Decoding Influence Narratives: Identifying Persuasion and Propaganda Techniques in Twitter Discourse on Violent Extremism

Shu Jia Chee[1], Sara Rubini[1], Paul Gill[1], Enrico Mariconti[1]

[1]University College London

shu.chee23@ucl.ac.uk, sara.rubini.19@ucl.ac.uk, paul.gill@ucl.ac.uk, e.mariconti@ucl.ac.uk

Abstract

Identifying persuasion and propaganda techniques is crucial for understanding message agendas and designing culturally resonant counter-messages, yet their overlapping nature complicates both qualitative annotation and automated detection efforts. Previous research often examined persuasion and propaganda separately, overlooking their frequent convergence in everyday communication and in digital contexts. To address this gap, our study developed a codebook of 20 unique influence techniques by integrating operationalisations from prior analyses of radical Islamic publications with insights from tweets discussing the 2020 Charlie Hebdo–related terror incidents. As techniques often co-occur, we clustered them to uncover underlying intent and context, distilling the 20 techniques into seven overarching strategies, four of which centred on authority and fallacy-based invocations. A cross-cultural analysis of tweets from five Asian Islamic subgroups then examined differences in susceptibility to these strategies, moving beyond the supply of extremist messaging to reveal the demand-side dynamics of what resonated with audiences. Results revealed significant variation: the Arabic subgroup, for instance, was more susceptible to religious appeals and misinformation-related fallacies than other subgroups. These findings demonstrate that influence strategies are not universally applicable, underscoring the importance of tailoring counter-propaganda messaging to the cultural values of specific audiences.

Introduction

In light of recent global crises such as the Israel-Hamas war and the US-China trade war, information campaigns involving persuasion, propaganda, and misinformation proliferated (Cherry, 2024). Yet distinguishing and classifying techniques used in these campaigns remains highly challenging due to ambiguous definitions and overlapping categories (Abdullah et al., 2022).

While seminal frameworks such as Cialdini's persuasion principles (Cialdini, 2006), Institute for Propaganda Analysis (IPA)'s classic propaganda techniques (Lee & Lee, 1939), and the more recent SemEval propaganda taxonomies (Da San Martino et al., 2019) provide valuable foundations, persuasion and propaganda techniques are non-exclusive and often overlap in definition and application thereby complicating efforts to draw clear distinctions (Blodgett et al., 2023). This also reflects a broader convergence: although persuasion and propaganda differ in their theoretical conceptualisations, they share the same goal of shaping attitudes (Stiff & Mongeau, 2016) and in practice often manifest in similar ways (Soules, 2015). The social media environment where intentions are ambiguous further amplifies this convergence, since the role once reserved for state actors and institutions can now be assumed by any user acting as a 'persuader' or 'propagandist.'

However, existing research has largely overlooked this shift. Examination of persuasion and propaganda usage in violent extremism, for example, have mainly focused on qualitative analysis of radical groups' (e.g., ISIS) media outputs (Zgryziewicz et al., 2016; Aggarwal, 2021), while automated detection has been applied mostly to political news by mainstream media (Abdullah et al., 2022; Teimas & Saias, 2023; Szwoch et al., 2024). Few studies have examined how everyday social media users employ these techniques. Even in computational contexts, progress has been limited. Despite advances from SemEval-2020 and transformer-based approaches, fine-grained technique detection efforts are hampered by poor inter-annotator agreement, unclear operationalisations, and the under-representation of certain techniques in existing codebooks (Abdullah et al., 2022; Blodgett et al., 2023). Reliable classification of propaganda techniques hence remains elusive, with F1 scores ranging between .18–.62 (Da San Martino et al., 2020a).

Given that conceptual and practical overlaps between persuasion and propaganda can create downstream challenges for annotation and detection, examining them together, rather than separately as past studies have done, offers greater analytic value. A standardised approach to

defining, classifying, and annotating influence techniques is therefore needed to improve interpretability and detection accuracy. Accordingly, the first aim of our study is to develop a coding framework to characterise influential narratives in the online extremism context. Specifically, we set out to answer the following research questions (RQs):

- How can influential narratives be systematically characterised through the identification of influence techniques in tweets? (**RQ1**)
- How can the identified influence techniques be refined to improve understanding of influential narratives? (**RQ2**)

In addition, while cultural values significantly shape responses to persuasion and propaganda, most research has presumed behavioural outcomes to be universally similar across cultural dimensions (Rodrigues et al., 2018). This leads to the second research aim: investigating whether cross-cultural differences exist in the reception of influence techniques in the online extremism.

- Are there cross-cultural differences in the susceptibility to influence techniques among Islamic subgroups in Asia? (**RQ3**)

To answer these RQs, we analysed the influence techniques found in tweets about the 2020 Charlie Hebdo terror incidents. Techniques that had multiple dimensions were expanded while those with overlapping definitions and operationalisations were integrated, resulting in the consolidation of seven Cialdini's persuasion principles and 25 IPA/SemEval's propaganda techniques into 20 unique influence techniques in our codebook. TThese were further clustered into seven overarching strategies to capture the synergistic effects of co-occurring techniques to enhance interpretation. A multinomial logistic regression was then conducted to examine differences in technique susceptibilities among five Asian Islamic subgroups (English, Turkish, Indonesian, Pakistani, and Arabic) to inform the design of targeted and resonant counter-messages.

Our contributions are threefold: First, we developed a comprehensive codebook that eliminates overlaps and reduces misclassification risks, advancing future annotation and detection of influence techniques. Second, by clustering co-occurring techniques into seven overarching strategies, we provide deeper insights into the intent and context behind their use. Unlike previous methods that might simply identify *loaded language* as the dominant technique, the current approach allows for more nuanced conclusions by showing how *loaded language* functions within generalisation and contrast-based strategies to instil fear and disparage opponents. Third, our comparative cross-cultural analysis offers actionable insights for designing counter-messages that account for cultural diversity within and across subgroups.

Literature Review

Theoretical Foundations of Persuasion and Propaganda. Persuasion and propaganda are often conflated yet differ in intent, process, and operational characteristics (Jowett & O'Donnell, 2012). Persuasion is generally defined as interactive and transactional communication that seeks desirable changes in attitudes by leveraging trust, reciprocity, and alignment with audience beliefs or "anchors" (Smith et al., 2014; Soules, 2015). It may reinforce existing views or shape new behaviours, but in all cases relies on perceived mutual benefit and audience acceptance (Jowett & O'Donnell, 2012). When communication serves only the communicator's interests and exploits audience trust, persuasion veers into propaganda (Soules, 2015). Traditionally associated with authoritarian regimes and wartime campaigns (O'Shaughnessy, 2004), propaganda has been described as agitative, fuelling division and anger (Ellul, 1965), or integrative, promoting conformity and passivity (Szanto, 1977). It has also been classified as *black*, *grey*, or *white*, depending on its degree of covertness and truthfulness (Vaughn et al., 1989). Contemporary scholarship largely regards propaganda as an insidious form of persuasion that serves the propagandist (Soules, 2015). Today it most closely resembles the agitative type, mobilising audiences through emotional manipulation, half-truths, and cultural symbols, often within or provoking crises (Farkas & Bastos, 2018; Oddo, 2019).

Persuasion and Propaganda Techniques. Correspondingly, persuasion and propaganda techniques refer to influence mechanisms used to reinforce or mobilise attitudes and behaviours (Vaughn et al., 1989). Cialdini (2001) described techniques as heuristics, or emotional and belief-based shortcuts that facilitate rapid decision-making. In high-volume online environments, especially during conflicts, these techniques can be exploited to enhance message effectiveness through pe mripheral influence (Soules, 2015). Scholars have proposed numerous taxonomies, and their application to extremist communication by terrorist groups such as ISIS, can demonstrate how these techniques are deployed in practice.

Social psychologists have advanced several persuasion taxonomies (Kellermann & Cole, 1994; Fogg, 2003; Cialdini, 2006; 2009). Among these, Cialdini's principles have been recognised as "universal persuasion principles" (Sofia et al., 2016) and the "weapons of influence" (Alslaity & Tran, 2020). These principles are: *reciprocity*, *commitment and consistency*, *social proof*, *authority*, *scarcity*, *liking*, and *unity*. Qualitative analysis of ISIS content like the Dabiq magazine and their social media posts found calls for *commitment* to action and *authority* appeals through scriptural and leadership references most common (Zgryziewicz et al., 2016; Qutteineh, 2017; Aggarwal,

2021). For example, 'Allah' and Qur'an verses were cited over 200 and 50 times in 30 *Dabiq* articles, often out of context to justify violence or legitimise Abu Bakr al-Baghdadi as the Caliph (Qutteineh, 2017).

Propaganda scholarship has produced parallel frameworks. Lee & Lee (1939)'s IPA techniques outlined seven "ABCs of propaganda analysis," later followed by Brown (1963) and Shabo (2008). The IPA framework remains most influential (Jowett & O'Donnell, 2015), comprising *name-calling*, *glittering generalities*, *transfer*, *testimonial*, *plain folks*, *card stacking*, and *bandwagon*. Applied to ISIS materials (Qutteineh, 2017; Saehuddin & Rdha, 2022; Lakomy, 2022), these techniques revealed frequent reliance on *name-calling*, *glittering generalities*, and *transfer* to disparage adversaries and promote moral superiority. To illustrate, recruitment videos had selectively quoted revered Islamic leaders and scholars to transfer legitimacy to ISIS's leadership claims. Conversely, they used negative transfer by branding the US government as 'satanic' and associating the Saudia Arabia regime with Zionism to discredit them (Lakomy, 2022).

With the advent of digital media and varied propaganda mediums, scholars extended classic frameworks for computational detection. Da San Martino et al. (2019, 2020) introduced 18 SemEval techniques that integrate rhetorical devices and logical fallacies. Some are repeated IPA categories (e.g., *name-calling*, *bandwagon*), while others like *appeals to authority* and *reductio ad Hitlerum* refine or expand them. These newer taxonomies provide finer granularity for automated detection tasks and complement traditional models in contemporary contexts. The full list of Cialdini, IPA, and SemEval techniques, along with their definitions can be found in Appendix A Table T1, showing that despite differences in theoretical conceptualisations and design contexts, overlaps in their applications are evident.

Culture and the 'Universality' of Influence Techniques. Culture can significantly shape how messages are framed and received (Soules, 2015). Cultural dimensions like collectivism-individualism and power distance (Hofstede, 2001) influence emotional and behavioural outcomes, with survey studies using Kaptein et al. (2012) *Susceptibility to Persuasive Strategies Scale* showing significant cultural variation in susceptibility to persuasion and propaganda (Rodrigues et al., 2018). For instance, East Asians (collectivistic) are more responsive to *reciprocity*, *consensus*, and *authority* than North Americans (individualistic) (Orji, 2016). In high power distance cultures, *transfer* and *repetition* are more effective, as authority is more readily accepted (Sample et al., 2018).

Importantly, cultural values are not monolithic. Within similar cultural dimensions, susceptibility differences emerge: Dutch participants, though from an individualistic culture like France, showed greater susceptibility to *authority*-based messages (Hornikx & Hoeken, 2007). Likewise, East Asian cultures emphasise harmony and modesty, encouraging *conformity* (Leung & Cohen, 2011), while Arabic cultures emphasise honour and collective strength, making *unity* and *flag-waving* more persuasive (Güngör et al., 2014). Variation exists even within cultural or religious subgroups. Alnunu et al. (2021) found that Arab Muslims' susceptibility to *authority* differed across regions (i.e., Arabic countries, non-Arabic Muslim countries, and Western societies) though responses to other techniques were consistent. Nonetheless, most evidence comes from self-reported survey data, which may not reflect actual behavioural responses.

Challenges in Technique Detection. Over the past decade, research in persuasion and propaganda detection has grown alongside interest in fake news. Early work primarily labelled entire news outlets or articles as propagandistic (Garimella et al., 2015), and analysed propaganda at the document level (Rashkin et al., 2017). A major advance came with SemEval-2020 Task 11, which introduced 18 propaganda techniques at the sentence and fragment level, along with an annotated corpus and BERT-based baselines (Da San Martino et al., 2020a). This benchmark has since fuelled machine learning studies using transformers and LLMs in technique detection (Blodgett et al., 2023).

Despite progress, fine-grained detection remains challenging. Even the best SemEval models achieved modest F1 scores around .52 for span identification[1] and .18–.62 for technique classification[2] (Da San Martino et al., 2020a). Later transformer ensembles reported similar ranges (.59–.60) (Abdullah et al., 2022; Teimas & Saias, 2023). Recent tests of LLMs on political news such as *gpt-3.5-turbo* and *gpt-4-0125-preview* performed comparably on coarse binary classification but struggled with span-level detection and multi-class technique classification, with F1 scores often ranging from .19–.30 (Sprenkamp et al., 2023; Jones, 2024; Szwoch et al., 2024). By contrast, coarse-grained models show stronger results for initial binary classification[3] of news, with XMLRoBERTa achieving .67–.82 at the paragraph level (Nikolaidis et al., 2023).

Several factors explain these challenges. First, significant data imbalance in the annotated corpora and datasets: underrepresented techniques like *whataboutism* and *red herring* yielded F1 scores below .30, while frequent ones like *loaded language* and *name-calling* exceeded .600 (Da San Martino et al., 2020a). Second, annotation subjectivity:

[1] Identification of the specific text segments in which propaganda is located

[2] Identification of the specific technique used in the highlighted text segments

[3] Identification of the presence or absence of propaganda technique

low inter-annotator agreement introducing inconsistencies that may compromise the reliability of the training data (Bonial et al., 2022; Blodgett et al., 2023). Third, overlaps between techniques: complexity increases when shifting from coarse to fine-grained detection, where confusion between similar and overlapping techniques (e.g., *whataboutism* vs. *doubt*) becomes more pronounced, turning the task into a multi-label classification problem (Nikolaidis et al., 2023). Finally, the scarcity of large, open-access, consistently annotated corpora further exacerbates the abovementioned issues and limits model performance (Abdullah et al., 2022; Da San Martino et al., 2020b).

Data

Our data comprised tweets related to the beheading of Samuel Paty on 16 October 2020, and the knife attack in Nice on 29 October 2020, after caricatures of Prophet Muhammad were re-published in the *Charlie Hebdo* magazine (BBC, 2020a). The dataset was collected in November 2020 using the Twitter API, retrieving standard fields like tweets, language, and retweet counts. To avoid biasing the dataset toward specific perspectives (i.e., pro-/anti- stances), generic search queries like 'France' and '#CharlieHebdo' were used. This yielded 560,807 tweets in 49 languages from 97,170 Southeast Asian and South Asian accounts, posted between 16 October–10 November 2020. The collection period coincided with widespread protests across the region (BBC, 2020b), providing a unique opportunity to analyse how users employed and engaged online persuasion and propaganda in response to the highly charged offline events.

Methodology

Data Pre-processing

As only 23.2% of tweets were geotagged and 58.7% of users reported recognisable locations, we used tweet language as a proxy for location. Our analysis focused on the top five languages (English, Turkish, Bahasa Indonesian, Urdu, Arabic), which comprised 87.0% of tweets. While languages like Turkish and Indonesian likely indicated geographic origin, English tweets were retained for their prevalence (47.5%, N=266,660) and possible appeal to a broader or Western audience, potentially reflecting different influence agendas (Lee, 2015). Non-English tweets were translated using the Opus-MT model via EasyNMT[4] for its computational efficiency and quality of translations[5] (Tiedemann & Thottingal, 2020). Tweets were then cleaned with the *tweet-preprocessor* library[6] and standardised with NLTK[7]. To focus on influential original messages, we excluded retweets and those with fewer than three words post-processing. From the remaining pool, the top 5% of tweets by retweet count in each language were coded, and 5,982 tweets containing at least one technique were retained for RQ1–RQ3 (Table 1).

Lang	Tweets	Original tweets	Post-processed	Coded tweets	Retained tweets
EN	266,660	116,664	86,845	4352	3518
TR	104,394	29,488	23,079	1154	857
IN	70,372	21,374	17,983	899	726
UR	29,793	16,368	11,368	568	478
AR	16,612	11,495	8859	443	403
Total	487,831	195,389	148,134	7,416	5982

Table 1. Tweets for qualitative coding of techniques

Coding Framework for Influence Techniques

To develop the coding framework, we adopted an abductive approach, combining deductive structures from existing analysis on ISIS media with inductive observations from our dataset. This was necessary given the shift in context from official propaganda materials to everyday users on Twitter. While existing schemas provided theoretical grounding, the iterative refinement with data allowed flexibility and clearer operational distinctions between techniques. This circular process of coding enabled both structure and adaptability (DeCuir-Gunby et al., 2011).

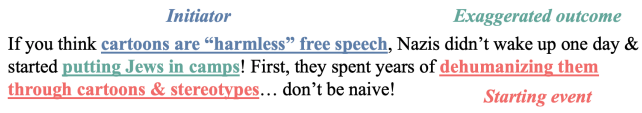

Figure 1. Example of a semantic frame for false dilemma

We drew upon Blodgett et al.'s (2023) schema for misinformation to incorporate objectivity and independence from external knowledge when designing our codebook. Annotating persuasion and propaganda presents challenges of personal bias and inter-annotator disagreement (Bonial et al., 2022). To assess reliability, 300 tweets were double-coded, achieving an averaged Matthews Correlation Coefficient (MCC) agreement of 0.64, before the remaining tweets were single-coded. To further mitigate bias, each technique was operationalised with semantic frames and linguistic markers observed in tweets. For example, *false dilemma* was defined through frames linking an initiator, a starting event, and an exaggerated outcome (Figure 1), while *social proof* included markers like universal

[4] https://github.com/UKPLab/EasyNMT
[5] Translation quality was assessed by native users using model outputs from 30 randomly selected Indonesian and Urdu tweets
[6] https://pypi.org/project/tweet-preprocessor/
[7] https://pypi.org/project/nltk/

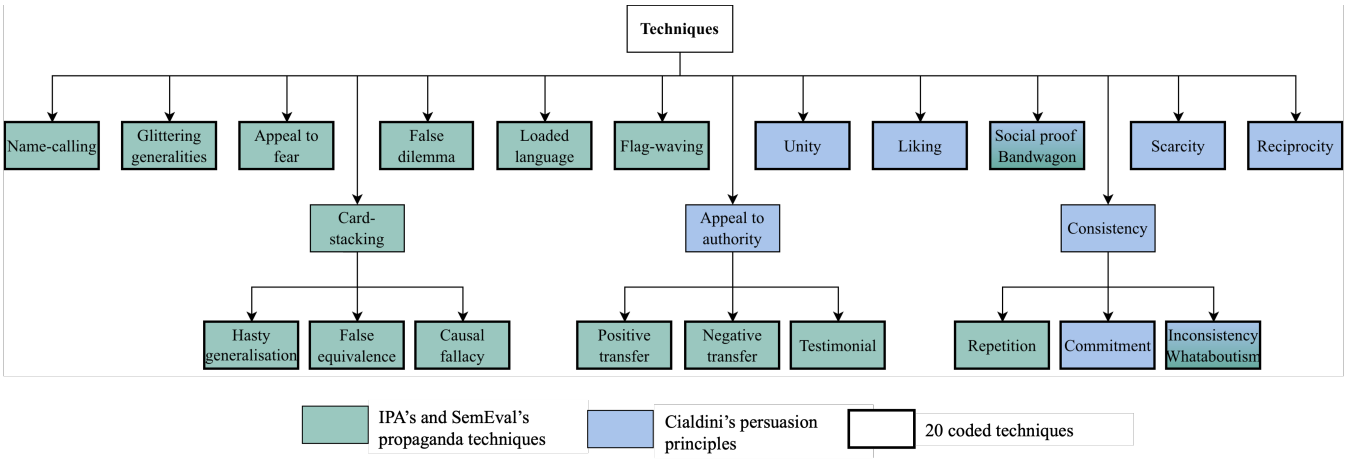

Figure 2. 20 techniques in the codebook

quantifiers("everyone") and superlatives ("most"). Jowett and O'Donnell's (2012a) devices, such as metaphors or symbolic rewards, were incorporated where relevant.

To ensure applicability across diverse international content, techniques requiring external knowledge (e.g., *red herring*, *straw man*) were excluded, and *card stacking* was adapted to identify fallacies observable without fact-checking. Drawing on Cook (2020) and Blodgett et al. (2023), we expanded the category to capture *hasty generalisation*, *false equivalence*, and *causal fallacy*. Overlaps across frameworks also necessitated refinement; these include the separation of *authority* into *positive transfer*, *negative transfer*, and *testimonial,* expansion of *commitment and consistency* to include *repetition* and *inconsistency,* and the integration of *liking* with *plain folks*, and *social proof* with *bandwagon* effects. Through this process, 20 distinct techniques were identified from 5,982 tweets (Figure 2; see Appendix A Table T2 for their operationalisations).

Refining Influence Techniques

As influence techniques are often used in combination (Jowett & O'Donnell, 2019), cluster analysis was applied to uncover naturally co-occurring patterns and consolidate the 20 coded techniques into more meaningful strategies. We implemented the k-modes algorithm[8] (Huang, 1997) for its efficiency and suitability for binary coding (Sharma & Gaud, 2015). Tweets were assigned to clusters based on the lowest dissimilarity, and cluster centres were iteratively updated until the objective function stabilised. Since k must be specified in advance, we first conducted hierarchical clustering with Jaccard distance and Ward's linkage to estimate the optimal number of clusters, followed by evaluating the k-modes objective function across values of k using Hamming distance to generate an elbow plot.

To select the optimal cluster solution, we assessed candidate partitions internally on two aspects: small within-cluster dissimilarities and stability of solution. Ideally, clusters should exhibit high intra-cluster compactness, where tweets within the same cluster are highly similar, and strong inter-cluster separation, where tweets in different clusters are distinctly dissimilar (Liu et al., 2013). Compactness and separation were measured through intra- and inter-cluster dissimilarities respectively, and overall cohesion was captured through silhouette scores. A bootstrap stability assessment with 100 samples was also conducted to ensure robustness against randomness in cluster initialisation and data variation, with stability quantified using the Adjusted Rand Index (Hennig, 2015).

Finally, candidate partitions with good internal validity were assessed on their explainability (i.e., theoretical alignment and expectations) and distinguishability (i.e., cluster distinctiveness). Chi-square tests were used to identify techniques that significantly contributed to cluster formation, while keyness analysis of n-grams with log likelihood ratios was conducted to identify distinctive keywords and hashtags (Rayson & Garside, 2000). To capture thematic context, we applied the Grievance Dictionary[9] (van der Vegt et al., 2021) with weight-based scoring to identify the dominant grievance categories in each cluster. Partitions that failed to demonstrate both internal validity and interpretability were reassessed until a final, robust cluster solution was selected. Additional robustness checks were performed using k-medoids and latent class analysis, with the results supporting the stability and validity of the final cluster solution (see Appendix B).

[8] https://pypi.org/project/kmodes/

[9] https://github.com/Isabellevdv/grievancedictionary

Technique	Operationalisation	Semantic Frame / Linguistic Markers	Example	Non-Example
Hasty generalisation	Generalise an entire group based on non-representative / single instances	Isolated events / examples AND subject of generalisation AND generalised claim; may include exaggerated /emotional language	*Samuel Paty*, the teacher, was beheaded by an Islamic terrorist in France. *Islam is full of terrorism & barbarism*…	#SamuelPaty... taught about 'freedom of expression', only to be murdered & beheaded by a barbaric Islamist terrorist.
False equivalence	Link two dissimilar events / groups as if they are equivalent	Use false analogies or metaphors to: compare two entities / events AND highlight superficial similarities and/or minimised differences	The *"Muslim question" today* is fast becoming what the *"Jewish question" was in 19th-century Europe*. Negative attitudes towards Muslim communities are *increasingly similar*	Everyone can #StandUp4HumanRights to stop racism, anti-Semitism, Islamophobia, xenophobia & all forms of intolerance...
Causal fallacy	Use of an oversimplified cause to explain an event	Misattribute cause-and-effect: assume cause AND simplify link AND observed effect; may include causal connectives (because) / simplifying conjunctions (thus)	Charlie Hebdo fueling anti-Islam sentiments in #France. Recep Tayyip Erdogan fueling religious extremism. *These two are to blame* for this *new terrorist attack* in #Nice	Recent barbaric attacks and beheadings in france on the name of religion are utterly shocking. Religion teaches tolerance, coexistence,...

Table 2. Snippet of the codebook for techniques expanded from card-stacking

Cross-cultural Differences in Susceptibility

To examine variation in susceptibility across groups, a multinomial logistic regression was conducted with technique clusters as the dependent variable and subgroup (English, Turkish, Indonesian, Urdu, Arabic) as the predictor. We included tweet volume per language as a control to account for disparities in subgroup size (e.g., 3,518 English vs. 403 Arabic tweets).

Results and Discussion

RQ1: Codebook of Influence Techniques

The abductive coding process streamlined 7 persuasion principles and 25 propaganda techniques into 20 influence techniques, informed by literature and tweet content. Our codebook is structured around technique operationalisation, markers, and illustrative cases, featuring clear headers to maintain an objective coding process and provide clarity. Table 2 provides a snippet of the codebook with the full codebook provided in Appendix A Table T3.

Loaded language was the most popular (>55%) technique, followed by *positive transfer* (≈30%) while *scarcity* and *false dilemma* were the least utilised (<6%) among the coded tweets. Overall, tweets that created false comparisons (*false equivalence*) received the highest number of retweets, although techniques generally did not differ in their usage of mentions, hashtags, and URLs.

Discussion. The prevalence of *loaded language* was consistent with prior annotations in political news (Da San Martino et al., 2020a), reflecting both its emotional salience and ease of identification. Likewise, the prevalence of *hasty generalisation* was likely reinforced by Twitter's character constraints and tendency of extremist discourse to rely on simplified assertions to provoke strong reactions (Elliott-Maksymowicz et al., 2021). In contrast, the less prevalent covert strategies such as *false dilemma* and *scarcity* echoed class imbalance issues noted in detection studies. These patterns suggest a preference for direct, emotionally charged, and easily recognisable techniques like *loaded language* and *name-calling* over subtler forms of influence. The findings underscore both the prominence of aggressive rhetorical strategies in online extremism and the challenges they pose for balanced detection across diverse techniques.

RQ2: Refining the Influence Techniques

Hierarchical clustering on a subset of tweets suggested k = 5–8, which was corroborated by k-modes objective values. Candidate partitions from k-modes clustering of this range showed low intra-cluster distance (≈0.18), moderate inter-cluster distance (≈0.31), and a silhouette score of ≈0.41, indicating compact yet moderately distinct clusters (Table). As dissimilarities measures were generally similar across different solutions, the interpretability assessment was carried out with the more stable 7- and 8-cluster solutions (ARI > 0.80), comparing their chi-square, keyness, and psycholinguistic outputs. While both were explainable, the 7-cluster solution produced clearer distinctions. The 8-cluster model offered added granularity (e.g., separating religious commitment from religious authority), but overlapping references to "Prophet," "Allah," and related grievances themes made several clusters harder to differentiate. As these refinements provided limited additional insight, the 7-cluster solution was selected to maximise interpretability while avoiding overfitting. Table 3 summarises the findings for the 7-cluster solution, along with inferred objective for each cluster.

k	Intra-Cluster Dist.	Inter-Cluster Dist.	Silhouette	ARI
5	0.188	0.318	0.409	0.765
6	0.191	**0.322**	0.414	0.739
7	**0.175**	0.316	**0.417**	0.802
8	0.178	0.315	0.415	**0.806**

Table 3. Comparison of cluster validity measures. Figures are rounded to 3 d.p., with best indicators in bold

	C1	C2	C3	C4	C5	C6	C7
Size	1122	1520	681	499	727	969	464
Cluster contribution (*Chi-sq*)	HastyGen; LoadLang; Flag-wave	PosTrans; GlitGen	Name-call; HastyGen; NegTrans Inconsistency	SocProof; Reciprocity; Commitment	Testimonial SocProof	Name-call; LoadLang; GlitGen PosTrans	HastyGen; CausalFal LoadLang; Fear;
Key n-grams (*keyness*)	Muslims Islamophobia Millions Red line Muslim ummah	Prophet Muhammad Allah Love Him Peace	Prophet Muhammad Hypocrites Fascist Enemy Colonialist	Boycott Boycott French products Thousands My appeal Muslims	I Says Erdogan Protest Campaign	Beheaded Muhammad Evil Macron Radical	Prophet Because If Target Responsible
Key hashtags (*keyness*)	#boycottfrance #armenianterror #boycot_france #macron_the_pig_of_france #group4palestine	#theprophetmuhammadpbuh #خاتم_النبيين_محمدﷺ #we_love_mohammad_صلى #respectmuhammad #prophetofhumanity	#racistmacron #racistwest #holocaust #fulani #franceenemyofislam	#boycottfranceproducts #removefrancefromunsc #team_defenders #france_spreadinghate #oicbanfrenchproducts	#boycottfranceproducts #we_love_mohammad_صلى #statement #islamophobia #هدبتنا_لماكرون#	#boycottfrance #boycott_french_products #terroristgeertwilders #charliehebdo #macron	#boycottfranceproducts #except_the_beloved_of_allah #expel_french_ambassador #isis #ottoman
Grievances (theme)	Honour, Soldier Threat	God, Honour, Soldier	Violence, Injustice, Impostor	Honour, God, Help	God, Planning, Organise	God, Hate, Threat	Threat, Paranoia, Impostor
Inference	**Leverage generalisations & religious sentiments to prompt action**	**Appeal to religious authority and sanctity**	**Appeal to historical grievances & hypocrisy**	**Leverage social pressure & duty to prompt action**	**Leverage authority to prompt action**	**Leverage contrast to praise or condemn**	**Instil fear with assertions and fallacies**

Table 4. Interpretation of the 7-cluster solution

C	Characteristic tweet	Inference
1	If the boycott succeeds, it means the end of the persecution of our brothers and sisters… Remember, as you boycott, you are protecting the lives and dignity of millions of Muslims #BoycottFrenchProducts	Leverage generalisations & religious sentiments to prompt action
2	#TheProphetMuhammadPBUH Prophet Muhammad ﷺ is the last messenger of God. He is the best of creation, a role model on how to be of exemplary character, kindness, mercy,…	Appeal to religious authority and sanctities
3	… your nation was tyrannical towards 36 nations and their millions of colonised Muslims; your history is of murder and genocide; … and you never apologised for it... [URL]	Appeal to historical grievances & hypocrisy
4	As a member of the board of directors…, I submitted an urgent request to the board to boycott and... I call on all my colleagues in cooperative societies to boycott in support of our Prophet #مقاطعه_المنتجات_الفرنسيه	Leverage social pressure & duty to prompt action
5	…"Al Meera", owner of the largest chain of stores, announces a #boycott_of_French_products... #Qatar University also announced the cancellation of the French Cultural Week event.	Leverage testimonies to prompt action
6	France, which committed genocide in Algeria, is attacking our Prophet who saved girls from being buried alive, ... and proclaimed no superiority of whites over blacks.	Leverage contrast to praise or condemn
7	The economic collapse that started in 2008 has peaked with COVID-19. This results in the rise of fascism on the US-EU axis, and ... We are just at the beginning. There will be more Macron and Wilders.	Instil fear with assertions and fallacies

Table 5. Characteristic tweets based on proximity to each cluster centroid

Among the seven strategies, two were authority-related (appealing to religious authority and sanctities, and leveraging testimonies to prompt action), while two others were fallacy-related (leveraging generalisations to prompt action, and instilling fear with assertions). Together, these strategies accounted for nearly two-thirds of all influential tweets (N=3,833). Characteristic tweets representing the core attributes of each cluster were also identified by their proximity to its centroid, illustrating these strategies and providing context for how influence techniques were orchestrated in the Twitter discourse (Table).

Discussion. The dominance of authority- and fallacy-based strategies was unsurprising given the strong cultural responsiveness to authority in Islamic cultures (Alnunu et al., 2021), where references to the Prophet and religious sanctities likely drove engagement in our context. Likewise, fallacious reasoning exploited cognitive biases to evoke fear and moral outrage, amplifying virality and misinformation (Blodgett et al., 2023). Contrast-based messaging, the third most common strategy (N=969; 18%), sharpened "us-versus-them" distinctions (Cialdini, 2016) by praising religious or political leaders while condemning opponents

like France and Macron, thereby reinforcing group identity and emotional intensity (Jowett & O'Donnell, 2015).

Across strategies, the underlying goal was seemingly mobilisation of Twitter users and political leaders. Of these strategies, three were especially explicit in their calls to action (N=2,348, 39.2%), prominently employing *flag-waving* and *social proofing* to inspire actions through appeals to religious sentiments, testimonies, and social pressure. Hashtags like #boycottfranceproducts and #expel_french_ambassador, along with n-grams referencing the "Muslim ummah" and collective action ("thousands," "campaign," "protest"), further support this observation, revealing calls for coordinated political or participatory responses across the global Islamic community.

RQ3: Cross-cultural Differences in Influence Susceptibility

The regression model converged successfully (likelihood ratio $\chi^2(24) = 462$, $p < .001$; pseudo $R^2 = .207$), suggesting that varied influence susceptibilities were found among the five Islamic subgroups. Arabic users were more susceptible to generalisations and religious sentiments (C1) and strategies that instil fear with assertions and fallacies (C7), while English and Pakistani (Urdu) users showed greater susceptibility to religious and sanctity appeals (C2) and strategies that leverage contrast (C6). By contrast, responses to historical grievances (C3) and social pressure and duty appeals (C4) were generally uniform across subgroups, although Turkish users were less susceptible to C3.

Discussion. The Arabic subgroup's heightened susceptibility to religious and fear-inducing strategies leveraging generalisations and fallacies suggests a preference for emotional and religious resonance over logical reasoning. This may be attributed to the communal religious framework ingrained in Arabic societies, which fosters receptivity to definitive narratives that reinforce religious identity and values (Güngör et al., 2014). Likewise, the susceptibility of English and Pakistani users to religious and sanctity appeals, and contrast-based strategies, highlights the role of binary framing and authoritative figures in shaping attitudes.

Consistent with Alnunu's et al. (2021), susceptibility to strategies invoking social pressure and religious duty was broadly uniform, reflecting norms of obligation and collective responsibility across Islamic cultures. Similarly, strategies invoking historical grievances resonated widely, evoking collective memory of atrocities and colonial legacies (e.g., France's colonialisation of African Muslims, based on top keywords like 'Fulani' and 'Algeria'). This shared remembrance, coupled with the concept of *ummah* which fosters a sense of vicarious suffering for the global community (Mandaville, 2001), can reinforce solidarity and amplify the moral outrage and resonance of such narratives. The exception was the Turkish subgroup, whose lower susceptibility may stem from secularist traditions (Yavuz & Öztürk, 2019) and enduring pride in the Ottoman legacy (Grigoriadis, 2009).

Conclusion

Our study offers three core contributions spanning codebook development, influence characterisation and identification, and counter-propaganda strategy. The lack of an operational coding framework has hindered both qualitative and automated analysis of influence techniques. We addressed the gap by developing a codebook tailored to social media posts, adapting insights from official propaganda materials from ISIS include semantic frames, linguistic markers, and examples. By expanding techniques and eliminating overlaps, we consolidated 20 distinct techniques that were prevalent and practical to code, improving annotation consistency and reducing misclassification risks, while offering a foundation for future detection in social media contexts.

Identifying single techniques alone may overlook the synergistic effects of multiple techniques as influence mechanisms and its broader strategic intent. Although 20 techniques were consolidated, its sheer number meant that detection will likely remain challenging, especially for less utilised techniques like *false dilemma* and *scarcity*. Clustering them into seven overarching strategies revealed how techniques like *loaded language* function within generalisations, fallacies, and contrasts to instil fear or discredit opponents. Our approach provides richer insights into influence mechanisms and may streamline future detection by focusing on strategy-level patterns, reducing the impact of class imbalance on prediction accuracy.

In addition, the cross-cultural differences in susceptibilities to influence strategies have practical implications for counter-propaganda. First, addressing universally resonant themes like social duty and historical grievances can maximise message outreach across subgroups, by stressing interfaith cooperation and highlighting reconciliation while acknowledging past suffering. Second, tailoring counter-messages to cultural contexts can improve resonance; for example, English and Pakistani audiences, more susceptible to religious and contrast-based appeals, may respond better to nuanced religious perspectives or shared-value messaging to counter 'us-versus-them' narratives. Third, the identification of populations vulnerable to fallacy-based techniques can guide counter-misinformation efforts. In this context, the Arabic subgroup's susceptibility to hasty generalisations and causal fallacies underscores the need for tailored countermeasures like fact-checking and inoculation strategies that challenge oversimplified narratives.

Despite these implications, several limitations must be acknowledged. Although the translation quality was evaluated by native users on a small subset of tweets, translation may have inevitably resulted in the loss of language-specific nuances, particularly for more implicit techniques like *false dilemma*, affecting their prevalence and annotation accuracy. Clustering also presented challenges: the 7-cluster solution's moderate silhouette score (0.42) suggests limited separation, and k-modes is sensitive to initialisation and tweet order, though checks on internal stability (ARI = 0.81) and external stability with other partitioning algorithm (ARI = 0.68 – 0.79) improved and confirmed robustness. The multinomial regression explained only modest variance (pseudo R^2 = 0.21), reflecting translation and coding challenges as well as inherent behavioural complexity. Finally, the study's focus on text-based identification overlooked multimodal elements of social media (e.g., images, videos), meaning techniques expressed through visual or interactive forms (e.g., *liking* from sarcasm conveyed through memes) were not captured, thereby reducing the comprehensiveness of our analysis.

Future studies should involve multiple native-language annotators to better evaluate translation quality and ensure cultural nuances are preserved, before exploring large language models (LLMs) for scalable annotation while balancing efficiency with human oversight. To address the sensitivity of k-modes to initialisation, clustering results should be compared with alternative algorithms, strengthening external validity through cross-method consistency. Generalisability could be further evaluated by replicating the coding and clustering process on datasets with ground-truth labels (e.g., SemEval) or in other contexts such as far-right extremism or the Israel–Hamas war, and by applying the codebook in a semi-automated or LLM-assisted annotation setting to demonstrate practical utility. Refining the framework through such applications may also justify adapting or removing weakly correlated techniques, thereby improving cluster cohesion and enhancing explanatory power in regression analyses.

References

Abdullah, M., Altiti, O., & Obiedat, R. 2022. Detecting Propaganda Techniques in English News Articles using Pre-trained Transformers. *2022 13th International Conference on Information and Communication Systems, ICICS 2022*. https://doi.org/10.1109/ICICS55353.2022.9811117

Aggarwal, N. K. 2021. In defence of Islam: How the Islamic State justifies violence. In K. Bhui & D. Bhugra (Eds.), *Terrorism, Violent Radicalisation, and Mental Health* (pp. 13–24). Oxford University Press. https://doi.org/10.1093/med/9780198845706.003.0002

BBC. 2020a. France attack: Three killed in 'Islamist terrorist' stabbings. *BBC News*. https://www.bbc.com/news/world-europe-54729957

BBC. 2020b. Anti-France protests: Muslims hold rallies worldwide as tensions rise. *BBC News*. https://www.bbc.com/news/world-54751920

Blodgett, A., Bonial, C., Hudson, T., & Voss, C. 2023. *Combined Annotations of Misinformation, Propaganda, and Fallacies Identified Robustly and Explainably (CAMPFIRE)*.

Bonial, C., Blodgett, A., Hudson, T., Lukin, S. M., Micher, J., Summers-Stay, D., Sutor, P., & Voss, C. R. 2022. The Search for Agreement on Logical Fallacy Annotation of an Infodemic. *2022 Language Resources and Evaluation Conference, LREC 2022*.

Brown, J. A. C. 1963. *Techniques of Persuasion: From Propaganda to Brainwashing*. Penguin.

Cherry, S. 2024. Modern Armed Conflicts: Disinformation Campaigns Shaping the Digital Information Landscape. *The Serials Librarian*, *85*(1–4), 19–31. https://doi.org/10.1080/0361526X.2024.2348140

Cialdini, R. B. 2001. Influence: Science and practice (4th ed.). In *New York: HarperCollins*.

Cialdini, R. B. 2006. Influence: The psychology of persuasion. *New York, NY, USA: HarperCollins Publishers.*

Cialdini, R. B. 2016. *Pre-suasion: a revolutionary way to influence and persuade*. Penguin Random House.

Cook, J. 2020. Deconstructing climate science denial. In *Research Handbook on Communicating Climate Change: Elgar Handbooks in Energy, the Environment and Climate Change*. https://doi.org/10.4337/9781789900408.00014

Da San Martino, G., Barrón-Cedeño, A., Wachsmuth, H., Petrov, R., & Nakov, P. 2020a. SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles. *14th International Workshops on Semantic Evaluation, SemEval 2020 - Co-Located 28th International Conference on Computational Linguistics, COLING 2020, Proceedings*. https://doi.org/10.18653/v1/2020.semeval-1.186

Da San Martino, G., Cresci, S., Barrón-Cedeño, A., Yu, S., Di Pietro, R., & Nakov, P. 2020b. A survey on computational propaganda detection. *IJCAI International Joint Conference on Artificial Intelligence*, *2021-January*. https://doi.org/10.24963/ijcai.2020/672

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. 2011. Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods*, *23*(2). https://doi.org/10.1177/1525822X10388468

Ellul, J. 1965. *Propaganda: The Formation of Men's Attitudes*. Vintage Books.

Farkas, J., & Bastos, M. 2018. IRA propaganda on Twitter: Stoking antagonism and tweeting local news. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3217804.3217929

Fogg, B. J. 2003. Persuasive Technology: Using Computers to Change What We Think and Do. In *Persuasive Technology: Using Computers to Change What We Think and Do*. https://doi.org/10.1016/B978-1-55860-643-2.X5000-8

Grigoriadis, I. N. 2009. Turkish National Identity. In *Trials of Europeanization* (pp. 123–153). Palgrave Macmillan US. https://doi.org/10.1057/9780230618053_6

Güngör, D., Karasawa, M., Boiger, M., Dinçer, D., & Mesquita, B. 2014. Fitting in or Sticking Together: The Prevalence and Adaptivity of Conformity, Relatedness, and Autonomy in Japan

and Turkey. *Journal of Cross-Cultural Psychology*, *45*(9). https://doi.org/10.1177/0022022114542977

Hennig, C. 2015. Clustering strategy and method selection. In *Handbook of Cluster Analysis*. https://doi.org/10.1201/b19706

Hofstede, G. 2001. Culture's Consequences: Comparing Values, Behaviors, Institutions and ... - Geert Hofstede - Google Books. *SAGE Publications*.

Hornikx, J., & Hoeken, H. 2007. Cultural differences in the persuasiveness of evidence types and evidence quality. *Communication Monographs*, *74*(4). https://doi.org/10.1080/03637750701716578

Huang, Z. 1997. A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. *Research Issues on Data Mining and Knowledge Discovery*.

Jones, D. G. 2024. Detecting Propaganda in News Articles Using Large Language Models. *Engineering: Open Access*, *2*(1), 1–12.

Jowett, G. S., & O'Donnell, V. 2012. What Is Propaganda, and How Does It Differ From Persuasion? In *Propaganda and Persuasion*.

Jowett, G. S., & O'Donnell, V. 2015. Propaganda and Psychological Warfare. In *Propaganda and Persuasion* (6th ed., pp. 231–312). SAGE.

Kaptein, M., De Ruyter, B., Markopoulos, P., & Aarts, E. 2012. Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. In *ACM Transactions on Interactive Intelligent Systems* (Vol. 2, Issue 2). https://doi.org/10.1145/2209310.2209313

Kellermann, K., & Cole, T. 1994. Classifying Compliance Gaining Messages: Taxonomic Disorder and Strategic Confusion. *Communication Theory*, *4*(1). https://doi.org/10.1111/j.1468-2885.1994.tb00081.x

Lakomy, M. 2022. Between the "Camp of Falsehood" and the "Camp of Truth": Exploitation of Propaganda Devices in the "Dabiq" Online Magazine. *Studies in Conflict and Terrorism*, *45*(10), 881–906. https://doi.org/10.1080/1057610X.2020.1711601

Leung, A. K. Y., & Cohen, D. 2011. Within- and Between-Culture Variation: Individual Differences and the Cultural Logics of Honor, Face, and Dignity Cultures. *Journal of Personality and Social Psychology*, *100*(3). https://doi.org/10.1037/a0022151

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., & Wu, S. 2013. Understanding and enhancement of internal clustering validation measures. *IEEE Transactions on Cybernetics*, *43*(3). https://doi.org/10.1109/TSMCB.2012.2220543

Mandaville, P. 2001. Reimagining Islam in diaspora: The Politics of Mediated Community. *Gazette*, *63*(2–3). https://doi.org/10.1177/0016549201063002005

Neubauer, L., Straw, I., Mariconti, E. et al. 2023. A Systematic Literature Review of the Use of Computational Text Analysis Methods in Intimate Partner Violence Research. *J Fam Viol* 38, 1205–1224. https://doi.org/10.1007/s10896-023-00517-7

Nikolaidis, N., Stefanovitch, N., & Piskorski, J. 2023. *On Experiments of Detecting Persuasion Techniques in Polish and Russian Online News: Preliminary Study*. 155–164.

Oddo, J. 2019. *The Discourse of Propaganda*. Penn State University Press. https://doi.org/10.1515/9780271082752

Orji, R. 2016. Persuasion and culture: Individualism-collectivism and susceptibility to influence strategies. *CEUR Workshop Proceedings*, *1582*.

O'Shaughnessy, N. J. 2004. Explaining Propaganda. In *Politics and propaganda: Weapons of mass seduction* (pp. 37–62). Manchester University Press.

Qutteineh, M. I. E. S. 2017. *The Language of Persuasion and Power in ISIS Political Discourse*. Hebron University.

Rayson, P., & Garside, R. 2000. *Comparing corpora using frequency profiling*. https://doi.org/10.3115/1117729.1117730

Rodrigues, L., Blondé, J., & Girandola, F. 2018. Social Influence and Intercultural Differences. In F. Colette (Ed.), *Advances in Culturally-Aware Intelligent Systems and in Cross-Cultural Psychological Studies* (pp. 391–413). Springer. https://doi.org/10.1007/978-3-319-67024-9_18

Saehudin, A., & Ridha, H. 2022. Coursebook-Based ISIS' Propaganda: A Critical Discourse Analysis of Arabic History Texts in ISIS' School Environments. *Insaniyat : Journal of Islam and Humanities*, *6*(2), 89–104. https://doi.org/10.15408/insaniyat.v6i2.25036

Sample, C., McAlaney, J., Bakdash, J., & Thackray, H. 2018. A cultural exploration of the social media manipulators. *European Conference on Information Warfare and Security, ECCWS*, *2018-June*.

Shabo, M. 2008. *Techniques of propaganda and persuasion*. Prestwick House Inc.

Sharma, N., & Gaud, N. 2015. K-modes Clustering Algorithm for Categorical Data. *International Journal of Computer Applications*, *127*(17), 1–6. https://doi.org/10.5120/ijca2015906708

Smith, E. R., Mackie, D. M., & Claypool, H. M. 2014. Social Psychology. In *Social Psychology*. https://doi.org/10.4324/9780203833698

Sofia, G., Marianna, S., George, L., & Panos, K. 2016. Investigating the role of personality traits and influence strategies on the persuasive effect of personalized recommendations. *CEUR Workshop Proceedings*, *1680*.

Soules, M. 2015. Introduction: The Spectrum of Persuasion. In *Media, Persuasion and Propaganda* (pp. 1–18). Edinburgh University Press. https://doi.org/10.1515/9780748644179

Sprenkamp, K., Jones, D. G., & Zavolokina, L. 2023. *Large Language Models for Propaganda Detection*.

Stiff, J. B., & Mongeau, P. A. 2016. Persuasive Communication, Third Edition. In *Guilford Publications*.

Szanto, G. H. 1977. *Theater & Propaganda*. University of Texas Press. https://doi.org/10.7560/780200

Szwoch, J., Staszkow, M., Rzepka, R., & Araki, K. 2024. Limitations of Large Language Models in Propaganda Detection Task. *Applied Sciences*, *14*(10), 4330. https://doi.org/10.3390/app14104330

Teimas, R., & Saias, J. 2023. Detecting Persuasion Attempts on Social Networks: Unearthing the Potential of Loss Functions and Text Pre-Processing in Imbalanced Data Settings. *Electronics (Switzerland)*, *12*(21). https://doi.org/10.3390/electronics12214447

Tiedemann, J., & Thottingal, S. 2020. OPUS-MT - Building open translation services for the World. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, EAMT 2020*.

van der Vegt, I., Mozes, M., Kleinberg, B., & Gill, P. 2021. The Grievance Dictionary: Understanding threatening language use. *Behavior Research Methods*, *53*(5), 2105–2119. https://doi.org/10.3758/S13428-021-01536-2

Vaughn, S., Jowett, G. S., & O'Donnell, V. 1989. Propaganda and Persuasion. *The American Historical Review*, *94*(1), 102. https://doi.org/10.2307/1862086

Yavuz, M. H., & Öztürk, A. E. 2019. Turkish secularism and Islam under the reign of Erdoğan. In *Journal of Southeast European and Black Sea* (Vol. 19, Issue 1). https://doi.org/10.1080/14683857.2019.1580828

Zgryziewicz, M. R., Grzyb, T., Fahmy, S., & Shaheen, J. 2016. *Daesh information campaign and its influence.*

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, this study aims to improve influence characterisation in extremist context and highlight the importance of examining propaganda susceptibilities within and across cultural subgroups.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, the abstract and introduction highlight the issues arising from conceptual and practical overlaps in influence techniques, which this study addressed through its key contributions.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, see the Methodology section. We justified the appropriateness our methods by referencing and refining existing frameworks and ensuring robustness in our analyses.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we controlled for unequal tweet counts in the subgroups, and identified nuances lost in translation as a limitation in our Conclusion section.
   (e) Did you describe the limitations of your work? Yes, see the Conclusion section.
   (f) Did you discuss any potential negative societal impacts of your work? No, we did not explicitly discuss the potential negative societal impacts. We focused on characterising harmful influence mechanisms that can manipulate or deceive audiences, and identifying populations that might be susceptible to them.
   (g) Did you discuss any potential misuse of your work? No, we did not explicitly discuss any potential misuse of our work (e.g., designing culturally resonant propaganda messages)
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, only aggregated results were reported, and only relevant segments of example tweets were shown to prevent identification of individual users or the original tweets. Access to the dataset was limited to research purposes and while not shareable under ethical guidelines, methodological choices were carefully documented to ensure transparency and support reproducibility.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA, our study is primarily empirical and exploratory rather than theoretical
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theortical results? NA, our study is primarily empirical and exploratory rather than theoretical
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA, the study did not involve machine learning experiments requiring reproducibility packages. Analyses were conducted with standard statistical methods, and links to packages were provided where applicable.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...

(a) If your work uses existing assets, did you cite the creators? Yes, we cited the authors of the EasyNMT translation package (Tiedemann & Thottingal, 2020) and the Grievance Dictionary (van der Vegt et al., 2021) and provided links to their GitHub repository.

(b) Did you mention the license of the assets? No, we did not mention the license in the main text. The EasyNMT package is under the apache-2.0 license, while the Grievance Dictionary is under the MIT license.

(c) Did you include any new assets in the supplemental material or as a URL? NA, as no new assets are included in the Appendix.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA, as this study did not use data from third parties.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? No, we did not explicitly discuss it in the main text. All personally identifiable information like user IDs and tweet IDs was anonymised, and we avoided showcasing examples of full tweets to prevent identification. Tweets may contain offensive language given the extremist context, but such content was analysed for their semantics and framing under ethical guidelines.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA, we did not curate or release new datasets

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...

(a) Did you include the full text of instruction given to participants and screenshots? NA, our study did not use crowdsourced or human research data

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA

(d) Did you discuss how data is stored, shared, and deidentified? NA

Appendix

A. Definitions, Operationalisations, and Codebook of the Influence Techniques

The complete lists of the seven Cialdini, seven IPA, and 18 SemEval techniques, along with their definitions and overlaps can be found in Table A, while the operationalisations of the 20 influence techniques identified from 5,982 tweets are provided in Table B. Finally, Table C details the full codebook of the 20 techniques.

	Technique	Definition	Overlap(s)		
			Cialdini	IPA	SemEval
Cialdini	*Commitment & consistency*	Individuals tend to align with previous behaviours to avoid dissonance			✓ (*repetition*)
	Social proof	Individuals tend to trust and adopt the values and beliefs of their peers		✓ (*band-wagon*)	✓ (*band-wagon*)
	Authority	Individuals tend to be influenced by authoritative figures perceived as trustworthy		✓ (*pos trans*) (testimonial)	✓ (*appeal to authority*)
	Reciprocity	Individuals tend to feel the obligation to repay in kind			
	Scarcity	Individuals tend to place higher value on limited resources			
	Liking	Individuals tend to be persuaded by those they like		✓ (*plain folks*)	
	Unity	Individuals tend to be persuaded by others that share a common identity			
IPA	*Name-calling*	Label opponents to create a 'negative-other' representation			✓ (*name-call*) (*reductio ad hitlerum*)
	Glittering generalities	Associate agenda with vague but colourful virtues to gain acceptance without critical scrutiny			✓ (*exaggerate*)
	Transfer	Use the prestige of a revered entity to legitimise another cause, or invoke negative associations with a disliked entity to provoke rejection	✓ (*authority*)		✓ (*appeal to authority*)
	Testimonial	Leverage the testimonies of respected individuals to shape perceptions toward a particular agenda	✓ (*authority*)		✓ (*appeal to authority*)
	Plain folks	Appear more relatable to the audience by presenting as an ordinary man	✓ (*liking*)		
	Card stacking	Selectively presents information to create the most favourable or unfavourable view			
	Bandwagon	Manipulates the 'herd instinct' where individuals follow the actions of the majority	✓ (*soc proof*)		✓ (*band-wagon*)
SemEval	*Red herring*	Introduce irrelevant information as distraction			
	Straw man	Misrepresent an opponent's argument as a weaker version to refute it			
	Whataboutism	Charge an opponent with hypocrisy			
	Causal over-simplification	Attribute a single or simple cause to a complex issue			
	Appeal to authority	Claim something to be true based on the words of an authority or expert	✓ (*authority*)	✓ (*pos trans*) (testimonial)	
	Black-and-white fallacy	Present two alternative options as the only possibilities			
	Name-calling	Label an opponent as something the audience fears, hates, or finds undesirable		✓ (*name-call*) (*neg trans*)	
	Loaded lang	Use emotionally charged negative language to influence an audience			
	Exaggeration / minimisation	Either represent something in an excessive or undermining manner		✓ (*glitter gen*)	
	Repetition	Repeat the same message to increase audience acceptance	✓ (*consistent*)		
	Appeal to fear	Support an idea by instilling fear against other alternatives			
	Slogans	Use brief and striking phrases that may include labelling and stereotyping			
	Flag-waving	Appeal to national and/or religious pride and sentiments			
	Bandwagon	Influence others to join in because "everyone is doing it"	✓ (*soc proof*)	✓ (*band-wagon*)	
	Reductio ad hitlerum	Compare an opponent's argument to that of individuals hated in contempt by audience		✓ (*name-call*) (*neg trans*)	
	Thought-terminating cliches	Use words that hinder critical thinking and meaningful discussions			
	Doubt	Question the credibility of something or someone			
	Obfuscation	Intentional use of unclear and obscure words to confuse an audience			

Table A. Definitions of the Influence techniques and their overlaps

Technique	Operationalisation (Deductive and Inductive[+])
Name-calling	Use of labels that carry negative religious connotations (e.g., *kafir*); mock judgmental capabilities; contain animal imagery; negative and simplistic stereotypes[+] (e.g., *Hindutva*)
Glittering generalities	Use of moral / virtue words or verbal symbols of power[^] to describe a cause, ideology, or individual's actions (e.g., truth, lions)
Fear	Use of existing prejudices[^] or exaggerated consequences[^] to instil anxiety
False dilemma	Use of slippery slope argument or 'lesser of two evils' metaphors[^]
Hasty generalisation[+]	Use of a single or few instances and/or events to generalise the entire cause, ideology, or groups
False equivalence[+]	Use of false analogies or metaphors[^] to link two dissimilar events or groups as if they are equivalent
Causal fallacy[+]	Use of an oversimplified cause[^] to explain an event
Loaded language	Use of high negative valent words and/or dysphemism[^] (e.g., 'propaganda' vs. 'persuasion') to describe an event or group
Flag-waving	Use of nationalistic or religious symbols, language, or values[^] (e.g., democracy, martyrdom)
Positive Transfer	Use of quotes and references of religious text and/or leaders
Negative Transfer	Use of quotes and references of widely disliked leaders or authorities
Testimonial	Use of statements or reports from authority or well-known figures
Unity	Use of communal language that fosters belonging; highlight a common enemy; express solidarity for a movement or cause[+]
Liking	Use of compliments[+]; familial or binding words; plain folk language[+] to relate to audience; humour to mock or ridicule opponent
Social proof	Use opinions of peers (including peer pressure[+]) and opinion leaders; highlight action as popular (*bandwagon*)
Repetition	Use of repeated stances or opinions from a previous sentence or part of a sentence for emphasis
Commitment	Use of public (online) commitment[+]; past actions; sense of duty or responsibility to uphold (assumed) beliefs and/or values[+]
Inconsistency	Use of accusations or *whataboutism* to highlight contradictory statements or actions of opponents
Scarcity	Use of time sensitivity[^] and consequences of inactions to generate implied m[+]
Reciprocity	Use of reciprocal words to frame the return of Allah's and/or the Prophet's favour; retaliation against opponents who undermined the religion or religious figures; symbolic punishment[^] or retribution for opponents; symbolic reward[^] for specific actions

Table B. Operationalisations of the 20 techniques. [+]Denotes inductive elements from tweet narratives that were adapted into the operationalisation of techniques; [^]denotes Jowett & O'Donnell's (2012) devices for maximising technique effects

Technique	Operationalisation	Semantic Frame / Linguistic Markers	Example	Non-Example
Name-calling	Label opponents / enemies with derogatory terms to degrade or dismiss them	Use labels that carry negative religious connotations (e.g., *kafir*); mock judgmental capabilities; contain animal imagery (e.g., pig); negative and simplistic stereotypes (e.g., *Hindutva*)	This Ustad is like MUI, just saying *Kafir*…Aqua and Vit mineral water are forbidden because they are products of *infidels*…	#boycottfrance… refrain from buying and selling these products of France…
Glittering generalities	Portray the moral superiority of the cause or aggrandise the actions of leaders	Use of virtue words (e.g., loyalty, glory) or verbal symbols of power (e.g., lion) to describe a cause, ideology, or individual's actions	It was *exalted to the heavens* as the *bastion of civilization*! He made chilling insults to the holy prophet of a religion, our Prophet (pbuh), who was sent as *a mercy to the worlds…*	Some say "My Prophet forgave those who insulted him". The Prophet as a person indeed never got angry when his person was insulted.
Appeal to fear	Use of prejudices / exaggerated consequences to instil anxiety	Prejudices / hyperbolic statements AND certainty and inevitability in consequences	… those who disbelieve are allies of one another. *If you do not* do so, there will be *fitnah (disorder)* in the land and *great corruption*…	He who kills a soul unless it be (in legal punishment) for murder or for causing disorder and corruption on the earth will be as if he had killed all humankind
False Dilemma	Present two alternative options as the only possibilities	Use of slippery slope argument (initiator AND starting event AND the exaggerated outcome) OR 'lesser of two evils' metaphors; may include use of either-or correlative conjunctions, or type-one-conditional tense	If you think *cartoons are "harmless" free speech*, Nazis didn't wake up one day & started *putting Jews in camps*! First, they spent years of *dehumanizing them through cartoons & stereotypes* ...	Charlie Hebdo is doing the same to Muslims as Nazi Germany's Der Stürmer did to Jews with its antisemitism.
Hasty generalisation	Generalise an entire group based on non-representative / single instances	Isolated events / examples AND subject of generalisation AND generalised claim; may include exaggerated /emotional language	*Samuel Paty*, the teacher, was beheaded by an Islamic terrorist in France. *Islam is full of terrorism & barbarism*…	#SamuelPaty...taught about 'freedom of expression', only to be murdered & beheaded by a barbaric Islamist terrorist.

False equivalence	Link two dissimilar events / groups as if they are equivalent	Use false analogies or metaphors to: compare two entities / events AND highlight superficial similarities and/or minimised differences	The *"Muslim question" today* is fast becoming what the *"Jewish question" was in 19th-century Europe*. Negative attitudes towards Muslim communities are *increasingly similar* to…	Everyone can #StandUp4HumanRights to stop racism, anti-Semitism, Islamophobia, xenophobia & all forms of intolerance...
Causal fallacy	Use of an oversimplified cause to explain an event	Misattribute cause-and-effect: assume cause AND simplify link AND observed effect; may include causal connectives (because) / simplifying conjunctions (thus)	Charlie Hebdo fueling anti-Islam sentiments in #France. Recep Tayyip Erdogan fueling religious extremism. *These two are to blame* for this *new terrorist attack*…	Recent barbaric attacks and beheadings in france on the name of religion are utterly shocking. Religion teaches tolerance, co-existence,...
Loaded language	Use of strong negative emotional indications to describe an event / group	Use of high negative valent words and/or dysphemism (e.g., 'propaganda' vs. 'persuasion'), hyperbole statements	Whoever wants to support his Messenger (peace be upon him), let him boycott French products because the *collapse of the economy* will be more painful for the *slaves of money*…	Alongside the boycott, we must introduce the French people to our religion and our Prophet. Hence, I announce the "This is Our Prophet" campaign
Flag-waving	Play on strong nationalistic or religious pride and sentiment to justify / promote a cause or belief	Use of nationalistic or religious symbols, language, or values (e.g., democracy, martyrdom)	*Sultans and victorious armies* took pride in belonging to Him. At the end of every struggle, his enemies were humiliated, and *those loyal to him*, like the Ottomans, were honored. *His Ummah* will also show Macron who he is insulting.	The way Sultan Abdülhamid II stopped the plan for a show that insulted Prophet Muhammad SAW in France when he led the Ottoman Turkey….
Positive Transfer	Transfer prestige associated with a revered entity to another entity	Use of quotes and references of religious text (e.g., Hadiths, Quran) and/or leaders (e.g., Prophet, Imam)	*"*The greatest jihad is to battle your own soul, to fight the evil within yourself.*"* - *Prophet Muhammad* (peace be upon him)	…We will not retreat from the boycott and will not stop defending our Prophet and supporting him, peace be upon him!
Negative Transfer	Transfer negative emotions associated with an entity to provoke rejection of another	Quoting and referencing widely disliked leaders or authorities	*Hitler*, I mean Erdogan, is now challenging the US to sanction it…	Imagine if Macron posted a picture of Hitler killing Jews on a government building? …
Testimonial	Leverages influence and/or endorsement of respected persons to shape public perception / behaviours toward a belief	A) Statements or reports from authority figures or well-known news media	*Erdogan said* "What is Macron's problem with Islam? What is his problem with Muslims? He needs mental checks….	The ruler in France has gone astray, he spends the whole day talking about Erdogan. He is a patient and truly needs mental treatment…
		B) Personal experiences / viewpoints and opinions written in first person; can be strengthened by supplementing it with personal identity, knowledge evidence, or expert's endorsement	@Yousuf_Alkamali *I live in France*, *and the boycott has greatly affected them*, and this is the only language that the capitalists understand. They...have no loyalty to a religion, a country,..	Boycotting French products will cause France a loss of 100 billion dollars. France exports 100 billion dollars worth of products to Muslim countries every year…
Unity	Fosters a sense of belonging or identity in the cause or movement	Use of communal language (e.g., we, ours) and/or identity terms, OR highlight a common enemy; may also express solidarity for a movement or cause	*We r Muslims*. We r fr different countries but *we r one soul one voice, one body & Muslim Umaah*…I salute all Muslims for showing *unity*. *Let's #boycottfrenchproducts* …	#Boycott_French_Products I hope every Muslim who cares about his religion and his beloved Prophet Muhammad ﷺ boycotts these products and fights them economically…
Liking	Reduce psychological distance by presenting in a likeable manner	Use of compliments; familial or binding words (e.g., brothers); plain folk language to relate to audience; humour or sarcasm to mock or ridicule opponent; may also include inclusive personal deictics (e.g., we, us)	In *our dear Kuwait*, its *noble people* have rallied in a popular campaign that the associations and shops have interacted with to #Boycott_French_Products…	Kuwait has become the first among Muslim countries to rise in defense of the prestige, the honor, the sanctity…
Social Proof	Use opinions or actions of others	A) Opinion and/or action of peers and opinion leaders AND encourage action	#we_love_prophet_muhammad_ﷺ❤️What do you lose by retweeting this, *let's see who skip without retweeting…*	…There is no doubt that He is the ideal of every activity for us who was sent by Allah…RETWEET for the sake of Allah

	to justify or promote a cause or action	B) Highlight action as popular by referencing to the mass of support (e.g., majority) AND encourage action; may include quantifiers (e.g., all, most), indefinite pronoun (e,g., everywhere) numbers and statistics as evidence	#Qatar's flagship Al Meera supermarket has removed all French products from its shelves after *calls for boycott grew louder across the Arab and Muslim world*... #BoycottFrance	Muadhin of Masjid Al Haram Sheikh Sami Rayes calls for the Boycott of French Products and chains after French President supported and encouraged the blasphemy…
Repetition	Repeat stances / opinions from a previous sentence or part of a sentence for emphasis	Repeated phrases AND (i) parallelism (similar sentence / phrase structures); OR (ii) logical flow; OR (iii) rhyme and rhythm for cohesion	Why do you *defend the honor of Prophet Muhammad*? *I defend him because* he was an extraordinary man... *I also defend him* as an extension of the love I have for all Muslims…	#Our_Prophet_is_a_Red_Line There is no good in us if we do not defend the best of prophets and messengers.
Commitment	A) Public commitment to a cause (online)	(i) Personal declaration (e.g., I am…) OR (ii) identity declaration (e.g., as a member of …) AND a cause	*I am [Name]* and *I'm a proud Muslim*. I *protest against* the disrespect of our beloved Prophet Hazrat Muhammad (S.A.W.W) and *fully support the campaign* #boycottfrance Retweet This and…	I hope a campaign called #Replace_French_Product_with_Turkish starts in the rest of the Arab and Islamic countries…
	B) Obligation to uphold beliefs and/or values	A sense of duty / responsibility AND an action; may include modals of obligation (e.g., should, must, have to)	It's *the duty of every muslim* to *take steps* to stop that blasphemous acts in france! *All muslim countries* if have any respect left *should* immediately *cut off all the relations*...	People of Pakistan are demanding expulsion of French ambassador @ImranKhanPTI Retweet!
Inconsistency	Accusations of hypocrisy	Juxtapose contradictory ideas next to each other; may include irony / sarcasm / whataboutism	Mocking black people = racism Mocking Jewish people = anti-semitism And for some reason...Mocking Muslims = free speech…	Where is the outrage over the murder of Samuel Paty and the attack on free speech in France...?
Scarcity	Create a sense of immediacy and importance to motivate action	Use of urgency language and/or temporal frame OR limited time opportunities AND calls to action	What are we waiting for? Why haven't we expelled the French ambassador from the country yet?	In order to defend his so-called secular values,…, which deeply offended Muslims around the world. Pakistan should expel French ambassador…
Reciprocity	Return favour or reprisal in kind	A) Received a favour from religion / religious authorities AND requirement to return in kind; may include reciprocal pronouns (e.g., each other) or verb (e.g., give, pay) + 'back'	The one person who *prayed for us* when he didn't even know us, *cried hours on end for us* and *the least we can do is be angry* about the fact that he has been mocked	We love the Messenger of Allah, peace be upon him, and we believe that our love for him is part of our faith…
		B) Anticipate a reward / positive outcome FROM performing an action or belief	We did not boycott and defend our Prophet because he needs us, peace be upon him, but because we need his intercession and to drink from his basin…	
		C) Action by adversary that undermined the religion / religious authorities AND retaliation against action	French Interior Minister: "We have closed more mosques than any previous administrations." And we will close all your stores in Islamic countries…	We are angry with President Macron who is arrogant and … But we are also angry with the terror that kills civilians. Anger cannot be expressed with terror and violence…

Table C. Codebook of persuasion and propaganda techniques

B. Robustness Checks of Cluster Solution

To assess the robustness of the clustering solution obtained via k-modes ($k = 7$), alternative distance-based (k-medoids) and model-based (LCA) clustering were conducted.

K-medoids. As a first robustness check, k-medoids[10] (Hamming distance, $k = 7$) was implemented with 100 random initialisations to reduce sensitivity to initialisation. Like k-modes, k-medoids supports non-Euclidean distance metrics but differs in how cluster representatives are defined, providing a meaningful methodological comparison. The resulting solution yielded a mean silhouette width of 0.38, indicating a moderate but relatively meaningful cluster structure, consistent with the primary k-modes solution (silhouette = 0.42). Visual inspection of the representative tweets in each cluster indicated similar dominant combinations of techniques, although minor differences were observed in the assignment of tweets at cluster boundaries. Quantitatively, the agreement between k-medoids and k-modes was high (ARI=0.79), indicating stability and replicability in the 7-cluster structure.

Latent Class Analysis (LCA). A second robustness check was implemented using LCA[11], which assigns class membership based on posterior probabilities rather than distance, making it methodologically distinct from k-modes and k-medoids. Models were estimated for $k = 2 – 8$ classes, with fit indices favouring a 6-class (BIC = 63,619.1, entropy = 0.56) over a 7-class solution (BIC = 62,942.8, entropy = 0.52). The LMR test was significant for both 6- and 7-class solutions, but non-significant for the 8-class solution ($p = .43$), suggesting that the addition of an eighth class did not improve model fit. Visual inspection of the latent classes indicated that the primary difference between the solutions was the subsumption of religious duty with authority to prompt action in the 6-class solution, with the 7-class solution providing a more granular separation of these techniques. Tweets with more ambiguous technique profiles (i.e., *hasty generalisations*, *false dilemma*) also exhibited lower posterior classification certainty and were more likely to be assigned differently compared to the k-modes solution. Nonetheless, the level of agreement between LCA and k-modes for the 7-class solution was moderate to high (ARI=0.68), indicating that the overall cluster structure is relatively robust and stable.

[10] https://CRAN.R-project.org/package=kmed

[11] https://github.com/dlinzer/poLCA