Changing Societies & Personalities, 2025
Vol. 9, No. 3, pp. 594–613
https://doi.org/10.15826/csp.2025.9.3.344

ARTICLE

Order, Loyalty, and Polarization in a Public Debate: Evidence From a Russian-Language Online Experiment

Dimitri Volchenkov
Texas Tech University, Lubbock, TX, USA

Daniel Korley
Texas Tech University, Lubbock, TX, USA

Alexander N. Lebedev
Institute of Psychology of the Russian Academy of Sciences, Moscow, Russia

ABSTRACT

We examine how presentation order structures polarization in a public debate. A large online debate on a proposed "Ministry of Happiness" is analyzed. The data for analysis was obtained from a two-stage experiment in which 307 (Stage 1) and then 200 (Stage 2) respondents from different age groups evaluated 50 basic and 10 additional statements about the creation of a "Ministry of Happiness." The contribution is a sequence-aware, audience-referenced reading of the same script. We trace how alignment with loyal or disloyal framing accumulates over the observed order, recover age-ordered cohorts directly from response patterns, rank the most divisive statements, and quantify sign-free responsiveness by age and gender. In terms of age categories, youth initially respond to statements about benefits and feasibility, though they quickly adjust after scrutiny or signs of distrust, ultimately receiving lower ratings when critical moments arise. Mid-life splits into a trust-forward profile and an oversight-oriented profile. The oldest cohort holds early, then declines in long negative runs. Polarization concentrates on a compact set of identity, control, and feasibility statements, with a sharp pivot at a mid-sequence trust cue. In the older module, women react more decisively than men at the same ages. The approach is descriptive and transparent, identifies where and for whom polarization accumulates, and is portable to other debates.

Received 19 February 2025
Accepted 26 September 2025
Published online 10 October 2025

dimitri.volchenkov@ttu.edu, dkorley@ttu.edu, lebedevan@ipran.ru

KEYWORDS
order effects, framing, polarization, loyalty index, net loyalty drift, age cohorts, gender differences

ACKNOWLEDGEMENTS
This work was supported by the Russian Science Foundation (RSF) under Grant № 23-18-00422, https://rscf.ru/project/23-18-00422/
We used the ChatGPT-5 Thinking system to assist with drafting and language polishing, to structure sections, and to suggest robustness checks and visualization options. All data handling, coding decisions, analyses, and interpretations were designed, executed, and verified by the human authors, who take full responsibility for the content.

Introduction

How does the sequence of statements in a public debate polarize an audience? In our previous work (Volchenkov & Lebedev, 2025), we reported a two-stage experiment that tracked polarization as a fixed debate script unfolded in a large online audience. In Stage 1, 307 volunteers from six Russian regions watched a moderated debate about a proposed "Ministry of Happiness" advanced by Senator Valentina Matvienko in 2023. From the transcript, three independent experts selected 50 statements (listed in the Appendix, in Russian) and coded each as loyal or disloyal. Each statement appeared on screen for 15 s, after which participants chose "agree", "cannot evaluate," or "disagree." In Stage 2, an older subsample of 200 respondents aged 40+ evaluated ten additional statements prepared after a preliminary pass to probe age-by-gender differences. Sessions were asynchronous with one-time token authorization, and we recorded response times. Before both stages, participants completed brief instructions and a training example. A short survey collected sex, age, education, and political self-identification. All sessions were individual to avoid social influence, lists were shuffled per person, participation was voluntary with electronic informed consent, and data were stored in de-identified form.

Building on that design, the present article treats content as fixed and reads reception over order. We extend the Loyalty Index introduced previously with sequence-sensitive summaries that trace how alignment accumulates across the script (Net Loyalty Drift) and we infer audience structure directly from response vectors using *t*-SNE and *k*-means, then carry those cohorts through all analyses. For the older module, we quantify sign-free engagement by age and gender with an absolute-sum responsiveness measure, and we estimate cohort-specific logistic agreement models with clustered errors to read axis weights descriptively. The discussion situates the observed fractures and pivots within social identity, motivated reasoning, and framing literatures (Kahneman & Tversky, 1979; Kunda, 1990; Lakoff, 2004; Taber & Lodge, 2006; Tajfel & Turner, 1979; Thibodeau & Boroditsky, 2011). Labels attach to statements, not to people.

This article employed an AI writing assistant (ChatGPT-5 Thinking) for drafting and language polishing and to propose analysis summaries. All procedures, coding, and statistical analyses were designed and carried out by the human authors, who verified every output and bear full responsibility for the manuscript.

Statement Loyalty Index

We use the Loyalty Index *L*, first introduced in Volchenkov & Lebedev (2025). For each statement *s*, we assign four signed integer scores in {–3, –2, –1, 0, +1, +2, +3}: *O*(*s*) for Optimism vs. Skepticism, *T*(*s*) for Trust vs. Distrust, *P*(*s*) for Patriotism vs. Critical Perception, and *B*(*s*) for Benefits vs. Problems. The sign encodes direction, and the absolute value encodes strength. Anchors and decision rules were fixed in advance. For *O*(*s*), explicit predictions of improvement or success score toward +3, explicit predictions of decline or failure toward –3; hedges, conditionals, and mixed forecasts reduce magnitude; absence of forecasting yields 0. For *B*(*s*), attention to positive consequences, advantages, or relief of burdens raises the score toward +3; attention to harms, costs, or risks lowers it toward –3; mere description without evaluative focus yields 0; magnitude follows salience and specificity rather than length. For *T*(*s*), assertions of institutional competence, fairness, or benevolent intent increase the score; allegations of incompetence, capture, or bad faith decrease it; conditional trust (e.g., "if audited, then …") is coded weakly positive unless the condition itself signals distrust. For *P*(*s*), civic identity cues that appeal to shared national belonging or collective duty move the score toward +3; distancing, cynicism about the polity, or derogation of collective identity move it toward –3; issue-general or universalist arguments that do not invoke identity are near 0. The index is the additive summary

$$L(s) = O(s) + T(s) + P(s) + B(s),\ L(\mathrm{s}) \in [-12, +12]; \tag{1}$$

positive values denote loyal content and negative values disloyal content. Statements with small absolute *O*, *T*, *P*, and *B* contribute little to *L*. The four-score representation preserves distinctions among forecast tone (*O*), attention frame (*B*), institutional stance (*T*), and civic identity (*P*). Two-axis schemes collapse these differences, while larger bases fragment signal and reduce coder agreement. The structure aligns with agenda-building traditions in political communication: Cobb & Elder's (1972) affective stream maps to *P*, their cognitive stream maps to *T*, and their analytic stream maps to *B*. Relative to Cobb and Elder's affective, cognitive, and analytic triad, our scheme adds an explicit temporal forecast axis *O*, distinct from attention to benefits (*B*) and from institutional trust (*T*). This axis makes explicit the temporal cue implicit in that triad and is the novel element of our four-score operationalization introduced in Volchenkov & Lebedev (2025). The scheme is scoped to judgments of a public institutional proposal. Content that is rhetorical, deontic, ironic, or off topic is treated as residual. Statements without evaluative or institutional content sit near zero on all four scores and add no signal. Multi-topic lines are split into meaning units before coding (Table 1).

Table 1

Axis Scores and Loyalty Index for Two Anchor Statements

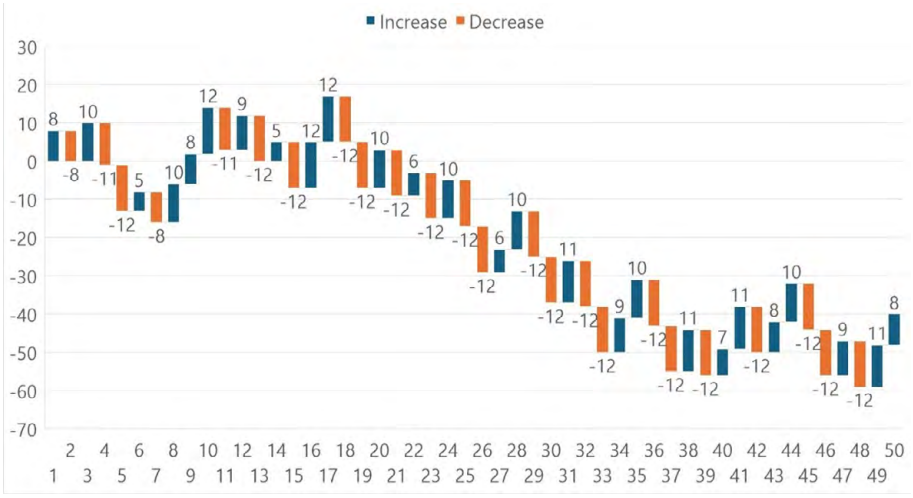
Axis (–3 … +3)	Statement 3	Statement 5
Optimism vs. Skepticism	+3	–3
Trust vs. Distrust	+2	–3
Patriotism vs. Critical Perception	+2	–3
Benefits vs. Problems	+3	–3
Loyalty Index (sum of four axes)	+10	–12

Loyalty Cascade for the Debate

Using the additive score *L* defined above, we accumulate it in presentation order to obtain a one-dimensional loyalty cascade. Additivity makes each step equal to the current statement's contribution to the running total; using an average would only rescale the same path. Figure 1 shows the cumulative loyalty path over the fixed script. The opening mixes short gains and losses; after #22 (a deliberately provocative endorsement of official media), the path tilts and a dense late block of critical statements in the low-30s and mid-40s drives the final decline. The figure helps locate anchors, plateaus, and zones where polarization accumulates.

Figure 1

Loyalty Cascade for the Debate Sequence



Note. Bars plot the cumulative sum of *L* in presentation order. Blue bars mark positive steps; orange bars mark negative steps; numbers indicate step size. Large and persistent steps indicate anchors; flat segments suggest plateaus or contrast. Source: developed by the authors.

To assess order sensitivity, we keep *O*, *T*, *P*, *B*, and *L* fixed and recompute the cascade for many random permutations and for three counterfactual scripts that place positives early, strictly alternate signs, or place negatives late. We summarize trajectories by four diagnostics: final level, distribution of same sign run lengths,

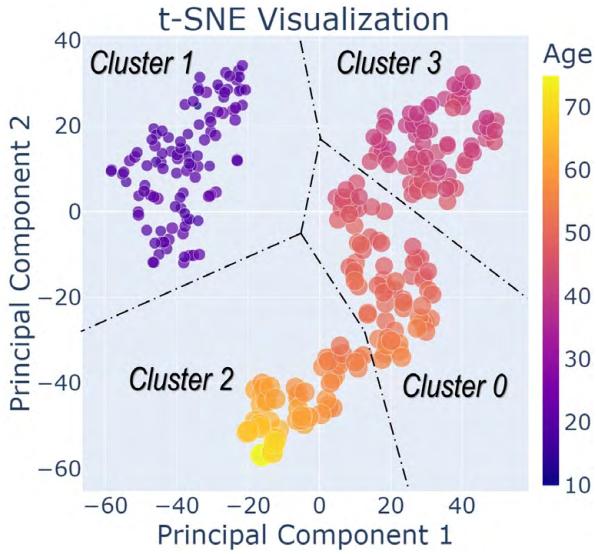
location of the first turning point, and rebound size after sign switches. In the observed sequence, runs are short (mostly length three) whereas random orders of the same content typically yield runs of five to six. A simple pre- and post-check around #22 shows no reliable shift in the average step relative to permutations. Steps after sign switches are not larger than steps inside runs, so contrast is weak. The robust feature is recovery shortfall: after long negative runs, the next positive step is much smaller than the loss it follows. An early-support then late-critique order climbs and then slides, finishing slightly below zero, showing how a dense block of late negatives erases early gains. Strict alternation hovers near zero with large swings and little drift. An early-critique then late-support order drops deeply and rebounds only partly, echoing the observed recovery shortfall. Taken together, the late negative finish reflects the placement and density of strongly disloyal statements in the final third; the cascade simply records that accumulation in order, without implying causation.

Audience Structure

We map respondent similarity using the 50-statement response vectors coded as +1, 0, –1. A *t*-SNE embedding (e.g., van der Maaten & Hinton, 2008) reveals four compact regions aligned with age (Figure 2). We then fit *k*-means (e.g., Lloyd, 1982) with $k = 4$ and identified four clusters, referred to by numeric IDs: Cluster 1 (ages 18–22), Cluster 3 (40–47), Cluster 0 (44–56), and Cluster 2 (52–75). The map provides a stable partition for subgroup analyses without prespecifying sociodemographic cutoffs.

Figure 2

t-SNE Map of 307 Participants Based on Response Similarity Across 50 Statements



Note. Colors show age. Dashed lines mark four *k*-means clusters (0–3). Proximity reflects similar agreement profiles. Source: developed by the authors.

Age organizes the map in Figure 2: four zones coincide with the previously defined cohorts, yielding a tight, internally uniform youth cluster (1), two distinct mid-age islands (3 and 0) with different trust and identity profiles, and a coherent senior ribbon (2) oriented to duty and stability. Cluster 3 tends to support social-benefit frames yet splits on oversight and pride, whereas Cluster 0 is more trust-forward and receptive to patriotic cues. Behavioral markers align with age: youth register the smallest neutral share (about 23%) and switch signs most often; the mid-age cohorts carry many neutrals (about 36% in Cluster 3 and 35% in Cluster 0), consistent with caution and ambivalence; the senior cohort sits between them on neutrality (about 29%) though shows the longest late negative runs.

The clearest fractures concentrate on a short set of statements: older cohorts agree with #48 ("culture must come first; we remember the USSR") and #47 ("censorship is needed"), while youth tend to disagree; youth are much more favorable to #9 (a specifically Russian path to reduce poverty) than the senior cohort; Cluster 3 is uniquely positive on #10 (feasibility and sincerity of the proposed ministry) compared to the rest. Naming clusters by age ranges therefore makes the visualization interpretable, supplies external validity for the grouping, and explains why two mid-age islands appear with distinct trust and identity profiles. These four data-driven clusters define the partition used in the analyses that follow.

Net Loyalty Drift

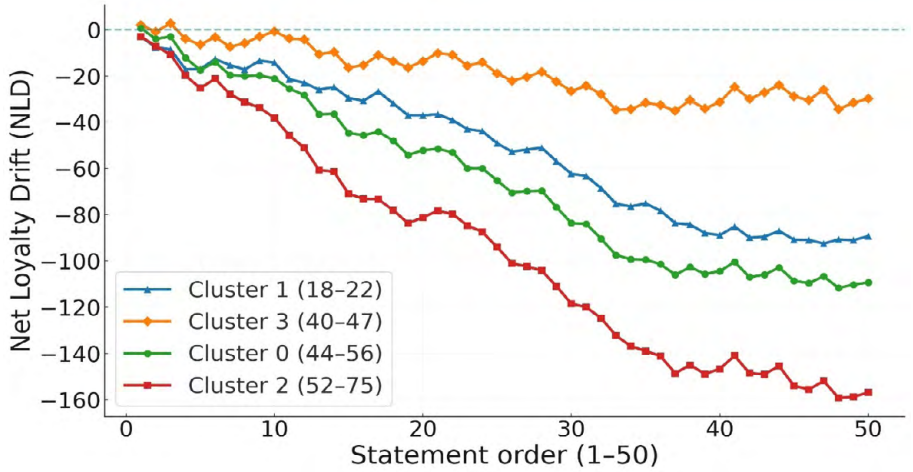
We keep the Loyalty Index L of each statement fixed as defined earlier. For cohort c and statement s in the original order (1–50), let $A_c(s)$ and $D_c(s)$ be the within-cohort shares of "agree" and "disagree." The statement-wise increment of loyalty score associated with the statement s is $L(s)\,[A_c(s) - D_c(s)]$. We define the Net Loyalty Drift (NLD) as the cumulative sum of these increments across the sequence:

$$NLD_c(t) = \sum_{s=1}^{t} L(s)[A_c(s) - D_c(s)]. \quad (2)$$

Positive increments mean the cohort aligns with loyal content or rejects disloyal content of s; negative increments mean alignment with disloyal content or rejection of loyal content; near-zero increments reflect split or neutral reactions. NLD (2) is a content-referenced trajectory that summarizes how a cohort accumulates alignment with the fixed labels of the 50 statements in the observed order.

Figure 3 overlays cohort trajectories with statement content held fixed. NLD (2) rises when a cohort aligns with loyal content or rejects disloyal content of statement s, falls in the opposite case, and is flat when "agree" and "disagree" balance or neutrality dominates. Steeper slopes indicate greater sensitivity to the current frame; large single steps mark *anchors* that move many respondents at once; a deep late decline signals that negative content placed near the end is hard to offset with a few supportive statements. These trajectories summarize group alignment with statement labels and do not label individuals.

Figure 3
Net Loyalty Drift (NLD) by Cohort



Note. Each curve is the cumulative sum, in presentation order, of a statement's LLL multiplied by the within-cohort opinion balance (share "agree" minus share "disagree," with neutrals as zero). Curves are shown for Cluster 1 (ages 18–22), Cluster 3 (40–47), Cluster 0 (44–56), and Cluster 2 (52–75). Source: developed by the authors.

Across cohorts the opening is mixed, and a prolonged negative drift follows statement #22, a provocative affirmation of trust in official media. Differences lie in depth and recovery. Cluster 1 (18–22) drops and rebounds quickly, showing many sharp corrections that track low neutrality; rebounds fade as late disloyal content accumulates. Cluster 3 (40–47) hovers near zero for much of the script and then declines, consistent with support for social-benefit framing coupled with a split between oversight and pride. Cluster 0 (44–56) holds longer under trust-forward and patriotic cues but yields when negative blocks stack, particularly across 32–37. Cluster 2 (52–75) is most inertial: once the late critical frame sets in, the curve falls in long runs with small recoveries and finishes lowest. The shared downward finish indicates that late segments concentrate critical cues whose effects accumulate faster than earlier supportive steps can offset.

Age also orders how the four semantic components shift. On Optimism–Skepticism, Cluster 1 gains on ##1, 3, and 10 and shows a strong rebound at #31, while Clusters 0 and 3 display situational optimism that turns skeptical when negatives repeat; Cluster 2 ends in sustained skepticism, particularly over ##32–37. On Trust–Distrust, statement #22 is the clearest pivot: Cluster 1 reads it distrustfully and bends down at once; Cluster 0, the most trust-forward earlier, loses less immediately and shows more support on trust cues such as ##16 and 20; Cluster 2 trusts cultural and protective themes early but turns distrustful when late critical frames concentrate; Cluster 3 mixes trust with demands for oversight and stays near zero until late. On Patriotism–Critical Perception, older cohorts are positive on #47 ("censorship is

needed") and #48 ("culture must come first; we remember the USSR"), while youth tend to reject both; Cluster 3 is warm to patriotic belonging in ##40–41 yet sympathetic to civic critique in #42 ("patriotism can include public protest against injustice"), which explains the two mid-age islands in Figure 2. On Benefits–Problems, youth respond strongly to benefit-framed #9 and to #10 about feasibility and sincerity; older cohorts are cooler on #9 and shift toward a problem focus when negatives cluster, producing the long downward segments seen for Clusters 0 and 2. In sum, age does not fix people as loyal or disloyal; it shapes how the same fifty statements are integrated over time, yielding distinct cumulative paths by cohort in Figure 3.

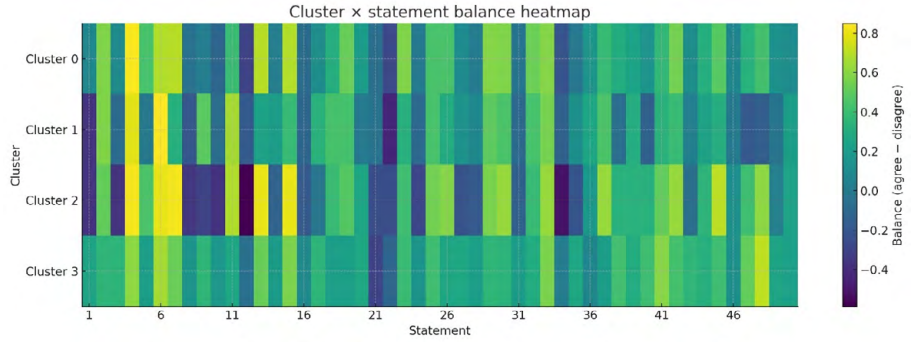
Polarization by Statement and Anchors in NLD

For each statement *s*, we keep its fixed Loyalty Index *L*(*s*) and, for each cohort *c*, compute the within-cohort balance of opinion $b_c(s) = A_c(s) - D_c(s)$, where $A_c(s)$ and $D_c(s)$ are the cohort shares "agree" and "disagree." Polarization is the dispersion of these four balances: we report (i) variance across cohorts, $V(s) = \frac{1}{4}\sum_c \left(b_c(s) - \bar{b}(s)\right)^2$ with $\bar{b}(s) = \frac{1}{4}\sum_c b_c(s)$; (ii) range $R(s) = \max_c b_c(s) - \min_c b_c(s)$; and (iii) the fracture index $F(s) = \frac{1}{6}\sum_{c<c'} |b_c(s) - b_{c'}(s)|$. We rank statements by *F*(*s*) because it rewards wide, multi-sided splits rather than a single outlier. Content weighted variants multiply by $|L(s)|$: $F_L(s) = |L(s)|F(s)$, $R_L(s) = |L(s)|R(s)$.

Figure 4 shows the resulting cohort-by-statement heat map; horizontal bands indicate within-cohort coherence, sharp vertical contrasts mark cross-cohort splits. The ten most divisive statements are ##49, 9, 3, 1, 47, 13, 40, 12, 10, and 7. A compact reading of these ten follows #48 (cultural memory and primacy of culture): older cohorts, especially Cluster 2, tend to agree; Cluster 1 is cooler or negative, a clear line on Patriotism vs. Critical Perception. #9 (a specifically Russian way to reduce poverty): Cluster 1 is most favorable, Cluster 2 is most skeptical, with Clusters 0 and 3 in between; the split lies on Benefits vs. Problems and Optimism vs. Skepticism. #3 (help to low-income groups if reporting and oversight; coded 3, 2, 2, 2 → +10): youth and Cluster 3 respond positively, older cohorts are more guarded; benefits combine with conditional trust. #1 (opening supportive, feasibility frame): endorsed in Clusters 1 and 3, cooler in 0 and 2; early optimism meets mid- and late-life caution. #47 ("censorship is needed"): agreement rises with age; Cluster 1 tends to reject the control frame; Cluster 2 often supports it; Cluster 3 is mixed, strengthening late divergence. #13 (categorical skepticism; –3, –3, –3, –3 →–12): youth resist the blanket claim; older cohorts show more agreement or neutrality; the fracture spans Optimism vs. Skepticism and Trust vs. Distrust. #40 (patriotism as belonging and pride): Clusters 2 and 3 are positive, Cluster 1 cooler, Cluster 0 supportive but less intensely so; an identity anchor on the right of the map. #12 (supportive with institutional prudence): youth and Cluster 3 endorse; older cohorts keep trust conditional. #10 (sincerity and feasibility of the ministry): Cluster 3 is notably positive, Cluster 1 favorable, Clusters 0 and 2 cooler; benefits meet perceived capacity. #7 (early critical claim): youth push back; older cohorts are divided; it amplifies early dispersion on problems and distrust.

Figure 4

Cluster by Statement Balance Heat Map



Note. Cells show $b_c(s) = A_c(s) - D_c(s)$. Rows are age cohorts: $C1$ = 18–22, $C3$ = 40–47, $C0$ = 44–56, $C2$ = 52–75. Columns follow the presentation order (1–50). Color encodes sign and magnitude. Horizontal bands denote within-cluster coherence; vertical contrasts mark polarization by statement.

Identifying anchors and turning points in NLD, we use the same inputs. With $b_c(s)$ as defined above, the *alignment increment* for cohort c on statement s is $\Delta_{c,s} = L(s)\, b_c(s)$; and the cohort's NLD cascade is defined by

$$NLD_{c,t} = \sum_{s \le t} \Delta_{c,s} \tag{3}$$

A *statement anchor* for cohort c has large $|\Delta_{c,s}|$ and drives the biggest single-step change; a *turning point* is a local extremum of NLD_c that ends a same-sign run.

Anchors in the NLD curves are the statements that shift a cohort most strongly at once. Early supportive cues act as anchors for youth (#1, #3, #10). A late block of critical material produces the largest downward anchors for the older cohorts, and it also drives the shared late decline for everyone. The clearest pivot is #22, a provocative affirmation of trust in official media, after which all curves tend to drift down. A rare strong positive anchor appears for the 40–47 cohort on an identity-patriotic cue (#41). A small set of late statements around the 30s sustains the longest negative runs, especially for the senior cohort. Turning points are where a cohort's curve changes direction. The principal turning point across cohorts is the downward bend after #22. Later brief peaks around the low 30s or early 40s are short-lived and followed by renewed decline. Youth recover fastest but the rebounds fade when late negatives cluster; the 40–47 cohort holds close to neutral for a long stretch before sliding; the 44–56 cohort resists longer under trust-forward and patriotic cues but yields when negative blocks stack; the 52–75 cohort falls in long runs with small recoveries once the late critical frame consolidates. Labels attach to statements, not to people. The figures record how age structures reception of the same content and how a few statements organize where cohorts part ways.

Axis Weights by Cohort: Logistic Agreement Model

We estimate how the four semantic components predict agreement at the person–statement level. Each statement carries four signed scores on a common –3 … +3 scale: *O* = Optimism vs. Skepticism, *T* = Trust vs. Distrust, *P* = Patriotism vs. Critical Perception, *B* = Benefits vs. Problems. For respondent *i* in cohort *c* evaluating statement *s*, the outcome is coded as "agree" (1) and "cannot evaluate / disagree" (0). We fit separate binomial generalized linear models (logit link) for each cohort with standard errors clustered by respondent (e.g., McCullagh & Nelder, 1989).

For cohort *c* the model specification is

$$logit\ Pr\ (agree_{i,s}|c) = a_c + \beta_{O,c}O(s) + \beta_{T,c}T(s) + \beta_{P,c}P(s) + \beta_{B,c}B(s), \quad (5)$$

where *logit p* = log[*p* / (1–*p*)]. Coefficients are log-odds; exp (*β*) gives the odds ratio. Axes are entered as coded (–3 … +3) so one unit is one step along that axis.

Across all cohorts Optimism (*O*) carries a negative weight: overtly optimistic forecasting lowers the odds of agreement once other content is held fixed. Benefits (*B*) carries a positive weight in every cohort and is strongest from midlife onward: concrete benefit framing reliably raises agreement, especially among the age group 40+. Patriotism (*P*) is most positive for the youngest cohort (18–22) and smaller elsewhere: identity cues resonate with youth but are uneven in older groups. Trust (*T*) adds little on average once the other axes are in the model: generic "trust" language has limited marginal pull relative to benefits, identity, and forecast tone. These axis weights are consistent with the age patterns seen in the maps and cascades: youth move on identity and benefits, mid-age and older cohorts reward tangible benefits and discount optimistic projection, and generalized trust contributes weakly once content is specified.

Age–Gender Loyalty Framing Responsiveness (AGLFR)

The final section analyzes the second part of the experiment: ten statements evaluated by an older sample (*N* = 200; ages 40–75). Unlike the youth sample, these respondents reported gender, which enabled age–gender contrasts. We keep each statement's fixed Loyalty Index *L*(*s*). For a subgroup *g* (men or women at a given age) let $A_g(s)$ and $D_g(s)$ be the shares "agree" and "disagree." We define the subgroup's responsiveness as the sign-free sum of absolute contributions across the statements:

$$AGLFR_g = \Sigma_s|L(s)(A_g(s) - D_g(s)|. \quad (6)$$

Agreement with loyal statements and disagreement with disloyal statements raise AGLFR; the opposite pair also raise it. Neutral or perfectly split reactions add zero. Thus AGLFR defined in (6) measures how strongly a subgroup reacts to loyalty-framed content, not which side it prefers.

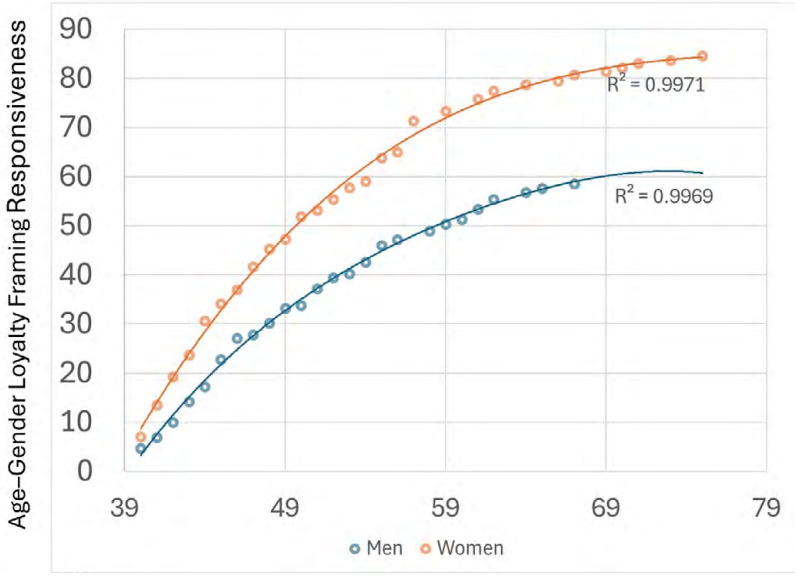
Figure 5 plots AGLFR by age for men and women (points) with smooth polynomial fits (lines). Within the 40–75 range the curves rise with age and then level off in the early seventies, indicating that responsiveness strengthens across later adulthood and approaches a ceiling for this item set. At every age, women score above men, and the

gap widens with age. The higher female curve reflects more decisive reactions—either endorsement of loyal cues (order, protection, civic duty) or firm rejection of disloyal cues (blanket distrust, categorical cynicism); the lower male curve reflects more neutrality or internal balance on the same cues. Most of the aggregate comes from statements with large $|L(s)|$ paired with clear subgroup responses; where a subgroup divides or remains neutral, contributions are near zero and the curve flattens.

This analysis is complemented to the main 50-statement debate, where Net Loyalty Drift tracked a signed accumulation over order. Here the sign is removed and only the magnitude of engagement is summarized for age–gender subgroups in the ten-statement module. Read together, the two summaries show where the debate ends and how strongly different groups engage with loyalty framing, without labeling individuals or inferring causes.

Figure 5

Age–Gender Loyalty Framing Responsiveness (by Age for Men and Women)



Note. Points show subgroup totals by age for men and women; smooth lines are polynomial fits. Higher values mean stronger cumulative responsiveness to loyalty-framed cues, not preference for a side. Source: developed by the authors

Discussion and Conclusion

Polarization in this debate is best understood as an interaction between fixed message content and audience predispositions observed over time. Our contribution is methodological and substantive. Methodologically, we work with content that is coded once and then held fixed, and we track reception as it unfolds in sequence. This yields transparent, reproducible descriptors that do not require latent variables or causal identification.

The four-score loyalty index introduced in our prior work serves here as a compact content space; the present study extends it in four ways. First, we make the temporal forecast dimension explicit, separating future-facing affirmations from other forms of supportive content. Second, we introduce a cumulative, order-sensitive description of reception (the loyalty cascade) and its cohort-specific counterpart (Net Loyalty Drift), which record how alignment with loyal or disloyal framing accumulates as the script proceeds. Third, we complement signed accumulation with a magnitude-only measure of engagement (Age–Gender Loyalty Framing Responsiveness) to capture how strongly subgroups react to loyalty-framed cues regardless of stance. Fourth, we quantify dispersion at the item level (variance, range, and a fracture index over cohort balances) to identify a compact "vocabulary of fractures," and we locate anchors and turning points in the trajectories without imposing a model of persuasion. Together, these tools provide a process-trace of polarization that is comparable across cohorts and interpretable down to specific statements and frames.

Age organizes both structure and dynamics. Youth engage decisively and adjust rapidly: early gains on feasibility and benefit frames are followed by quick corrections when control or distrust cues appear, though a dense late block of disloyal content still pulls the path down. Mid-life respondents bifurcate into two islands with distinct trust and identity profiles: one starts trust-forward and patriotic yet yields under sustained critique; the other supports social-benefit claims but hesitates where pride and oversight conflict. The senior cohort is steady early and then shows the deepest sustained decline once late critical statements cluster, with the clearest recovery shortfall after long negative runs. These age-graded patterns fit motivated reasoning and identity-congruent interpretation—congenial frames are assimilated, dissonant frames resisted until counterevidence passes a threshold (Kunda, 1990; Taber & Lodge, 2006)—and align with value-alignment perspectives in which identity and authority cues are read differently across generations (Haidt, 2012).

Polarization concentrates on a small set of statements. Identity and control themes divide generations (older participants endorse cultural primacy and tighter information control; youth reject them), while reform-and-benefit themes divide in the opposite direction (youth and younger mid-life are more favorable to concrete, feasible gains; older cohorts shift to a problem-first reading when critical frames repeat). A deliberately provocative trust cue mid-sequence is the clearest pivot: younger respondents bend downward at once, older respondents later—consistent with generational gaps in institutional trust and with framing work showing that diagnostic labels steer interpretation (Thibodeau & Boroditsky, 2011). The long late slide across cohorts mirrors gain–loss asymmetries: once the debate becomes problem-laden and risk-framed, loss sensitivity dominates and supportive framing loses traction (Kahneman & Tversky, 1979). Social Identity Theory clarifies why the same patriotic or protectionist language yields opposite readings across age groups—statements are filtered through in-group norms and identity commitments (Tajfel & Turner, 1979). In Lakoff's terms, audiences inhabit different frame grammars; identical words recruit different value complexes across cohorts (Lakoff, 2004).

Gender further stratifies engagement among older respondents. With content fixed, women show consistently higher sign-free responsiveness than men at the same ages, and the gap widens with age. This indicates stronger, more unified reactions, either endorsement of loyal protection-and-duty cues or firm rejection of cynical blanket distrust—while men's aggregate responses more often cancel out through internal splits. The finding coheres with moral-psychological accounts in which communal protection and social order grow in salience for many women over the life course, and with evidence that identical frames can polarize men internally while producing clearer stances among women.

The broader contribution is a flexible grammar for debates. Holding content fixed and reading trajectories over order yields a mid-level account between qualitative narrative and causal identification: where polarization grows, which statements carry disproportionate weight, and how subgroup paths diverge or converge—without labeling individuals or imputing hidden traits. Classic frameworks supply mechanisms (identity cleavages, motivated resistance and tipping, value-bound framing), while our measures provide the process trace that shows when in a discourse those mechanisms bite and which statements activate them. For applied fields in civic communication, public consultation, and media studies, the value is diagnostic rather than prescriptive: scripts can be audited ex post for order sensitivity, anchor density, and cohort fractures; sign-free modules can gauge engagement magnitude; and audiences can be segmented empirically by response profiles rather than by assumed ideology. None of this asserts causation; it clarifies how a common agenda yields distinct cumulative receptions across a real sequence of arguments. Future work should test generality across topics and venues, preregister order audits, and examine whether cohort and gender patterns persist when content foregrounds justice, autonomy, risk, or security rather than culture and control; bridging research and practice will require careful ethics, since these tools are for diagnosis and understanding, not manipulation.

References

Cobb, R. W., & Elder, C. D. (1972). *Participation in American politics: The dynamics of agenda-building*. Allyn & Bacon.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall/CRC.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lakoff, G. (2004). *Don't think of an elephant!: Know your values and frame the debate*. Chelsea Green.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.

Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE*, *6*(2), Article e16782. https://doi.org/10.1371/journal.pone.0016782

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Volchenkov, D., & Lebedev, A. N. (2025). Diskursivnaia poliarizatsiia i vozrastno-gendernye tsennosti: kombinirovannyi analiz reaktsii na initsiativu "Ministerstvo schast'ia" [Discursive polarization and age–gender value structures: A combined analysis of reactions to the "Ministry of Happiness" initiative]. *Proceedings of the Institute of Psychology of the Russian Academy of Sciences*, *5*(2), 3–15.

Appendix

Key Statements and Their Loyalty Assessment (in Russian)

Statement 1 (loyal; Loyalty Index = 8): Друзья, В. И. Матвиенко несколько раз в своих выступлениях предлагала создать в нашей стране так называемое Министерство счастья, которое будет рассматривать все мероприятия, проводимые руководством страны, на предмет того, делают ли эти мероприятия людей более счастливыми или они бесполезны. Я думаю, что это хорошее предложение.

Statement 2 (disloyal; Loyalty Index = –8): Несколько раз мне это попадалось в СМИ. Если я правильно понимаю, Матвиенко предлагает, чтобы это министерство оценивало любые решения, законопроекты, постановления правительства и так далее на предмет того, как они повлияют на всеобщее счастье. Но непонятно, как это будет делаться. Я знаю, что такие министерства есть в некоторых странах, например, в ОАЭ и в Нигерии. Но их практика говорит о том, что введение Министерства счастья слабо коррелирует с тем, что люди чувствуют, потому что показатели говорят об экономике, а не о том, счастливы ли люди на самом деле или нет.

Statement 3 (loyal; Loyalty Index = 10): Я думаю, что если создается государственная организация, то она обязана будет отчитываться перед обществом, так как получает финансирование. Поэтому для какой-то части населения, хотя бы для малообеспеченной, будут разрабатываться какие-то проекты, которые принесут пользу. Думаю, что такое министерство обязательно будет способствовать улучшению жизни.

Statement 4 (disloyal; Loyalty Index = –11): А что будет делать это министерство? Какие будет решать проблемы, чтобы добиться всеобщего счастья? Мы будем с бедностью таким образом бороться? Я лично этого не понимаю. Вообще-то вопрос этот, наверное, должен был бы сопровождаться какой-то дополнительной информацией для населения о том, что собой представляет этот проект. Иначе это лишь очередная акция популизма.

Statement 5 (disloyal; Loyalty Index = –12): А я вот не думаю, что оно с бедностью будет бороться. Ведь бедность в стране выгодна, потому что бедный человек, который живет на 15 тысяч, куда идет? Он идет воевать или обслуживать богатых людей.

Statement 6 (loyal; Loyalty Index = 5): Друзья, мне кажется, что мы уходим от темы. Начали с Министерства счастья, а ушли в какое-то Министерство бедности. Если вернуться к теме, то я хочу сказать, что нет понятных критериев счастья. Для одного счастье – петь под гитару в походе, а для другого – съездить на Мальдивы.

Statement 7 (disloyal; Loyalty Index = –8): Мне тоже вот не совсем ясно, что это за министерство будет? И так армия чиновников превышает все мыслимые и немыслимые по численности пределы. Мне кажется, что это подмена понятий, манипуляция. Опять же это все для чиновников, которые будут жить хорошо в этом министерстве.

Statement 8 (loyal; Loyalty Index = 10): Я тоже хочу сказать, потому что со многими здесь не могу согласиться. Идея помочь людям стать менее бедными и как-то их поддержать мне нравится. Что в этом плохого? А на самом деле злобные разговоры с критикой идеи В. Матвиенко о создании подобного министерства – это просто попытка оппозиции раздуть из мухи слона.

Statement 9 (loyal; Loyalty Index = 8): Я считаю, что в России вообще должно быть какое-то свое понимание того, как можно бедность уменьшить. Я знаю, что какие-то меры сейчас принимаются, чтобы с бедностью бороться. Какие-то есть выплаты, социальная поддержка и так далее. И с каждым годом это увеличивается. Поэтому я лично поддерживаю эту идею.

Statement 10 (loyal; Loyalty Index = 12): Про то, что Министерство счастья – это вполне возможный проект, сомнений не может быть. Это же не желтая пресса придумала, это сказано третьим лицом государства по уровню полномочий. И я не думаю, что в данном случае Матвиенко преследовала какие-то личные корыстные цели, например, возглавить это министерство. У нее и так высокая должность. Думаю, это было искреннее желание помочь людям.

Statement 11 (disloyal; Loyalty Index = –11): Позвольте я по поводу как раз бедности и счастья хочу сказать, поскольку это уже не первый раз звучит. Во-первых, никакой корреляции, насколько я представляю, между небедностью и счастьем не существует. Говорить, что бедный человек заведомо несчастен, это все-таки манипуляция. Потому что бедный он или не бедный, каждый для себя сам определяет. Идея с министерством совсем не понятна. Зачем это?

Statement 12 (loyal; Loyalty Index = 9): Кстати, возможно, вы не знаете или не помните, но сама Матвиенко, когда обсуждали пенсионную реформу, сообщила, что лично у нее пенсия, оказывается, всего 25 тысяч рублей. Так что я думаю, что про Министерство счастья она вполне искренне это сказала. Разве нет?

Statement 13 (disloyal; Loyalty Index = –12): Ну не может у нее быть пенсия 25 тысяч рублей. Это смешно. Здесь вы что-то не то говорите.

Statement 14 (loyal; Loyalty Index = 5): Почему же? Даже вот у научных сотрудников зарплата сегодня очень маленькая, но есть некие надбавки, которые на пенсионных накоплениях не отражаются. Эти надбавки часто превышают зарплату. То же самое может быть и здесь.

Statement 15 (disloyal; Loyalty Index = –12): Нет, оклад главы Совета Федерации не может быть таким, чтобы из него пенсия получилась в 25 тысяч (рублей).

Statement 16 (loyal; Loyalty Index = 12): По поводу какой-то корысти Матвиенко это, скорее всего, очередной фейк. Я лично по поводу фейков могу сказать, что те, кто их сознательно сочиняет и вбрасывает в социальные сети, это люди, которые всегда чем-то недовольны. Они во всем видят негатив, особенно это те, у кого жизнь не сложилась. Проще говоря, это всегда неудачники, которые хотят хайпануть и прежде всего на успешных людях. Лучше всего доверять официальным СМИ, они уж точно фейки не распространяют. У них в этом нет необходимости...

Statement 17 (loyal; Loyalty Index = 12): А я думаю, что на самом деле все-таки должна быть такая организация, которая будет заниматься вопросами борьбы с бедностью, потому что президент постоянно эту задачу ставит. Он вот точно заинтересован, чтобы бедных не было.

Statement 18 (disloyal; Loyalty Index = –12): А может быть, это был такой популистский ход? Может быть, это чисто слова, сказанные для аудитории, такая индивидуальная фишка: «Вот Матвиенко говорит о Министерстве счастья, значит, она за народ, за счастье всех людей».

Statement 19 (disloyal; Loyalty Index = –12): Очень может быть, так как в Госдуме, например, депутаты, которые принимают законы, получают за законопроекты определенные деньги, и это деньги немалые. И им нужно, чтобы их фамилии все время были на слуху. То есть они получают дополнительные деньги кроме своей зарплаты. Поэтому они часто предлагают то, за что Дума явно не проголосует.

Statement 20 (loyal; Loyalty Index = 10): А это официальный источник или Интернет? Мне лично такого никогда не попадало. Интернету вообще нельзя доверять.

Statement 21 (disloyal; Loyalty Index = –12): Почему нельзя? По-моему, только интернету и можно доверять. Ну не телевизору же?

Statement 22 (loyal; Loyalty Index = 6): А чем вам не нравится телевизор? Официальная информация всегда надежнее. Она же проверенная.

Statement 23 (disloyal; Loyalty Index = –12): И мне кажется, что люди в Думе просто хотят чем-то выделяться, то есть одна – министерством счастья, другой – законом против геев, например, третий еще чем-то. Вот Милонова знают как великого борца с ЛГБТ-сообществом[7], а другого знают еще как-то. И все они как-то выделяются, и это как раз элемент просто собственного пиара на фоне других.

Statement 24 (loyal; Loyalty Index = 10): Друзья, министерство счастья или несчастья – неважно, но вообще это министерство интересная штука. Матвиенко ведь очень активная женщина. Она же до этого работала в Петербурге. Сейчас все жалуются на нового губернатора, а при Матвиенко ведь было все хорошо. Что-то такого я не помню, чтобы на нее жаловались. Сейчас какие-то отдельные меры предпринимаются, но почему бы их в обязанности отдельного министерства не включить?

Statement 25 (disloyal; Loyalty Index = –12): Не уверен. Мне кажется, что министерством счастья все-таки должен являться парламент. И это должно быть свободным творчеством на благо людей в рамках законодательного органа. И не надо сваливать это на какое-то одно министерство. Занимайтесь с помощью депутатских своих полномочий этим делом.

Statement 26 (disloyal; Loyalty Index = –12): Я еще хочу добавить. Я считаю, что если будет министерство бедности, то бедность сама себя не изживет. Думаю, что если будет министерство борьбы с бедностью, то бедности будет больше. Потому что любое ведомство стремится увеличить степень своего контроля и расширить то, что оно регулирует. И оно объективно будет

стремиться не к тому, чтобы бедных стало меньше, а к тому, чтобы бедных стало больше, потому что тогда у этого министерства вырастают полномочия, финансирование, статус.

Statement 27 (loyal; Loyalty Index = 6): А вот у меня другое мнение... Сначала кажется, что это наивная такая идея, но такое Министерство могло бы заниматься, например, тем, что можно измерять: экономические показатели, эмоциональное состояние населения и др. Я знаю примеры и в США, и в других странах. Там есть такие программы для помогающих профессий. Я считаю, что это очень нужно, и это министерство могло бы возобновить старую традицию СССР по профилактике здоровья и не только физического, но и эмоционального.

Statement 28 (loyal; Loyalty Index = 10): А я эту идею поддерживаю. Мне кажется, что сейчас у нас в стране многое становится более прозрачным. Точно не знаю, но у меня больше позитива в плане развития страны в будущем. Сейчас намечаются огромные перспективы.

Statement 29 (disloyal; Loyalty Index = –12): Не уверен. Сегодня вообще нельзя ничего сказать о том, что с нами будет лет через десять.

Statement 30 (disloyal; Loyalty Index = –12): Почему через десять? Сегодня нельзя сказать даже о том, что будет через пару лет. Все очень непонятно.

Statement 31 (loyal; Loyalty Index = 11): Ну и почему у вас такой пессимизм? Мне кажется, что все очень даже ясно, что с нами будет. Будем развиваться и даже еще более быстрыми темпами, чем раньше. И экономика будет расти, а уровень жизни уже вырос, и дефицита нет, как было в 90-е.

Statement 32 (disloyal; Loyalty Index = –12): И все-таки… Мне кажется, что мы ушли от темы. У меня возникает вопрос: а каковы будут задачи у этого министерства? Проводить замеры? Может быть, в какой-нибудь Финляндии это имеет смысл, но в наших реалиях я это воспринимаю как структуру чисто репрессивную. Я не могу от нее ожидать ничего другого. Если это будут замеры, то я не думаю, что будет чем похвастаться перед Финляндией, например. То есть надо будет как-то подкручивать, улучшать показатели. Я не очень понимаю, как по-другому это может работать в нынешних реалиях...

Statement 33 (disloyal; Loyalty Index = –12): Так и будет работать – скручивать, накручивать, убирать ненужное. Поймите, никто вам правду не скажет. Люди боятся говорить правду. А как определить, ты счастлив или нет? Критерий счастья у каждого свой.

Statement 34 (loyal; Loyalty Index = 9): А я вот думаю, что в новом министерстве могут работать люди, которые будут работать профессионально. Сегодня ведь есть такие структуры в силовых ведомствах, например, которые ловят «оборотней в погонах», и они неплохо работают. Разве нет? Ведь у нас Минфин поддерживает экономику, а Министерство промышленности развивает производство. Поэтому говорить, что министерство счастья – это полная ерунда, наверное, неправильно.

Statement 35 (loyal; Loyalty Index = 10): Знаете ли, вот сегодня все вроде как отвечают за счастье, но нет какого-то конкретного ответственного лица или подразделения... Непонятно к кому вообще можно обратиться. То есть

была бы польза, если бы это была отдельная структура, в которую можно прийти, зайти на сайт и понять, что она разработала, чтобы кто-то конкретный нес ответственность за решения. Сейчас у нас, между прочим, обратная связь с населением уже неплохо налажена...

Statement 36 (disloyal; Loyalty Index = –12): Я сейчас слушаю и вот, что меня зацепило. Представилось мне следующее. Например, ввели параметры счастья и вот... Семья влияет на ощущение счастья? Значит – уголовная ответственность за развод! Низкая рождаемость? Аборты запретим! То есть не увеличим социальную защиту для одиноких женщин с детьми, а просто аборты запретим!

Statement 37 (disloyal; Loyalty Index = –12): Да и если так пойдет, то у нас еще сделают Министерство пропаганды и Министерство патриотизма, о необходимости которых уже высказывались некоторые депутаты, и это ведь будет ужасно! Разве нет?

Statement 38 (loyal; Loyalty Index = 11): Знаете, если такое Министерство счастья появится, то должны иметь место и пропаганда, и его популяризация, но к этому нужно подходить очень грамотно. Хотя я лично думаю, что пропаганда нужна. Она в любом обществе есть. А у нас ее явно не хватает. Нужно пропагандировать ценности общества, а не просто о них говорить.

Statement 39 (disloyal; Loyalty Index = –12): Да вы что такое говорите? У нас и так пропаганды слишком много! Нужны независимые СМИ! А их практически нет!

Statement 40 (loyal; Loyalty Index = 7): Я думаю, что чувство счастья формируется, и оно зависит не только от материальных условий жизни, но и, например, от чувства причастности к чему-то важному для человека. Я думаю, что в понятие счастья обязательно должно входить чувство патриотизма. И это решаемый вопрос. Сейчас люди с фронта возвращаются с боевым опытом. Предлагается, разумеется, после специального обучения педагогическим навыкам их привлекать к работе в школах и вузах. И мне кажется, это тоже очень важно и даст положительный эффект.

Statement 41 (loyal; Loyalty Index = 11): Я согласен, но ведь патриотизм тоже по-разному можно оценивать, как и счастье. Вот есть патриоты, которые только кричат о любви к Родине, а на самом деле они ее защищать на фронте не торопятся. Сейчас решается судьба России и каждый должен что-то сделать, чтобы победить. И оказалось, что весь мир против России, потому что правда на нашей стороне. Да что я говорю, это все сегодня понимают. По крайней мере должны понимать.

Statement 42 (disloyal; Loyalty Index = –12): Патриотизм разный бывает. Это точно! Но одно дело – это принимать все, что сверху скажут, а другое – выступать против несправедливости, которой сегодня много потому, что есть и очень богатые, и очень бедные! Патриоты – это и те, кто может публично высказать свое мнение и не боится.

Statement 43 (loyal; Loyalty Index = 8): С бедностью, конечно, нужно бороться, но государство не может сразу всех материально обеспечивать

на высоком уровне. Поэтому, например, выходить на митинги – это даже аморально. Работать нужно больше, выполнять задачи общества.

Statement 44 (loyal; Loyalty Index = 10): Я думаю, что митинги устраивать – это не аморально, а бессмысленно. Это только время терять. Можно со многими вещами не соглашаться, но попасть в иностранные агенты и даже в тюрьму я лично считаю глупостью.

Statement 45 (disloyal; Loyalty Index = –12): А я вот не согласен с вами. Когда затопило Оренбургскую область, то люди выступили против местных начальников и правильно сделали. Иначе им вообще бы никакой помощи не было. Власть нужно контролировать и не бояться, потому что ей доверять нельзя.

Statement 46 (disloyal; Loyalty Index = –12): Какие митинги? Побойтесь вы бога! Люди всего боятся. Не то, что выступить с требованиями, а просто говорить что-то против начальства. Никто никуда у нас не пойдет. Эти времена уже прошли давно! Все контролирует цензура.

Statement 47 (loyal; Loyalty Index = 9): А почему нет? Цензура очень даже нужна. Все эти «голые вечеринки», когда люди на фронте гибнут, разве нормально? В кино одна порнуха, по телевизору тоже пошлость одна. Юмористы такое выдают, что в СССР их давно бы уже пересажали. Культура резко упала.

Statement 48 (disloyal; Loyalty Index = –12): Да, я с вами согласен. Мы родились в СССР и всё помним. Фильмы были хорошие и люди тоже. А сейчас каждый старается больше заработать любыми средствами. И даже не заработать, а хапнуть побольше. И люди уже ничего не стесняются. И с этим надо что-то делать.

Statement 49 (loyal; Loyalty Index = 11): Я думаю, что если министерство счастья и появится, то оно должно прежде всего заняться именно культурой.

Statement 50 (loyal; Loyalty Index = 8): Да ладно вам! Не смешите! У нас нищих только официально 10 миллионов. Какая тут культура у людей, у которых денег даже на коммуналку и еду не хватает.