0000073

Krishnarajan S, Tolstrup J (2023) Pre-war experimental evidence that Putin's propaganda elicited strong support for military invasion among Russians. *Science Advances* 9(45): eadg1199 https://doi.org/10.1126/sciadv.adg1199

Kteily N, Bruneau E, Waytz A et al (2015) The Ascent of Man: Theoretical and Empirical Evidence for Blatant Dehumanization. *Journal of Person-*

ality and Social Psychology 109(5): 901–931 https://doi.org/10.1037/pspp0000048

Kteily N, Cotterill S, Sidanius J et al (2014) Not One of Us: Predictors and Consequences of Denying Ingroup Characteristics to Ambiguous Targets. *Personality and Social Psychology Bulletin* 40(10): 1231–1247 https://doi.org/10.1177/0146167214539708

Kühberger A (1998) The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes* 75(1): 23–55 https://doi.org/10.1006/obhd.1998.2781

Kunda Z (1990) The Case for Motivated Reasoning. *Psychological Bulletin* 108(3): 480–498 https://doi.org/10.1037/0033-2909.108.3.480

Kuzman T, Ljubešić N (2025) LLM Teacher-Student Framework for Text Classification With No Manually Annotated Data: A Case Study in IPTC News Topic Classification. *IEEE Access* 13: 35621–35633 https://doi.org/10.1109/ACCESS.2025.3544814

Lahtinen V, Dietrich T, Rundle-Thiele S (2020) Long live the marketing mix. Testing the effectiveness of the commercial marketing mix in a social marketing context. *Journal of Social Marketing* 10(3): 357–375 https://doi.org/10.1108/JSOCM-10-2018-0122

Lakoff G (2004) *Don't Think of an Elephant!: Know Your Values and Frame the Debate*. Chelsea Green Publishing, White River Junction, VT

Lamond G (1996) Coercion, Threats, and the Puzzle of Blackmail. In: Simester AP, Smith ATH (eds) *Harm and Culpability*. Oxford University Press, pp 215–238. https://doi.org/10.1093/acprof:oso/9780198260578.003.0018

Lang A (2000) The limited capacity model of mediated message processing. *Journal of Communication* 50(1): 46–70 https://doi.org/10.1111/j.1460-2466.2000.tb02833.x

Lazer DMJ, Baum MA, Benkler Y et al (2018) The science of fake news. *Science* 359(6380): 1094–1096 https://doi.org/10.1126/science.aao2998

Leach CW, Ellemers N, Barreto M (2007) Group Virtue: The Importance of Morality (vs. Competence and Sociability) in the Positive Evaluation of In-Groups. *Journal of Personality and Social Psychology* 93(2): 234–249 https://doi.org/10.1037/0022-3514.93.2.234

Lenk T (2026) Towards a deeper understanding of information manipulation: Proposing a multilevel framework for the analysis of manipulative narratives. *Humanities and Social Sciences Communications* 13(1): 343 https://doi.org/10.1057/s41599-026-06656-8

Leonardi PM (2008) Indeterminacy and the Discourse of Inevitability in International Technology Management. *Academy of Management Review* 33(4): 975–984 https://doi.org/10.5465/amr.2008.34422017

Lerner JS, Keltner D (2001) Fear, Anger, and Risk. *Journal of Personality and Social Psychology* 81(1): 146–159 https://doi.org/10.1037/0022-3514.81.1.146

Levendusky M (2009) *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. University of Chicago Press, Chicago https://doi.org/10.7208/chicago/9780226473673.001.0001

Levendusky MS, Malhotra N (2016) (Mis)perceptions of Partisan Polarization in the American Public. *Public Opinion Quarterly* 80(S1): 378–391 https://doi.org/10.1093/poq/nfv045

Levin IP, Gaeth GJ (1988) How Consumers Are Affected by the Framing of Attribute Information Before and After Consuming the Product. *Journal of Consumer Research* 15(3): 374–378 https://doi.org/10.1086/209174

Levin IP, Schneider SL, Gaeth GJ (1998) All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes* 76(2): 149–188 https://doi.org/10.1006/obhd.1998.2804

Levine M, Prosser A, Evans D et al (2005) Identity and Emergency Intervention: How Social Group Membership and Inclusiveness of Group Boundaries Shape Helping Behavior. *Personality and Social Psychology Bulletin* 31(4): 443–453 https://doi.org/10.1177/0146167204271651

Lewandowsky S (2021) Climate Change, Disinformation and How to Combat It. *Annual Review of Public Health* 42: 1–21 https://doi.org/10.1146/annurev-publhealth-090419-102409

Lewandowsky S, Linden S van der (2021) Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology* 32(2): 348–384 https://doi.org/10.1080/10463283.2021.1876983

Lickel B, Miller N, Stenstrom DM et al (2006) Vicarious Retribution: The Role of Collective Blame in Intergroup Aggression. *Personality and Social Psychology Review* 10(4): 372–390 https://doi.org/10.1207/s15327957pspr1004_6

Linde C (1993) *Life stories: The creation of coherence*. Oxford University Press, New York https://doi.org/10.1093/oso/9780195073720.001.0001

Linden S van der (2022) Misinformation: Susceptibility, Spread, and Interventions to Immunize the Public. *Nature Medicine* 28(3): 460–467 https://doi.org/10.1038/s41591-022-01713-6

Lombardi Vallauri E (2021) Presupposition, Attention and Cognitive Load. *Journal of Pragmatics* 183: 15–28 https://doi.org/10.1016/j.pragma.2021.06.022

Luguri J, Strahilevitz LJ (2021) Shining a Light on Dark Patterns. *Journal of Legal Analysis* 13(1): 43–109 https://doi.org/10.1093/jla/laaa006

Luhtanen R, Crocker J (1992) A Collective Self-Esteem Scale: Self-Evaluation of One's Social Identity. *Personality and Social Psychology Bulletin* 18(3): 302–318 https://doi.org/10.1177/0146167292183006

Luke TJ, Alceste F (2020) The Mechanisms of Minimization: How Interrogation Tactics Suggest Lenient Sentencing Through Pragmatic Implication. *Law and Human Behavior* 44(4): 266–285 https://doi.org/10.1037/lhb0000410

Lynn M (1991) Scarcity Effects on Value: A Quantitative Review of the Commodity Theory Literature. *Psychology & Marketing* 8(1): 43–57 https://doi.org/10.1002/mar.4220080105

Macagno F (2024) Presuppositional Fallacies. *Argumentation* 38: 109–140 https://doi.org/10.1007/s10503-023-09625-6

Mack A, Rock I (1998) *Inattentional Blindness*. The MIT Press, Cambridge, MA

MacKay DG, Shafto M, Taylor JK et al (2004) Relations between Emotion, Memory, and Attention: Evidence from Taboo Stroop, Lexical Decision, and Immediate Memory Tasks. *Memory & Cognition* 32(3): 474–488 https://doi.org/10.3758/BF03195840

Mackie DM, Smith ER (2018) Intergroup Emotions Theory: Production, Regulation, and Modification of Group-Based Emotions. In: *Advances in Experimental Social Psychology*, vol 58. Academic Press, pp 1–69. https://doi.org/10.1016/bs.aesp.2018.03.001

Mahon JE (2016) The Definition of Lying and Deception. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, Winter 20160th edn. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/lying-definition/

Malcolm F (2023) Analysing Extremism. *Ethical Theory and Moral Practice* 26: 321–327 https://doi.org/10.1007/s10677-023-10370-8

Markowitz DM, Hancock JT (2016) Linguistic Obfuscation in Fraudulent Science. *Journal of Language and Social Psychology* 35(4): 435–445 https://doi.org/10.1177/0261927X15614605

Marwick AE, Lewis R (2017) Media Manipulation and Disinformation Online. Technical report, Data & Society Research Institute, New York. https://datasociety.net/research-library/media-manipulation-and-disinformation-online/

Mason L (2015) "I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science* 59(1): 128–145 https://doi.org/10.1111/ajps.12089

Mason L (2016) A Cross-Cutting Calm: How Social Sorting Drives Affective Polarization. *Public Opinion Quarterly* 80(S1): 351–377 https://doi.org/10.1093/poq/nfw001

Mathur A, Acar G, Friedman MJ et al (2019) Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proceedings of the ACM*

on Human-Computer Interaction 3(CSCW): 1–32 https://doi.org/10.1145/3359183

Mazziotta A, Mummendey A, Wright SC (2011) Vicarious Intergroup Contact Effects: Applying Social-Cognitive Theory to Intergroup Contact Research. *Group Processes & Intergroup Relations* 14(2): 255–274 https://doi.org/10.1177/1368430210390533

McCauley C, Moskalenko S (2017) Understanding Political Radicalization: The Two-Pyramids Model. *American Psychologist* 72(3): 205–216 https://doi.org/10.1037/amp0000062

McCombs ME, Shaw DL (1972) The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly* 36(2): 176–187 https://doi.org/10.1086/267990

McCoy J, Rahman T, Somer M (2018) Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities. *American Behavioral Scientist* 62(1): 16–42 https://doi.org/10.1177/0002764218759576

Media Manipulation Casebook (n.d.) *Methods: The Media Manipulation Life Cycle*. https://mediamanipulation.org/methods/ Accessed 28 Aug 2026

Medzihorsky J, Lindberg SI (2024) Walking the Talk: How to Identify Anti-Pluralist Parties. *Party Politics* 30(3): 420–434 https://doi.org/10.1177/13540688231153092

Meleady R (2021) "Nudging" Intergroup Contact: Normative Social Influences on Intergroup Contact Engagement. *Group Processes & Intergroup Relations* 24(7): 1180–1199 https://doi.org/10.1177/13684302211016047

Mepham K, Vörös A, Stadtfeld C (2025) Network Polarization: The Study of Political Attitudes and Social Ties as Dynamic Multilevel Networks. *Network Science* 13: e7 https://doi.org/10.1017/nws.2025.2

Metzger MJ, Flanagin AJ, Eyal K et al (2003) Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. *Communication Yearbook* 27(1): 293–335 https://doi.org/10.1080/23808985.2003.11679029

Meyerowitz BE, Chaiken S (1987) The Effect of Message Framing on Breast Self-Examination Attitudes, Intentions, and Behavior. *Journal of Personality and Social Psychology* 52(3): 500–510 https://doi.org/10.1037/0022-3514.52.3.500

Michaels D (2006) Manufactured Uncertainty: Protecting Public Health in the Age of Contested Science and Product Defense. *Annals of the New York Academy of Sciences* 1076(1): 149–162 https://doi.org/10.1196/annals.1371.058

Michaud N (2018) Appeal to Tradition. In: Arp R, Barbone S, Bruce M (eds) *Bad Arguments: 100 of the Most Important Fallacies in Western*

Philosophy. Wiley, pp 121–124. https://doi.org/10.1002/9781119165811.ch19

Milgram S (1963) Behavioral Study of Obedience. *Journal of Abnormal and Social Psychology* 67(4): 371–378 https://doi.org/10.1037/h0040525

Miller RL, Brickman P, Bolen D (1975) Attribution versus Persuasion as a Means for Modifying Behavior. *Journal of Personality and Social Psychology* 31(3): 430–441 https://doi.org/10.1037/h0076539

Miskimmon A, O'Loughlin B, Roselle L (2013) *Strategic Narratives: Communication Power and the New World Order*. Routledge, New York https://doi.org/10.4324/9781315871264

Mobayed T, Sanders JG (2022) Moral Foundational Framing and Its Impact on Attitudes and Behaviours. *Behavioral Sciences* 12(5): 118 https://doi.org/10.3390/bs12050118

Moffitt B (2016) *The Global Rise of Populism: Performance, Political Style, and Representation*. Stanford University Press, Stanford, CA https://doi.org/10.11126/stanford/9780804796132.001.0001

Mohammed D (2023) Argument by Association: On the Transmissibility of Commitment in Public Political Arguments. *Topoi* 42(2): 625–634 https://doi.org/10.1007/s11245-023-09914-5

Moody SJ (2019) Contextualizing Macro-Level Identities and Constructing Inclusiveness through Teasing and Self-Mockery: A View from the Intercultural Workplace in Japan. *Journal of Pragmatics* 152: 145–159 https://doi.org/10.1016/j.pragma.2019.05.023

Moons WG, Mackie DM (2007) Thinking Straight While Seeing Red: The Influence of Anger on Information Processing. *Personality and Social Psychology Bulletin* 33(5): 706–720 https://doi.org/10.1177/0146167206298566

Moran T, Nudler Y, Bar-Anan Y (2023) Evaluative Conditioning: Past, Present, and Future. *Annual Review of Psychology* 74: 245–269 https://doi.org/10.1146/annurev-psych-032420-031815

Mudde C, Kaltwasser CR (2017) *Populism: A very short introduction*. Oxford University Press, New York https://doi.org/10.1093/actrade/9780190234874.001.0001

Musen MA, Protégé Team (2015) The Protégé Project: A Look Back and a Look Forward. *AI Matters* 1(4): 4–12 https://doi.org/10.1145/2757001.2757003

Musolff A (2015) Dehumanizing Metaphors in UK Immigrant Debates in Press and Online Media. *Journal of Language Aggression and Conflict* 3(1): 41–56 https://doi.org/10.1075/jlac.3.1.02mus

Nascimento IJB do, Pizarro AB, Almeida JM et al (2022) Infodemics and health misinformation: a systematic review of reviews. *Bulletin of the World Health Organization* 100(9): 544–561 https://doi.org/10.2471/BLT.21.287654

Newman H (2022) Foreign Information Manipulation and Interference Defence Standards: Test for Rapid Adoption of the Common Language and Framework DISARM. Hybrid CoE Research Report 7, European Centre of Excellence for Countering Hybrid Threats and NATO Strategic Communications Centre of Excellence, Helsinki. https://stratcomcoe.org/publications/foreign-information-manipulation-and-interference-defence-standards-test-for-rapid-adoption-of-the-common-language-and-framework-disarm-prepared-in-cooperation-with-hybrid-coe/253

Nimmo B, Hutchins E (2023) Phase-Based Tactical Analysis of Online Operations. Technical report, Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2023/03/phase-based-tactical-analysis-of-online-operations

Noggle R (1996) Manipulative Actions: A Conceptual and Moral Analysis. *American Philosophical Quarterly* 33(1): 43–55 https://www.jstor.org/stable/20009846

Noor M, Vollhardt JR, Mari S et al (2017) The social psychology of collective victimhood. *European Journal of Social Psychology* 47(2): 121–134 https://doi.org/10.1002/ejsp.2300

Nuolijärvi P, Tiittula L (2011) Irony in Political Television Debates. *Journal of Pragmatics* 43(2): 572–587 https://doi.org/10.1016/j.pragma.2010.01.019

O'Keefe DJ (1999) How to Handle Opposing Arguments in Persuasive Messages: A Meta-Analytic Review of the Effects of One-Sided and Two-Sided Messages. *Communication Yearbook* 22: 209–249 https://doi.org/10.1080/23808985.1999.11678963

O'Keefe DJ (2016) *Persuasion: Theory and Research*. 3rd edn. SAGE Publications, Thousand Oaks, CA

O'Keefe DJ, Hale SL (1998) The Door-in-the-Face Influence Strategy: A Random-Effects Meta-Analytic Review. *Communication Yearbook* 21(1): 1–33 https://doi.org/10.1080/23808985.1998.11678947

O'Keefe DJ, Jensen JD (2007) The Relative Persuasiveness of Gain-Framed and Loss-Framed Messages for Encouraging Disease Prevention Behaviors: A Meta-Analytic Review. *Journal of Health Communication* 12(7): 623–644 https://doi.org/10.1080/10810730701615198

OECD (2019) *Good Governance for Critical Infrastructure Resilience*. OECD Reviews of Risk Management Policies. OECD Publishing, Paris https://doi.org/10.1787/02f0e5a0-en

Ognyanova K, Lazer D, Robertson RE et al (2020) Misinformation in Action: Fake News Exposure Is Linked to Lower Trust in Media, Higher Trust in Government When Your Side Is in Power. *Harvard Kennedy School Misinformation Review* 1(4): 1–19 https://doi.org/10.37016/mr-2020-024

Opotow S (1990) Moral Exclusion and Injustice: An Introduction. *Journal of Social Issues* 46(1): 1–20 https://doi.org/10.1111/j.1540-4560.1990.tb00268.x

Oswald S, Lewiński M (2014) Pragmatics, Cognitive Heuristics and the Straw Man Fallacy. In: Herman T, Oswald S (eds) *Rhetoric and Cognition: Theoretical Perspectives and Persuasive Strategies*. Peter Lang, Bern, pp 313–343.

Oxford Internet Institute (n.d.) *Computational Propaganda*. https://www.oii.ox.ac.uk/research/projects/computational-propaganda/ Accessed 28 Aug 2026

Pamment J (2020) The EU's Role in Fighting Disinformation: Crafting a Disinformation Framework. Technical report, Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2020/09/the-eus-role-in-fighting-disinformation-crafting-a-disinformation-framework/

Pamment J, Smith V (2022) Attributing Information Influence Operations: Identifying Those Responsible for Malicious Behaviour Online. Technical report, NATO Strategic Communications Centre of Excellence and European Centre of Excellence for Countering Hybrid Threats, Riga. https://stratcomcoe.org/publications/attributing-information-influence-operations-identifying-those-responsible-for-malicious-behaviour-online/244

Pereira A, Berent J, Falomir-Pichastor JM et al (2015) Collective Punishment Depends on Collective Responsibility and Political Organization of the Target Group. *Journal of Experimental Social Psychology* 56: 4–17 https://doi.org/10.1016/j.jesp.2014.09.001

Perkins AW, Forehand MR (2012) Implicit Self-Referencing: The Effect of Nonvolitional Self-Association on Brand and Product Attitude. *Journal of Consumer Research* 39(1): 142–156 https://doi.org/10.1086/662069

Persson P (2018) Attention Manipulation and Information Overload. *Behavioural Public Policy* 2(1): 78–106 https://doi.org/10.1017/bpp.2017.10

Peterson E, Kagalwala A (2021) When Unfamiliarity Breeds Contempt: How Partisan Selective Exposure Sustains Oppositional Media Hostility. *American Political Science Review* 115(2): 585–598 https://doi.org/10.1017/S0003055420001124

Pettigrew TF, Tropp LR (2006) A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology* 90(5): 751–783 https://doi.org/10.1037/0022-3514.90.5.751

Petty RE, Cacioppo JT (1986) The Elaboration Likelihood Model of Persuasion. In: Berkowitz L (ed) *Advances in Experimental Social Psychology*, vol 19. Academic Press, New York, pp 123–205. https://doi.org/10.1016/S0065-2601(08)60214-2

Phi K, Salek Faramarzi N, Wang C et al (2024) Paying Attention to Deflections: Mining Pragmatic Nuances for Whataboutism Detection in Online Discourse. In: *Findings of the Association for Computational Linguistics: ACL 2024*. Association for Computational Linguistics, pp 12628–12643. https://doi.org/10.18653/v1/2024.findings-acl.750

Piskorski J, Dimitrov D, Dobranić F et al (2025) SlavicNLP 2025 Shared Task: Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media. In: *Proceedings of the 10th Workshop on Slavic Natural Language Processing*. Association for Computational Linguistics, pp 254–275. https://doi.org/10.18653/v1/2025.bsnlp-1.27

Piskorski J, Stefanovitch N, Bausier V-A et al (2023a) News Categorization, Framing and Persuasion Techniques: Annotation Guidelines. Technical Report JRC132862, European Commission Joint Research Centre, Ispra, Italy. https://knowledge4policy.ec.europa.eu/text-mining/news-categorization-framing-persuasion-techniques-annotation-guidelines_en

Piskorski J, Stefanovitch N, Da San Martino G et al (2023b) SemEval-2023 Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup. In: Ojha AK, Doğruöz AS, Da San Martino G et al (eds) *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Association for Computational Linguistics, Toronto, Canada, pp 2343–2361. https://doi.org/10.18653/v1/2023.semeval-1.317

Posner MI (1992) Attention as a cognitive and neural system. *Current directions in psychological science* 1(1): 11–14 https://doi.org/10.1111/1467-8721.ep10767759

Prati F, Crisp RJ, Meleady R et al (2016) Humanizing Outgroups through Multiple Categorization: The Roles of Individuation and Threat. *Personality and Social Psychology Bulletin* 42(4): 526–539 https://doi.org/10.1177/0146167216636624

Pratkanis AR (2000) Altercasting as an Influence Tactic. In: Terry DJ, Hogg MA (eds) *Attitudes, Behavior, and Social Context: The Role of Norms and Group Membership*. Lawrence Erlbaum Associates, Mahwah, NJ, pp 201–226.

Pretorius D (2019) SLUIT NOU AAN! South African Union Defence Force Recruitment Posters from the Second World War. *South African Historical Journal* 71(1): 41–69 https://doi.org/10.1080/02582473.2019.1574311

Priolo D, Pelt A, St Bauzel R et al (2019) Three Decades of Research on Induced Hypocrisy: A Meta-Analysis. *Personality and Social Psychology Bulletin* 45(12): 1681–1701 https://doi.org/10.1177/0146167219841621

Prooijen J-W van, Douglas KM (2018) Belief in Conspiracy Theories: Basic Principles of an Emerging Research Domain. *European Journal of Social Psychology* 48(7): 897–908 https://doi.org/10.1002/ejsp.2530

Propp V (2010) *Morphology of the Folktale, 2nd edn. (ed: Wagner LA; trans: Scott L)*. University of Texas Press, Austin, TX

Proust J (2008) Epistemic Agency and Metacognition: An Externalist View. *Proceedings of the Aristotelian Society* 108(3): 241–268 https://doi.org/10.1111/j.1467-9264.2008.00245.x

Publications Office of the European Union (n.d.) *EuroVoc: The EU's Multilingual and Multidisciplinary Thesaurus*. https://eur-lex.europa.eu/browse/eurovoc.html Accessed 28 Aug 2026

Radnitz S, Mylonas H (2022) Introduction: Theorizing Fifth-Column Politics. In: Mylonas H, Radnitz S (eds) *Enemies Within: The Global Politics of Fifth Columns*. Oxford University Press, pp 1–24. https://doi.org/10.1093/oso/9780197627938.003.0001

Rasiah P (2010) A Framework for the Systematic Analysis of Evasion in Parliamentary Discourse. *Journal of Pragmatics* 42(3): 664–680 https://doi.org/10.1016/j.pragma.2009.07.010

Raven BH (1993) The Bases of Power: Origins and Recent Developments. *Journal of Social Issues* 49(4): 227–251 https://doi.org/10.1111/j.1540-4560.1993.tb01191.x

Regan DT (1971) Effects of a Favor and Liking on Compliance. *Journal of Experimental Social Psychology* 7(6): 627–639 https://doi.org/10.1016/0022-1031(71)90025-4

Reichberg GM, Syse H (2018) Threats and Coercive Diplomacy: An Ethical Analysis. *Ethics & International Affairs* 32(2): 179–202 https://doi.org/10.1017/S0892679418000138

Reicher S, Hopkins N (1996) Seeking Influence through Characterizing Self-Categories: An Analysis of Anti-Abortionist Rhetoric. *British Journal of Social Psychology* 35(2): 297–311 https://doi.org/10.1111/j.2044-8309.1996.tb01099.x

Ricœur P (1984) *Time and Narrative, vol 1 (trans: McLaughlin K, Pellauer D)*. University of Chicago Press, Chicago

Rivlin-Angert N, Mor-Lan G (2025) The Enemy from Within: A Study of Political Delegitimization Discourse in Israeli Political Speech. In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Suzhou, China, pp 16634–16647. https://doi.org/10.18653/v1/2025.emnlp-main.841

Roberts ME (2018) *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press, Princeton, NJ https://doi.org/10.23943/9781400890057

Roberts R, Davidai S (2022) The Psychology of Asymmetric Zero-Sum Beliefs. *Journal of Personality and Social Psychology* 123(3): 559–575 https://doi.org/10.1037/pspi0000378

Rogers RW (1975) A Protection Motivation Theory of Fear Appeals and Attitude Change. *The Journal of Psychology* 91(1): 93–114 https://doi.org/10.1080/00223980.1975.9915803

Rogers T, Zeckhauser R, Gino F et al (2017) Artful Paltering: The Risks and Rewards of Using Truthful Statements to Mislead Others. *Journal of Personality and Social Psychology* 112(3): 456–473 https://doi.org/10.1037/pspi0000081

Roozenbeek J, Linden S van der, Goldberg B et al (2022a) Psychological Inoculation Improves Resilience Against Misinformation on Social Media. *Science Advances* 8(34): eabo6254 https://doi.org/10.1126/sciadv.abo6254

Roozenbeek J, Traberg CS, Linden S van der (2022b) Technique-Based Inoculation Against Real-World Misinformation. *Royal Society Open Science* 9(5): 211719 https://doi.org/10.1098/rsos.211719

Roselle L, Miskimmon A, O'Loughlin B (2014) Strategic narrative: A new means to understand soft power. *Media, War & Conflict* 7(1): 70–84 https://doi.org/10.1177/1750635213516696

Rothman AJ, Salovey P (1997) Shaping Perceptions to Motivate Healthy Behavior: The Role of Message Framing. *Psychological Bulletin* 121(1): 3–19 https://doi.org/10.1037/0033-2909.121.1.3

Rothschild ZK, Landau MJ, Sullivan D et al (2012) A Dual-Motive Model of Scapegoating: Displacing Blame to Reduce Guilt or Increase Control. *Journal of Personality and Social Psychology* 102(6): 1148–1163 https://doi.org/10.1037/a0027413

Różycka-Tran J, Boski P, Wojciszke B (2015) Belief in a Zero-Sum Game as a Social Axiom: A 37-Nation Study. *Journal of Cross-Cultural Psychology* 46(4): 525–548 https://doi.org/10.1177/0022022115572226

Rubin M, Badea C (2007) Why Do People Perceive Ingroup Homogeneity on Ingroup Traits and Outgroup Homogeneity on Outgroup Traits? *Personality and Social Psychology Bulletin* 33(1): 31–42 https://doi.org/10.1177/0146167206293190

Rubin M, Badea C (2012) They're All the Same!... but for Several Different Reasons: A Review of the Multicausal Nature of Perceived Group Variability. *Current Directions in Psychological Science* 21(6): 367–372 https://doi.org/10.1177/0963721412457363

Sandri D, Grigoli F, Gorodnichenko Y et al (2023) Keep Calm and Bank On: Panic-Driven Bank Runs and the Role of Public Communication. Tech. rep. 31644, National Bureau of Economic Research. https://doi.org/10.3386/w31644

Sani F, Bowe M, Herrera M et al (2007) Perceived Collective Continuity: Seeing Groups as Entities That Move Through Time. *European Journal*

of Social Psychology 37(6): 1118–1134 https://doi.org/10.1002/ejsp.430

Sanz-Menéndez L, Cruz-Castro L (2019) The Credibility of Scientific Communication Sources Regarding Climate Change: A Population-Based Survey Experiment. *Public Understanding of Science* 28(5): 534–553 https://doi.org/10.1177/0963662519840946

Saul JM (2012) *Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics*. Oxford University Press, Oxford https://doi.org/10.1093/acprof:oso/9780199603688.001.0001

Schatz RT, Lavine H (2007) Waving the Flag: National Symbolism, Social Identity, and Political Engagement. *Political Psychology* 28(3): 329–355 https://doi.org/10.1111/j.1467-9221.2007.00571.x

Scheff TJ (1988) Shame and Conformity: The Deference-Emotion System. *American Sociological Review* 53(3): 395–406 https://doi.org/10.2307/2095647

Schelling TC (1966) *Arms and Influence*. Yale University Press, New Haven, CT https://doi.org/10.2307/j.ctt5vm52s

Schokkenbroek JM, Snaphaan T (2025) Love as Bait: A Scoping Review and Crime Script Analysis of Online Romance Scams. *Trauma, Violence, & Abuse* Advance online publication: https://doi.org/10.1177/15248380251361046

Schreurs S (2021) Those Were the Days: Welfare Nostalgia and the Populist Radical Right in the Netherlands, Austria and Sweden. *Journal of International and Comparative Social Policy* 37(2): 128–141 https://doi.org/10.1017/ics.2020.30

Schultz PW, Nolan JM, Cialdini RB et al (2007) The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science* 18(5): 429–434 https://doi.org/10.1111/j.1467-9280.2007.01917.x

Séville A (2017) From 'One Right Way' to 'One Ruinous Way'? Discursive Shifts in 'There Is No Alternative'. *European Political Science Review* 9(3): 449–470 https://doi.org/10.1017/S1755773916000035

Shavell S (1993) An Economic Analysis of Threats and Their Illegality: Blackmail, Extortion, and Robbery. *University of Pennsylvania Law Review* 141(5): 1877–1903 https://doi.org/10.2307/3312577

Sherif M, Harvey OJ, White BJ et al (1961) *Intergroup Conflict and Cooperation: The Robbers Cave Experiment*. University Book Exchange, Norman, OK

Shorten R (2011) Towards a Political Rhetoric of Wrongdoing: The Case of Moral Equivalence. *Journal of Political Ideologies* 16(2): 195–219 https://doi.org/10.1080/13569317.2011.575684

Shulman HC, Dixon GN, Bullock OM et al (2020) The Effects of Jargon on Processing Fluency, Self-Perceptions, and Scientific Engagement. *Journal*

of Language and Social Psychology 39(5–6): 579–597 https://doi.org/10.1177/0261927X20902177

Simons DJ, Chabris CF (1999) Gorillas in our midst: Sustained inattentional blindness for dynamic events. *perception* 28(9): 1059–1074 https://doi.org/10.1068/p281059

Skurka C (2019) You Mad? Using Anger Appeals to Promote Activism Intentions and Policy Support in the Context of Sugary Drink Marketing to Kids. *Health Communication* 34(14): 1775–1787 https://doi.org/10.1080/10410236.2018.1536943

Smeekes A, Verkuyten M (2013) Collective Self-Continuity, Group Identification and In-Group Defense. *Journal of Experimental Social Psychology* 49(6): 984–994 https://doi.org/10.1016/j.jesp.2013.06.004

Smeekes A, Wildschut T, Sedikides C (2021) Longing for the "Good Old Days" of Our Country: National Nostalgia as a New Master-Frame of Populist Radical Right Parties. *Journal of Theoretical Social Psychology* 5(2): 90–102 https://doi.org/10.1002/jts5.78

Smith ER, Mackie DM (2016) Group-Level Emotions. *Current Opinion in Psychology* 11: 15–19 https://doi.org/10.1016/j.copsyc.2016.04.005

Solt S (2015) Vagueness and Imprecision: Empirical Foundations. *Annual Review of Linguistics* 1: 107–127 https://doi.org/10.1146/annurev-linguist-030514-125150

Sontag S (1978) *Illness as Metaphor*. Farrar, Straus and Giroux, New York

Spear AD (2023) Epistemic Dimensions of Gaslighting: Peer-Disagreement, Self-Trust, and Epistemic Injustice. *Inquiry* 66(1): 68–91 https://doi.org/10.1080/0020174X.2019.1610051

Spears R (2021) Social Influence and Group Identity. *Annual Review of Psychology* 72: 367–390 https://doi.org/10.1146/annurev-psych-070620-111818

Stephan WG, Ybarra O, Rios Morrison K (2009) Intergroup Threat Theory. In: Nelson TD (ed) *Handbook of Prejudice, Stereotyping, and Discrimination*. Psychology Press, New York, pp 43–60.

Stern C, Ondish P (2018) Political Aspects of Shared Reality. *Current Opinion in Psychology* 23: 11–14 https://doi.org/10.1016/j.copsyc.2017.11.004

Stocking SH, Holstein LW (2009) Manufacturing Doubt: Journalists' Roles and the Construction of Ignorance in a Scientific Controversy. *Public Understanding of Science* 18(1): 23–42 https://doi.org/10.1177/0963662507079373

Stone J, Aronson E, Crain AL et al (1994) Inducing Hypocrisy as a Means of Encouraging Young Adults to Use Condoms. *Personality and Social Psychology Bulletin* 20(1): 116–128 https://doi.org/10.1177/0146167294201012

Stone J, Wiegand AW, Cooper J et al (1997) When Exemplification Fails: Hypocrisy and the Motive for Self-Integrity. *Journal of Personality and Social Psychology* 72(1): 54–65 https://doi.org/10.1037/0022-3514.72.1.54

Strutzenberg CC, Wiersma-Mosley JD, Jozkowski KN et al (2017) Love-Bombing: A Narcissistic Approach to Relationship Formation. *Discovery, The Student Journal of Dale Bumpers College of Agricultural, Food and Life Sciences* 18(1): 81–89 https://doi.org/10.54119/discovery.zxgc9960

Stukal D, Sanovich S, Bonneau R et al (2022) Why Botter: How Pro-Government Bots Fight Opposition in Russia. *American Political Science Review* 116(3): 843–857 https://doi.org/10.1017/S0003055421001507

Sumner P, Vivian-Griffiths S, Boivin J et al (2014) The Association between Exaggeration in Health Related Science News and Academic Press Releases: Retrospective Observational Study. *BMJ* 349: g7015 https://doi.org/10.1136/bmj.g7015

Sunstein CR, Vermeule A (2009) Conspiracy Theories: Causes and Cures. *Journal of Political Philosophy* 17(2): 202–227 https://doi.org/10.1111/j.1467-9760.2008.00325.x

Supran G, Oreskes N (2017) Assessing ExxonMobil's climate change communications (1977–2014). *Environmental research letters* 12(8): 084019 https://doi.org/10.1088/1748-9326/aa815f

Susser D, Roessler B, Nissenbaum H (2019) Online Manipulation: Hidden Influences in a Digital World. *Georgetown Law Technology Review* 4(1): 1–45 https://georgetownlawtechreview.org/online-manipulation-hidden-influences-in-a-digital-world/GLTR-01-2020/

Svolik MW (2019) Polarization versus democracy. *Journal of Democracy* 30(3): 20–32 https://doi.org/10.1353/jod.2019.0039

Swann Jr. WB, Gómez Á, Huici C et al (2010) Identity Fusion and Self-Sacrifice: Arousal as a Catalyst of Pro-Group Fighting, Dying, and Helping Behavior. *Journal of Personality and Social Psychology* 99(5): 824–841 https://doi.org/10.1037/a0020014

Tajfel H, Turner JC (1979) An Integrative Theory of Intergroup Conflict. In: Austin WG, Worchel S (eds) *The Social Psychology of Intergroup Relations*. Brooks/Cole, pp 33–47.

Tangney JP, Stuewig J, Mashek DJ (2007) Moral Emotions and Moral Behavior. *Annual Review of Psychology* 58: 345–372 https://doi.org/10.1146/annurev.psych.56.091103.070145

Tannenbaum MB, Hepler J, Zimmerman RS et al (2015) Appealing to Fear: A Meta-Analysis of Fear Appeal Effectiveness and Theories. *Psychological Bulletin* 141(6): 1178–1204 https://doi.org/10.1037/a0039729

Technology and Social Change Project (n.d.) *The Media Manipulation Casebook: Code Book*. https://mediamanipulation.org/code-book/ Accessed 28 Aug 2026

Tessler MH, Bakker MA, Jarrett D et al (2024) AI Can Help Humans Find Common Ground in Democratic Deliberation. *Science* 386(6719): eadq2852 https://doi.org/10.1126/science.adq2852

Thibodeau PH, Boroditsky L (2011) Metaphors We Think With: The Role of Metaphor in Reasoning. *PLOS ONE* 6(2): e16782 https://doi.org/10.1371/journal.pone.0016782

Thoma D, Becker K, Kißler A (2023) Presuppositions Are More Persuasive than Assertions if Addressees Accommodate Them: Experimental Evidence for Philosophical Reasoning. *Applied Psycholinguistics* 44(5): 816–843 https://doi.org/10.1017/S0142716423000292

Tilley J, Bejan T, Hobolt SB (2026) Partisan (In)Tolerance and Affective Polarization. *British Journal of Political Science* 56: e36 https://doi.org/10.1017/S0007123426101550

Todd AR, Bodenhausen GV, Richeson JA et al (2011) Perspective Taking Combats Automatic Expressions of Racial Bias. *Journal of Personality and Social Psychology* 100(6): 1027–1042 https://doi.org/10.1037/a0022308

Tomz M, Weeks JLP (2021) Military Alliances and Public Support for War. *International Studies Quarterly* 65(3): 811–824 https://doi.org/10.1093/isq/sqab015

Toulmin SE (2003) *The Uses of Argument*. 2nd edn. Cambridge University Press, Cambridge https://doi.org/10.1017/CBO9780511840005

Turner JC, Wetherell MS, Hogg MA (1989) Referent Informational Influence and Group Polarization. *British Journal of Social Psychology* 28(2): 135–147 https://doi.org/10.1111/j.2044-8309.1989.tb00855.x

Turner MM, Richards AS, Bessarabova E et al (2020) The Effects of Anger Appeals on Systematic Processing and Intentions: The Moderating Role of Efficacy. *Communication Reports* 33(1): 14–26 https://doi.org/10.1080/08934215.2019.1682175

Turner MM, Banas JA, Rains SA et al (2010) The Effects of Altercasting and Counterattitudinal Behavior on Compliance: A Lost Letter Technique Investigation. *Communication Reports* 23(1): 1–13 https://doi.org/10.1080/08934211003598759

Tversky A, Kahneman D (1981) The Framing of Decisions and the Psychology of Choice. *Science* 211(4481): 453–458 https://doi.org/10.1126/science.7455683

Tyler TR (1988) What Is Procedural Justice? Criteria Used by Citizens to Assess the Fairness of Legal Procedures. *Law & Society Review* 22(1): 103–135 https://doi.org/10.2307/3053563

Tyler TR (1994) Governing amid Diversity: The Effect of Fair Decisionmaking Procedures on the Legitimacy of Government. *Law & Society Review* 28(4): 809–831 https://doi.org/10.2307/3053998

Tyler TR (2000) Social Justice: Outcome and Procedure. *International Journal of Psychology* 35(2): 117–125 https://doi.org/10.1080/00207590 0399411

Tyler TR (2006) Psychological Perspectives on Legitimacy and Legitimation. *Annual Review of Psychology* 57: 375–400 https://doi.org/10.1146/a nnurev.psych.57.102904.190038

United Nations (2019) *United Nations Strategy and Plan of Action on Hate Speech*. https://www.un.org/en/hate-speech/un-strategy-and-pla n-of-action-on-hate-speech Accessed 28 Aug 2026

Valtorta RR, Baldissarri C, Volpato C (2026) Metaphors Matter: The Effect of Biological Dehumanizing Language on Indirect Aggression. *Current Psychology* 45: 884 https://doi.org/10.1007/s12144-026-09290-4

Verhulsdonk I, Nai A, Karp JA (2022) Are Political Attacks a Laughing Matter? Three Experiments on Political Humor and the Effectiveness of Negative Campaigning. *Political Research Quarterly* 75(3): 720–737 https://doi.org/10.1177/10659129211023590

Vescio TK, Sechrist GB, Paolucci MP (2003) Perspective Taking and Prejudice Reduction: The Mediational Role of Empathy Arousal and Situational Attributions. *European Journal of Social Psychology* 33(4): 455–472 https://doi.org/10.1002/ejsp.163

Vezzali L, Di Bernardo GA, Stathi S et al (2019) Using Intercultural Videos of Direct Contact to Implement Vicarious Contact: A School-Based Intervention That Improves Intergroup Attitudes. *Group Processes & Intergroup Relations* 22(7): 1059–1076 https://doi.org/10.1177/13684 30218809885

Wagoner B, Brescó de Luna I (2021) Collective Grief: Mourning Rituals, Politics and Memorial Sites. In: Køster A, Holte Kofod E (eds) *Cultural, Existential and Phenomenological Dimensions of Grief Experience*. Routledge, pp 197–213. https://doi.org/10.4324/9781003099420-17

Walter N, Tukachinsky R, Pelled A et al (2019) Meta-Analysis of Anger and Persuasion: An Empirical Integration of Four Models. *Journal of Communication* 69(1): 73–93 https://doi.org/10.1093/joc/jqy054

Walton D (1992) *Slippery Slope Arguments*. Oxford University Press, Oxford

Walton D (1996) The Straw Man Fallacy. In: Benthem J van, Eemeren FH van, Grootendorst R et al (eds) *Logic and Argumentation*. North-Holland, Amsterdam, pp 115–128.

Walton D (1997) *Appeal to Pity: Argumentum ad Misericordiam*. State University of New York Press, Albany, NY

Walton D (1998) *Ad Hominem Arguments*. Studies in Rhetoric and Communication. University of Alabama Press, Tuscaloosa, AL

Walton D (2004) Classification of Fallacies of Relevance. *Informal Logic* 24(1): 71–103 https://doi.org/10.22329/il.v24i1.2133

Walton D (2005) *Fundamentals of Critical Argumentation*. Cambridge University Press, Cambridge https://doi.org/10.1017/CBO9780511807039

Walton D (2008) Personal Attack in Argumentation. In: *Informal Logic: A Pragmatic Approach*, 2nd edn. Cambridge University Press, Cambridge, pp 170–208.

Walton D (2015) The Basic Slippery Slope Argument. *Informal Logic* 35(3): 273–311 https://doi.org/10.22329/il.v35i3.4286

Walton D, Hyra C (2018) Analogical Arguments in Persuasive and Deliberative Contexts. *Informal Logic* 38(2): 213–262 https://doi.org/10.22329/il.v38i2.4805

Walton D, Reed C, Macagno F (2008) *Argumentation Schemes*. Cambridge University Press, Cambridge https://doi.org/10.1017/CBO9780511802034

Wang MY, Silva DE (2018) A Slap or a Jab: An Experiment on Viewing Uncivil Political Discussions on Facebook. *Computers in Human Behavior* 81: 73–83 https://doi.org/10.1016/j.chb.2017.11.041

Wardle C, Derakhshan H (2017) Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Report DGI(2017)09, Council of Europe, Strasbourg. https://rm.coe.int/information-disorder-report-2017/1680766412

Waxman MC (2014) The Power to Threaten War. *Yale Law Journal* 123: 1626–1691

Weber M (1947) *The Theory of Social and Economic Organization (ed: Parsons T; trans: Henderson AM, Parsons T)*. Oxford University Press, New York

Weinstein EA, Deutschberger P (1963) Some Dimensions of Altercasting. *Sociometry* 26(4): 454–466 https://doi.org/10.2307/2786148

Westfall J, Van Boven L, Chambers JR et al (2015) Perceiving Political Polarization in the United States: Party Identity Strength and Attitude Extremity Exacerbate the Perceived Partisan Divide. *Perspectives on Psychological Science* 10(2): 145–158 https://doi.org/10.1177/1745691615569849

Wheeler SC, Petty RE, Bizer GY (2005) Self-Schema Matching and Attitude Change: Situational and Dispositional Determinants of Message Elaboration. *Journal of Consumer Research* 31(4): 787–797 https://doi.org/10.1086/426613

White H (1973) *Metahistory: The Historical Imagination in Nineteenth-Century Europe*. Johns Hopkins University Press, Baltimore

Whitehouse H, Lanman JA (2014) The Ties That Bind Us: Ritual, Fusion, and Identification. *Current Anthropology* 55(6): 674–695 https://doi.org/10.1086/678698

Whitmeyer JM (2002) The Compliance You Need for a Cost You Can Afford: How to Use Individual and Collective Sanctions? *Social Science Research* 31(4): 630–652 https://doi.org/10.1016/S0049-089X(02)00017-0

Wildschut T, Bruder M, Robertson S et al (2014) Collective Nostalgia: A Group-Level Emotion That Confers Unique Benefits on the Group. *Journal of Personality and Social Psychology* 107(5): 844–863 https://doi.org/10.1037/a0037760

Witte K, Allen M (2000) A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Campaigns. *Health Education & Behavior* 27(5): 591–615 https://doi.org/10.1177/109019810002700506

Wodak R (2015) *The Politics of Fear: What Right-Wing Populist Discourses Mean*. Sage, London https://doi.org/10.4135/9781446270073

Wood JL, James M, Ó Ciardha C (2014) 'I Know How They Must Feel': Empathy and Judging Defendants. *The European Journal of Psychology Applied to Legal Context* 6(1): 37–43 https://doi.org/10.5093/ejpalc2014a5

Worchel S, Lee J, Adewole A (1975) Effects of Supply and Demand on Ratings of Object Value. *Journal of Personality and Social Psychology* 32(5): 906–914 https://doi.org/10.1037/0022-3514.32.5.906

Wright SC, Aron A, McLaughlin-Volpe T et al (1997) The Extended Contact Effect: Knowledge of Cross-Group Friendships and Prejudice. *Journal of Personality and Social Psychology* 73(1): 73–90 https://doi.org/10.1037/0022-3514.73.1.73

Ye S, Attali D, Ghazi M et al (2026) Systematic Review and Meta-analysis of the Evidence for an Illusory Truth Effect and Its Determinants. *Nature Communications* 17: 3270 https://doi.org/10.1038/s41467-026-70041-x

Zappavigna M (2022) Social Media Quotation Practices and Ambient Affiliation: Weaponising Ironic Quotation for Humorous Ridicule in Political Discourse. *Journal of Pragmatics* 191: 98–112 https://doi.org/10.1016/j.pragma.2021.12.003

Zillmann D (1999) Exemplification Theory: Judging the Whole by Some of Its Parts. *Media Psychology* 1(1): 69–94 https://doi.org/10.1207/s1532785xmep0101_5

Zomeren M van, Leach CW, Spears R (2010) Does Group Efficacy Increase Group Identification? Resolving Their Paradoxical Relationship. *Journal of Experimental Social Psychology* 46(6): 1055–1060 https://doi.org/10.1016/j.jesp.2010.05.006

Zwaan RA, Langston MC, Graesser AC (1995) The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. *Psychological Science* 6(5): 292–297 https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

Zwaan RA, Radvansky GA (1998) Situation Models in Language Comprehension and Memory. *Psychological Bulletin* 123(2): 162–185 https://doi.org/10.1037/0033-2909.123.2.162