An Open-Ended Pandemic? Cross-Cutting Cleavages and the Avoidance or Ambivalence of Public Opinion

Joshua D. Clinton* Menglin (Miley) Liu†

May 8, 2026

Abstract

A complication when studying the effect of cross-cutting social divisions on public opinion is the difficulty of determining whether the more nosier and moderate responses of cross-pressured individuals reflect genuine ambivalence due to competing considerations or evidence of satisficing response behavior that avoids dealing with internal conflict. A prominent example of both the existence of cross-cutting cleavages and the importance of discerning the meaning of responses by cross-pressured individuals resulting from the consequential cleavages between partisanship compliance with public health suggestions related to masking and vaccination that sometimes aligned with partisan identity during the COVID-19 pandemic. By combining forced-choice thermometer scores with open-ended responses, we show how the results of each question type helps inform interpret the meaning of cross-pressured survey behavior. We show that individuals who experienced conflicting cues based on their partisan identity and public health recommendations gave survey responses that were more consistent with ambivalence than avoidance. We argue that integrating open-ended and forced-choice questions – perhaps by leveraging LLMs – helps interpret the meaning of survey responses when respondents possess internal tensions that may result in them being either ambivalent or reluctant to engage with that conflict when taking a survey.

*Abby and Jon Winkelried Professor, Department of Political Science, Vanderbilt University. Email: josh.clinton@vanderbilt.edu, Website: www.joshclinton.com

†Assistant Professor of Computational Social Science, The Chinese University of Hong Kong, Shenzhen. Email: menglinliu@cuhk.edu.cn, Website: https://menglinmileyliu.github.io

In an increasingly polarized and fragmented society, what happens when new, cross-cutting cleavages emerge (Lipset and Rokkan, 1967)? While some have hypothesized that a shared threat can overcome polarization by fostering common purpose (Sherif et al., 1961; Kam and Ramos, 2008), the COVID-19 pandemic revealed the opposite: partisan differences fractured along new lines tied to compliance with public health recommendations on masking and vaccination (Clinton et al., 2021; Gadarian, Goodman and Pepinsky, 2023). Characterizing such divisions is difficult because being cross-pressured may relate to how respondents answer questions involving that internal conflict. Faced with dissonance between competing identities, individuals may respond with genuine ambivalence—or with satisficing behavior that avoids confronting the conflict altogether (Krosnick, 1991). Discerning between the two is critical for understanding the nature of political divisions and the composition of the polity (Ansolabehere, Rodden and Snyder, 2008; Fowler et al., 2023; Broockman and Lauderdale, 2025; Hillygus and Shields, 2008), but it is also difficult because the two possibilities are observationally equivalent in forced-choice survey data—especially in online surveys where respondents are compensated upon completion and may be prone to satisfice (Berinsky, Margolis and Sances, 2014; Peer et al., 2022; Stagnaro et al., 2025).[1]

The COVID-19 pandemic provided a rare, but unfortunate, opportunity to examine this problem. Public health guidance framed compliance as a moral act with life-and-death consequences (Macedo and Lee, 2025), and elite messaging likely deepened the resulting asymmetry in opinions and behavior (Clinton et al., 2021). Democratic elites delivered a unified pro-compliance message, while Republican elites were divided—some endorsed the measures, others challenged their necessity, and still others remained silent (Green et al., 2020; Box-Steffensmeier and Moses, 2021). Such asymmetry was not only associated with policy views (Lenz, 2012), but also likely related to the evaluative reactions partisans formed toward compliers and non-compliers. For compliant Democrats, partisan identity and personal behavior reinforced one another. For non-compliant Democrats and many Republicans, they did not.

Using this context, we assess whether moderate and disparate opinions toward compliers and non-compliers in forced-choice items more likely reflect genuine ambivalence or strategic avoidance in cross-pressured respondents. Pairing forced-choice thermometer scores with open-ended responses that are analyzed using both human coders and large language models allows to ex-

[1] A related interpretive challenge arises in the treatment of "don't know" responses. Berinsky (2004) demonstrates that respondents who select "don't know" differ systematically from those who offer substantive answers, and that excluding them can distort estimates of public opinion. Our concern is analogous but distinct: rather than asking whether non-response is politically meaningful, we ask whether moderate responses from cross-pressured individuals reflect genuine attitudinal content or the functional equivalent of non-engagement.

amine the similarity and differences in responses across these types of item both within individuals and between various levels of cross-pressure to evaluate whether responses are more likely to arise from ambivalence or avoidance (Zaller, 1992; Grice, 1975). We show that the responses that cross-pressured individuals provided reveal neither the echos of co-partisan cues nor cognitive disengagement. Instead, the moderate and disperse forced-choices we observe reflect the variation we observe in the open-ended responses – revealing distinctive language across the various groups and language that is both less emotionally intense and more evaluatively complex than individuals for whom their partisan and compliance identities are reinforcing.

More broadly, does our approach reveal that survey responses by cross-pressured individuals during the COVID-19 pandemic reflect substantively meaningful ambivalence, but it also demonstrates the importance of combining multiple assessments in survey research given increasing societal divisions and cleavages to ensure that our measures capture genuine engagement rather than than avoidance. Especially with the advent of LLM-based coding, incorporating open-ended responses when there is the potential for satisficing may help ensure that the responses reflect respondent cognition rather than mere click-throughs.

1 Divisions, Opinions, and Survey Responses

Scholars have long questioned how well survey responses reflect the underlying attitudes of respondents. Converse (1964), for example, argued that most citizens hold "non-attitudes" rather than coherent and stable belief systems, and Zaller (1992)'s receive–accept–sample model suggests that even if respondents have coherent and stable beliefs, the survey responses they provide may reflect whichever cues or considerations are most accessible and salient at the moment the respondent is interviewed.

The difficulty of interpreting the meaning of a survey response is arguably greatest in the presence of new, cross-cutting cleavages that can create internal tensions for individuals (Lipset and Rokkan, 1967; Powell, 1976). It is well-established that political parties structure both the social identities (Green, Palmquist and Schickler, 2002) and expressed opinions (Zaller, 1992) of individuals, but new issues can create tensions and pressures (Hillygus and Shields, 2008; Brader and Marcus, 2013) that may affect how cross-pressured individuals respond to survey questions about that internal conflict. Because the considerations that come to mind may expose internal tensions and/or ambiguity, it becomes difficult to know whether the moderate and/or inconsistent

answers to survey questions by cross-pressured individuals reflects genuine ambivalence or an effort to avoid confronting internal conflict through satisficing.

The extant literature suggests that both behaviors are certainly plausible. Ambivalence occurs when individuals hold competing but simultaneously accessible considerations, producing internal conflict and evaluative complexity (Thompson and Zanna, 1995; Lavine, 2001; Keele and Wolak, 2008; Lavine, Johnston and Steenbergen, 2012). Conflicting considerations coexist in memory and shape political judgments through their simultaneous activation; moderate forced-choice responses, in this case, may indicate authentic internal conflict. Satisficing, by contrast, reflects minimal cognitive effort: respondents offer superficial or default answers to avoid the burden of deliberation or the discomfort of dissonance (Festinger, 1957; Converse, 1964; Krosnick, 1991). Here, moderate responses signal not conflict but disengagement—a strategy to sidestep incompatible beliefs. The response patterns may appear similar, but the underlying mechanisms differ: moderate responses born of genuine internal conflict are politically meaningful in ways that disengaged responses are not. Disentangling these possibilities requires moving beyond forced-choice responses alone.

Researchers have previously used open-ended survey responses to probe the authenticity of attitudes expressed in closed-ended formats (Converse, 1970; Schuman and Presser, 1981; Bishop, Oldendick and Tuchfarber, 1984; Bishop, Tuchfarber and Oldendick, 1986) and, more recently, to assess respondent engagement with experimental manipulations (Ziegler, 2022) and to reveal substantive distinctions obscured by forced-choice measures (Allamong et al., forthcoming). Indeed, open-ended responses can surface nuanced perspectives that closed-ended formats systematically miss, particularly when opinions are multifaceted (Bernhard-Harrer and Pfaff, 2025). We extend this tradition by integrating forced-choice and open-ended items to determine whether the responses of cross-pressured individuals more likely reflect avoidance or ambivalence. Individuals whose identities and behavior are self-reinforcing should express extreme and concentrated forced-choice ratings and be willing and able to respond to open-ended questions using thematically coherent and emotionally intense language. Cross-pressured individuals, by contrast, should express more moderate and dispersed forced-choice ratings, but the character of their open-ended responses should differ depending on whether the moderation and variation in forced-choice responses is a result of avoidance/satisficing or genuine ambivalence.

If cross-pressured individuals are merely satisficing and selecting responses to finish the survey to collect their rewards at the end, their open-ended responses should reveal more disengagement than those with well-defined, reinforcing beliefs. Satisficing behavior may manifest in terms of

higher item non-response, shorter and less complex open-ended responses, and a greater reluctance to articulate positions (Krosnick et al., 2002; Saraç, 2024); (see Roberts et al., 2019, for a comprehensive review of satisficing indicators). Responses due to genuine ambivalence arising from conflicting pressures should exhibit different patterns. Compared to those who are not, cross-pressured individuals should be less likely to use moralistic or generalized language and more likely to draw on personal circumstances (Mulligan, 2013). If they are actively weighing conflicting considerations when responding, their responses may also be longer and less coherent narrative than those with reinforcing identities.

2 Assessing Social Divisions in the Pandemic

The COVID-19 pandemic provides an opportunity to test these predictions because compliance with public health recommendations on masking and vaccination created cross-pressures that were unusually visible, morally charged, and unevenly distributed across partisan lines—and because understanding public opinion about compliance became central to many efforts at managing the pandemic. To explore the nature of cross-cutting divisions in a polarized society, we interviewed 2974 respondents between 11/19/21 and 12/1/21 using an online survey with both forced-choice and open-ended questions asking about the pandemic and pandemic-related policies.[2] Included in this survey were questions about public health policies related to the pandemic, vaccination and masking, their own personal behavior, and their personal experience with the pandemic and its' effects. Because we are interested in the responses of individuals experiencing a conflict between their partisanship and compliance with pandemic policies, we focus on the 2163 respondents who either self-identify with the Democratic or Republican parties (or who lean toward either).

To assess the nature of social division in light of the new compliance-based cleavages and the increasing politicization of pandemic-related behavior and opinions (Clinton et al., 2021; Gadarian,

[2]The study was reviewed by redacted IRB and determined to be exempt. We contracted with Bovitz, Inc. to obtain a sample of approximately 3,000 respondents recruited against Census benchmarks (on age, gender, race/ethnicity, education, and income) to generate a nationally representative sample of US adults. Respondents from Bovitz's standing Forthright panel (volunteers who have signed up to take market and academic surveys) were invited to participate in our survey; the link given potential participants informed them of the projected study length and the rewards associated with their participation. On the study information page, they were invited to express their views on important issues in the country. Respondents were compensated using Bovitz's standard protocol, which includes the respondent's choice of rewards (e.g., cash payments, gift cards, or charitable donations); this was the equivalent of $2.17 for a 10 minute survey, well above the $7.25/hour federal minimum wage in 2021. No deception was employed in the study. Field dates were November 19–December 1, 2021. Data from respondents who failed either of two attention checks, completed the survey in less than five minutes, or received a Qualtrics recaptcha score under 0.5 (indicating a potential bot) were marked as invalid and are not analyzed herein. We also removed those who took longer than 30 minutes - leaving us with 2669 out of the original 3000 respondents.

Goodman and Pepinsky, 2023), we asked respondents about their opinions toward those that did and did not comply with vaccination and masking policies using thermometer score questions commonly used to measure emotional affect and affective polarization (Iyengar et al., 2019; Druckman and Levendusky, 2019). Figure 1 shows the online questions respondents were asked.

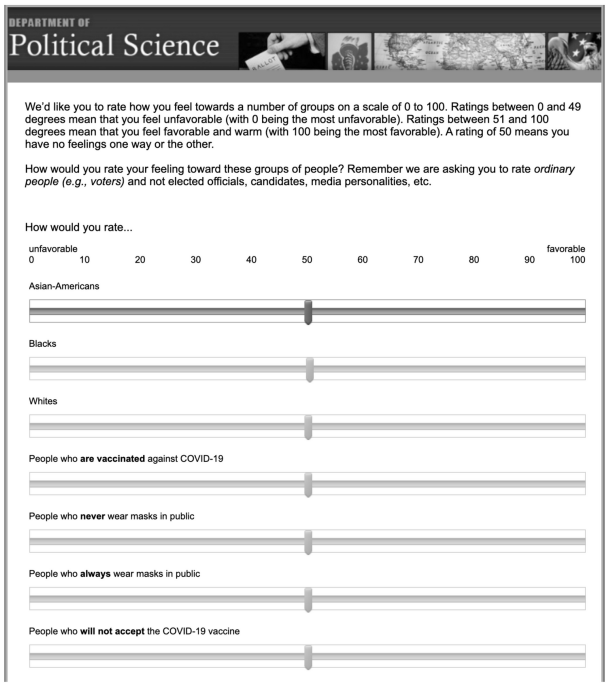

Figure 1: Thermometer Score Questions as Seen By a Respondent on a Computer. Default score were set to 50.

To determine whether an individual is cross-pressured, we use the asymmetry in elite partisan messaging and compliance with masking and vaccination behavior. Because Democrat elites, but not Republican elites, were unified in supporting masking and vaccination policies, compliance with public health recommendations reinforced the partisan identity for compliant Democrats while creating cross-pressures of varying magnitude for every other combination of partisanship and behavior. Non-compliant Democrats were cross-pressured because their personal behavior puts them at odds with the messaging and actions of their co-partisans. Republicans faced a related but distinct form of conflict: ambiguous elite messaging and greater variation in compliance among co-partisans create an environment where the "correct" position is unclear. Self-reported compliance reflect this partisan asymmetry: among Democrats, 1,058 (85%) report being vaccinated and 192

(15%) report not. Among Republicans, 567 (52%) report being vaccinated and 528 (48%) report not.[3]

It is difficult to determine which cross-pressured group is most cross-pressured, but they should all appreciably differ from the self-reinforcing pressures of partisanship and compliance that compliant Democrats experienced. As a result, the extreme and concentrated responses of compliant Democrats provide a useful baseline for how individuals with reinforcing identities should react to survey questions about those identities. Relative to of compliant Democrats, respondents who are cross-pressured should be both more moderate and also more varied.

Figure 2 plots the distribution of the *difference* in thermometer scores that each respondent assigns to the compliant and non-compliant group—for masking, vaccination, and partisanship—broken out by the respondent's own party and compliance status.[4] Compliant Democrats show a highly concentrated distribution with very large positive differences that reflect the fact that they rate compliers much more warmly than non-compliers. For nearly every other group, however, the distribution is centered near zero, indicating that many cross-pressured respondents see little difference between the two groups. Even among Republicans who mask, where the distribution tilts positive, there is substantial probability mass around zero.

But what does the concentration of probability mass near zero for cross-pressured groups mean? Does this mean that these responses have less extreme affect toward the various groups - perhaps due to their internal conflict? Or does that internal conflict result in such respondents choose to select a middling response rather than resolving how that meaning affects their views toward the groups. The pattern of varied, but middling responses could either reflect genuine ambivalence associated with competing internal pressures, or else the results of satisficing respondents choosing to avoid the question and choose answers near the default response of 50. Despite the large substantive differences in what responses due to avoidance or ambivalence means about social cleavages, both mechanisms would be expected to produce this pattern of moderate and varied forced-choice responses.

Another way to characterize the response patterns of cross-pressured groups is to examine the implied affective ordering by considering the fraction of respondents who rate the out-group either

[3]We see similar numbers when asking about masking behavior. 1,032 Democrats report wearing masks and 218 report not doing so. In contrast, 424 Republicans report wearing masks and 671 report not doing so. See Appendix A for the joint distribution.

[4]Because individuals may vary in their compliance with each behavior, compliance is defined with respect to each behavior (i.e., compliant Democrats for masking questions are Democrats who always mask, and compliant Democrats for vaccination questions are Democrats who report being vaccinated.

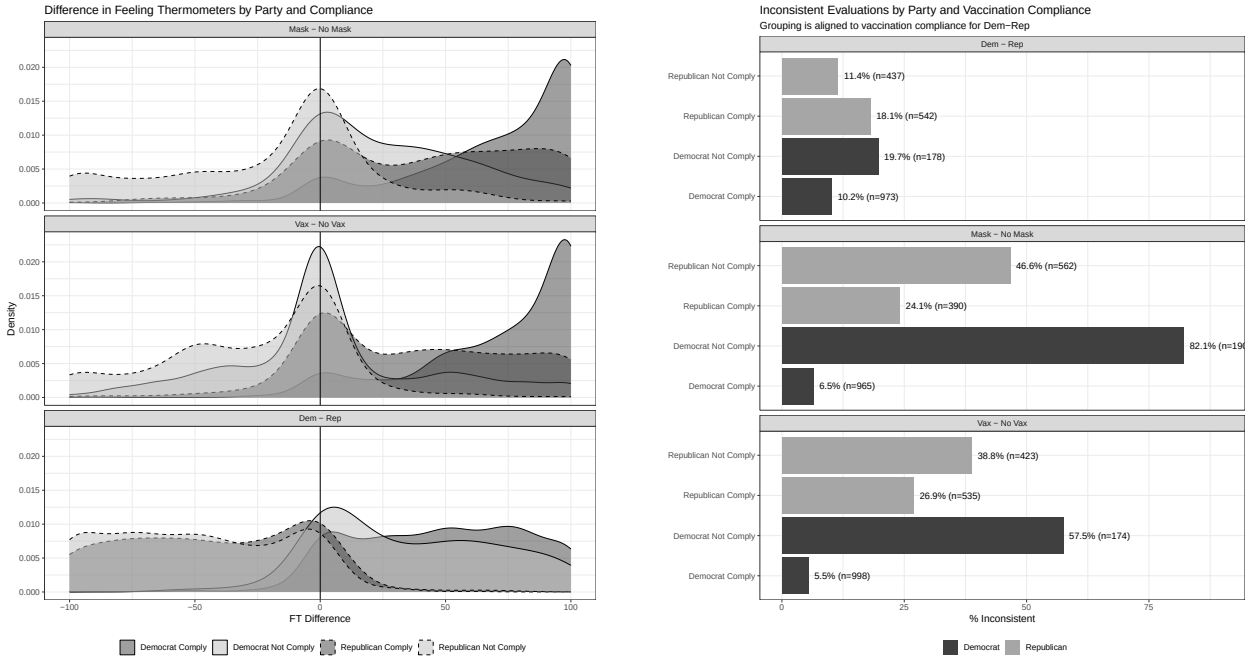

Figure 2: Left: Distribution of feeling thermometer differences by party and compliance status. Positive values indicate warmer feelings toward the compliant group (maskers, vaccinated) or toward co-partisans (Democrats for Democrats, Republicans for Republicans). Compliance is behavior-specific. Right: Percentage of misordered respondents. Inconsistency is based on whether the out-group thermometer score is rated greater than or equal to the in-group thermometer score when group status is defined by partisanship (top), masking (middle) and vaccination status (bottom). Compliance is defined by vaccination status when looking at partisan differences.

as warmly, or more warmly, than their in-group. As the right-panel of Figure 2 shows, only 10% of compliant Democrats rate Republicans as warmly or more warmly than they do fellow Democrats. While this is somewhat surprising – and perhaps reflecting a baseline level of survey response error – this percentage is similar for non-compliant Republicans but nearly twice as much for cross-pressured respondents. Nearly 20% of vaccinated Republicans and unvaccinated Democrats rate the opposing party at least as warmly as their own. But does this reflect greater ambivalence toward the out-party because the individual is cross-pressured, or does it reflect a higher level of survey response error due to avoidance?

For compliance-based comparisons, the rates of ordinal reversals are even higher. Among vaccinated Republicans, 27% rate the unvaccinated more warmly than the vaccinated and 58% of unvaccinated Democrats rate the vaccinated more warmly even though they are personally unvaccinated. In contrast, 39% of unvaccinated Republicans and 68% of unvaccinated Democrats rate those who are vaccinated at least as warmly as they rate those who are not.[5]

[5]The rate of reversals is even higher for masking: 47% of Republicans who self-report not wearing masks and 82%

Resolving the observational equivalence is substantively consequential. If respondents are choosing middling and possibly inconsistent responses because they are satisficing when answering survey questions rather than working to resolve their internal conflict, the resulting survey data is uninformative about the nature and implications of cross-pressures for social cleavages and opinion formation. But if the seemingly conflicted responses are more consistent with genuine ambivalence, the implications are quite different. If cross-pressured individuals become more ambivalent when forced to confront internal conflict between their identities, existing social divisions may not be as calcified as they might appear and new issues may disrupt the status quo through the creation of new coalitions with weaker divisions and alternative connections. Resolving whether the patterns we find are more consistent with avoidance or ambivalence, however, requires going beyond the forced-choice data to examine how respondents self-describe their positions.

3 Beyond Forced Choice: Revealing Ambivalence Through Open-Ended Responses

To resolve the indeterminacy of interpreting the responses from cross-pressured individuals, we leverage information contained in open-ended responses to questions that asked respondents about those who did and did not comply with public health recommendations. In particular, each survey respondent was asked about either masking or vaccination using the following set of questions:[6]

- What words or phrases come to mind when you think of someone **who wears (does not wear) a mask** in public places?
- What words or phrases come to mind when you think of someone who **has (not) been** vaccinated against COVID-19?

The goal is to use the resulting responses to discern whether the patterns noted above are more likely to reflect genuine ambivalence or survey satisficing behavior by respondents seeking to avoid confronting internal conflict (Roberts et al., 2014). As Table 1 summarizes, if the moderate and varied responses among the forced-choice questions we observed are the results of survey satisficing by respondents seeking to avoid their internal conflict, we should observe low-effort responses that are largely unrelated to the cross-pressured considerations they face. In contrast, if the results reflect

of unmasked Democrats report as much warmth toward masked individuals as they do toward the unmasked.

[6]To ease respondent burden, respondents were randomly asked either about masking or about vaccinations. In each case they were asked for their opinion about those who did and did not comply.

genuine ambivalence we should observe meaningful effort in the open-ended responses. The nature of that effort is unclear. One possibility is that the dominance of partisanship for creating cues and frameworks could result in partisan considerations anchoring responses even among cross-pressured individuals because it is the most accessible consideration. Even though they may express more moderate views when answering forced-choice questions, the increased cognitive load associated with responding to open-ended questions may result in them resorting to more accessible notions related to partisanship. Alternatively, the responses to the open-ended questions may instead reflect ambivalence in ways that mirror the more moderate, and varied, responses they provide on the forced-choice questions. If so, responses from the different cross-pressured groups will be more similar within groups than they are between groups because the considerations each group faces is structured by both partisanship and compliance status. Moreover, we might expect such individuals to expend more, not less, effort when expressing their conflicted opinions.

Table 1: Three Predictions About Cross-Pressured Open-Ended Responses

Prediction	**Avoidance (Satisficing)**	**Partisan Echoing**	**Genuine Ambivalence**
Engagement	High item non-response; short or meaningless answers.	Comparable effort to co-partisans.	Comparable or greater effort than compliant Democrats.
Partisan Vocabulary	No distinctive vocabulary (generic, low-effort language).	Vocabulary similar among co-partisans	Each group develops distinctive vocabulary reflecting competing considerations.
Cognitive & Emotional Content	Shallow topics; flat, disengaged affect.	Affect and topics mirror co-partisan patterns.	Broader topic engagement with regulated intensity: lower emotional intensity combined with positive reframing and moral justification.

Because open-ended responses are more demanding for respondents because they require respondents to generate, organize, and report their considerations rather than just select among provided responses, we evaluate the prevalence of non-effort satisficing using measures based on item non-response, mechanical text features, and (human-coded) subjective ratings of response

quality.[7]

To begin, Table 2 presents the percentage of responses that were judged to be we examine the percentage of"uncodeable" responses due to gibberish text, item non-response, or a failure to engage the question by at least one human coder.[8]

Table 2: Percentage of Uncodable Open-Ended Responses by Group

	Group	Toward Compliers	Toward Non-Compliers	*N*
Vaccination				
	Democrat & Vaccinated	2.1%	3.9%	490
	Democrat & Not Vaccinated	12.2%	14.3%	98
	Republican & Vaccinated	5.3%	5.0%	282
	Republican & Not Vaccinated	6.5%	6.4%	219
Masking				
	Democrat & Mask	3.3%	3.1%	483
	Democrat & No Mask	5.2%	6.4%	96
	Republican & Mask	4.4%	5.9%	204
	Republican & No Mask	4.1%	8.2%	294

Note: Compliers are vaccinated (vaccination domain) or masked (masking domain). Uncodable responses are those flagged by human coders as gibberish, blank, or otherwise failing to engage the prompt.

In general, the pattern suggests engagement rather than avoidance. Consistent with the fact that compliant Democrats possess a coherent and reinforcing set of considerations to draw from when responding, they have the lowest percentage of uncodable responses – with percentages ranging from 1.7 to 3.7 percent across the four open-ended prompts.[9] Cross-pressured groups have a higher percentage of uncodable responses — ranging from 4% among masked Republicans to 14% among unvaccinated Democrats — but the vast majority of respondents provided a codable response.[10]

Analyzing several additional measures support this conclusion.[11] Mechanical measures based on the number of whitespace-delimited tokens (*Word Count*), and the ratio of unique tokens to

[7]These responses were collected in December 2021, before the widespread availability of large language models that might have allowed respondents to outsource the cognitive work of composing a response (Westwood, 2025).

[8]Three humans evaluated each response and if any of them thought the response was "uncodeable" the response was flagged as such.

[9]The extremely high level of response is likely do to the fact that we prompted respondents for an answer (which they presumably were inclined to provide to ensure payment for completion) and also the possible effects of taking the survey during the pandemic when the opportunity costs of taking surveys was presumably lower due to the restrictions in place.

[10]Predicting the probability of an uncodable response shows that those with more extreme expressions on the feeling thermometer question are less likely to give an uncodable response (see appendix Table E.1), but the substantive effects of a maximal difference in extremity is only a slight increase in the probability of an uncodable response.

[11]We also looked at the overall time respondents took to complete the survey and found comparable completion times across the groups (median range: 10.5–11.6 minutes).

total tokens (*Lexical Diversity*) are useful (Wasow, 2026), but we also categorize responses using a five-point scale (*Response Quality*) ranging from: (1) "No effort. Blank, gibberish, or irrelevant, " to (5) "Thoughtful, developed, and expressive response."[12]

To decompose the variation in response effort, we regress each effort measure on indicators for each cross-pressure group (with compliant Democrats as the baseline), the extremity of the respondent's feeling thermometer rating, and demographic controls.[13] In general, there are few systematic differences in effort between compliant Democrats and cross-pressured respondents: nearly every coefficient is statistically indistinguishable from zero, and the few that are distinguishable suggest that cross-pressured individuals write *more* words, not fewer (Appendix Tables L.1–L.3). These results do not establish that cross-pressured individuals are actively engaging with competing considerations, but they do suggest that the responses of cross-pressured individuals differ from those of compliant Democrats in ways that suggest that they are dealing with their internal conflict by exerting minimal effort.

3.1 Do Cross-Pressured Respondents Simply Echo Their Co-Partisans?

Cross-pressured individuals do not respond to internal conflict by avoiding questions related to that conflict, but examining the characteristics of the answers they provide is essential for evaluating whether their responses reflect genuine ambivalence. One possibility is that even though they are cross-pressured, they may resolve that by relying on readily accessible partisan cues rather than grappling with conflicting considerations when dealing with the higher cognitive load of answering open-ended responses. Put differently, even though they answer the forced-choice items differently, perhaps they resort to dominant partisan cues when asked to answer more demanding open-ended

[12] All coding was performed by undergraduate students who were provided with a spreadsheet of responses and a coding rubric and told to assess the open-ended according to the rubric and their interpretation of the rubric. They were explicitly told not to use an LLM to assist them in their coding. See Appendix D for the full coding rubric, but the specific prompt we provided was: "On a scale from 1 (very low effort) to 5 (very high effort), how much effort did the respondent seem to put into the following answer? Consider how specific, thoughtful, and elaborated the response is. 1: No effort. Blank, gibberish, or irrelevant. 2: Very minimal response; short and vague. 3: Brief but interpretable response. 4: Sentence or multi-word phrase with some context or explanation. 5: Thoughtful, developed, and expressive response." To assist in this evaluation, we also used the same rubric and GPT-4o-mini to generate an analogous LLM-based coding (Krippendorff, 2018; Chang et al., 2023; Mellon et al., 2024; Heseltine and Clemm von Hohenberg, 2024). Appendix N reports details of our runs using GPT-4o-mini as well as the exact code we used.

[13] Controls include age, education, gender, race/ethnicity, COVID-19 exposure, and device type. For coded response quality, we additionally include coder fixed effects to account for systematic differences across evaluators. See Appendix L for full results.

questions (Zaller, 1992). If so, we should see similar considerations being expressed by partisans regardless of their compliance status.

To begin, Figure 3 compares the highest and lowest log-odds ratio for word usage for each group relative to those used by compliant Democrats when describing vaccination-related (top) and masking-related (bottom) groups.[14] Although this analysis highlights differences in word use relative to compliant Democrats, it is useful for assessing the extent to which those differences appear substantively meaningful and interpretable.

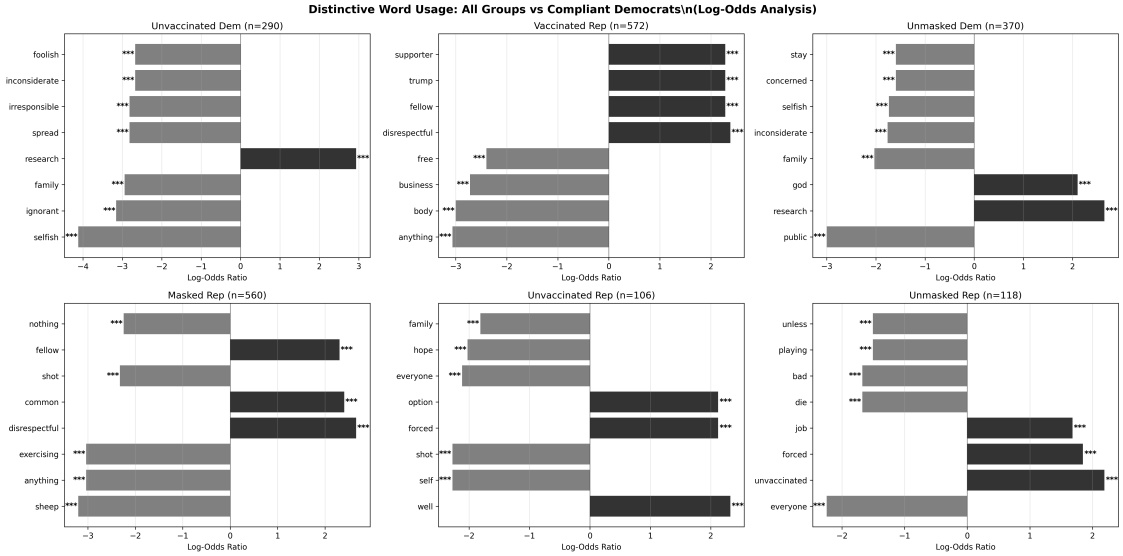

Figure 3: Distinctive Language Usage by All Groups vs Compliant Democrats. Horizontal bars show log-odds ratios for the most distinctive words used by each group compared to compliant Democrats. Analysis uses Porter stemming to group word variations while displaying complete words for readability. Dark bars indicate overused words (positive log-odds); gray bars indicate underused words (negative log-odds). Only statistically significant differences ($p<0.05$) are shown.

It is immediately clear that the language of cross-pressured groups differs from that used by compliant Democrats in different, but interpretable, ways. Relative to vaccinated Democrats, unvaccinated Democrats are more likely to use words related to skeptical inquiry (e.g., "research" and "god") and less likely to use terms of moral condemnation like "selfish," "ignorant," and "irresponsible." This is consistent with ambivalence arising from an interpretation of vaccine hesitancy as a matter of uncertainty or personal freedom rather than moral failure. As one respondent explained: *"it is my body my choice.....I think they need to leave well enough alone and let the people decide*

[14] As in our other analyses, we combined responses toward both compliers and non-compliers when analyzing the opinions each respondent expresses toward vaccination and masking.

what is right for their own body."[15]

Vaccinated Republicans are relatively more likely to use community-oriented language (e.g., "disrespectful," "fellow," "supporter," and "trump") and avoid terms like "anything," "body," and "free." Such respondents appear to ground their compliance within conservative community values, for as one respondent noted when describing those who have been vaccinated: *"They are doing what they've been told to do and protecting themselves, loved ones, and society in general."*[16]

Unvaccinated Republicans, in contrast, are more likely to emphasize government coercion (e.g., "forced," "option") and less likely to use terms like "everyone" and "hope." Non-compliance is often framed as principled resistance; as one respondent wrote: *"Everyone has a choice whether to be vaccinated or not. No one should be forced into doing so. It is wrong and unconstitutional. Stop mandating vaccines."*[17]

Parallel patterns emerge when for masking: unmasked Democrats express the vocabulary of skeptical inquiry observed among unvaccinated Democrats, masked Republicans emphasize community considerations, and the responses of unmasked Republicans often frame non-compliance in terms of principled resistance.[18] The fact that the considerations being expressed vary by partisan-compliance groups while also exhibiting a broad similarity across masking and vaccine behaviors suggests that these language patterns are associated with similar considerations across compliance contexts rather than issue-specific reactions to masking and vaccination.

While suggestive of the different considerations being used by the partisan-compliance groups, it cannot assess the degree to which the language being used is similar or dissimilar because it is purposefully designed to identify the words that are more or less frequently used relative to compliant Democrats. To characterize the variation and similarity in open-ended responses more systematically, we compute Term Frequency-Inverse Document Frequency (TF-IDF) vectors for each individual's responses, calculate the pairwise cosine similarity between all response pairs, and estimate a dyad-level regression that decomposes how much of the text similarity between respondent pairs is attributable to shared partisanship, shared compliance status, and their interaction.[19] If parti-

[15] `ResponseId: R_1f7Rb1TRi7JWgjb`. Complete response available in replication dataset.
[16] `ResponseId: R_27Uan40Yrdlmh6i`. Complete response available in replication dataset.
[17] `ResponseId: R_3jTp6hsV6o43Daf`. Complete response available in replication dataset.
[18] See the bottom row of Figure 3 for details.
[19] Standard errors are two-way clustered on respondent i and respondent j using Cameron, Gelbach and Miller (2011). A nonparametric permutation test that shuffles party labels within compliance groups yields consistent results (Appendix Figure G.1).

san considerations structure the open-ended responses, sharing a party should predict greater text similarity; if compliance-related considerations matter more, sharing a compliance status should be the stronger predictor.

Table 3 presents the results. In the baseline specification (columns 1 and 3), both shared partisanship and shared compliance status are positively and significantly associated with pairwise text similarity. For vaccination, sharing a party ($\beta = 0.012$, $p < 0.001$) contributes roughly twice as much to text similarity as sharing a compliance status ($\beta = 0.006$, $p < 0.001$), whereas for masking, shared compliance ($\beta = 0.015$, $p < 0.001$) slightly exceeds the party effect ($\beta = 0.012$, $p < 0.001$). The interaction term is small and insignificant in both domains, indicating that the two identity dimensions contribute roughly additively to language similarity.

Table 3: Dyad-Level Regression: Decomposing Pairwise Text Similarity

	Dependent Variable: Pairwise Cosine Similarity			
	Vaccination		Masking	
	(1)	(2)	(3)	(4)
Same Party	0.012***	0.007**	0.012***	0.008***
	(0.002)	(0.002)	(0.001)	(0.001)
Same Compliance	0.006***	-0.002	0.015***	0.010***
	(0.001)	(0.002)	(0.002)	(0.002)
Same Party × Same Compliance	0.003	0.004*	-0.002	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
FT Distance		-0.020***		-0.016***
		(0.003)		(0.002)
Constant	0.023***	0.042***	0.019***	0.033***
	(0.001)	(0.003)	(0.001)	(0.002)
R^2	0.008	0.016	0.013	0.020
Observations	188,805	174,936	218,791	203,841
Clusters (resp. i)	614	591	661	638
Clusters (resp. j)	614	591	661	638

Note: *p<0.05; **p<0.01; ***p<0.001. Two-way clustered standard errors (Cameron, Gelbach & Miller, 2011) on respondent i and respondent j in parentheses.

Same Party = 1 if both respondents share partisan affiliation; Same Compliance = 1 if both share compliance status. FT Distance = $|\text{FT}_{\text{diff},i} - \text{FT}_{\text{diff},j}|$ where FT_{diff} is the respondent's feeling thermometer differential toward target groups (vax−novax or mask−nomask, scaled −1 to 1).

Cosine similarity computed on TF-IDF vectors (min_df=3, max_df=0.8, English stop words removed) at the individual response level.

Including the absolute difference in feeling thermometer differentials as a control (columns 2 and 4) reassuringly reveals that affective distance is strongly and negatively associated with text similarity. As we would expect if the open-ended responses of each respondent was meaningfully related to forced-choice thermometer scores, respondents with more dissimilar thermometer scores

also have more dissimilar open-ended responses ($\beta = -0.020$, $p < 0.001$ for vaccination; $\beta = -0.016$, $p < 0.001$ for masking). Controlling for the difference in forced-choice affect distance fully absorbs the compliance effect related to vaccination (which drops from 0.006 to -0.002, $p = 0.32$), suggesting that compliance-driven similarity in vaccination language is largely mediated by affective orientation. For masking, the compliance effect remains significant even controlling for affect, suggesting that shared compliance status captures language similarity beyond what thermometer scores explain.

Stepping back, the dyad regression confirms that although co-partisans share some common vocabulary, so too do those sharing a similar compliance status. It is not the case that respondents resort to partisan cues when dealing with internal cross-pressures. Instead, the dominant pattern is that both features matter, resulting in group-specific language that reflects the particular tensions each group faces and which is revealed by the log-odds analysis.

3.2 Content Differences in Cross-Pressured Responses

Having found that cross-pressured individuals engage rather than avoid open-ended responses and they respond using language that seemingly varies by both partisanship and compliance status, we now analyze the open-ended responses in more detail to better discern the nature of the internal pressures and considerations that are associated with the more ambivalent and varied affect being expressed by cross-pressured individuals. We begin by examining the considerations cross-pressured individuals raise before turning to the emotional and moral content of their responses.

LDA-based measures of topic diversity show no significant differences in the number of topics that cross-pressured individuals raise (Appendix Table H.3), but there are important differences in the types of considerations being invoked. Using LLM-coded content categories derived from an initial round of human coding, we classify each response according to whether it contains evaluative language (mentions of helpfulness, incompetence, or ineffectiveness) or appeals to personal choice (individual autonomy, freedom, or personal responsibility).[20]

Figure 4 shows that compliant Democrats are most likely to use evaluative language (42% in the vaccination domain and 46% in masking) and non-compliant Republicans are most likely to

[20]See Appendix H for LDA methodology and Appendix Table I.1 for the full breakdown of content themes by group, including chi-square tests. The content coding scheme was derived from themes identified in an initial round of human open coding.

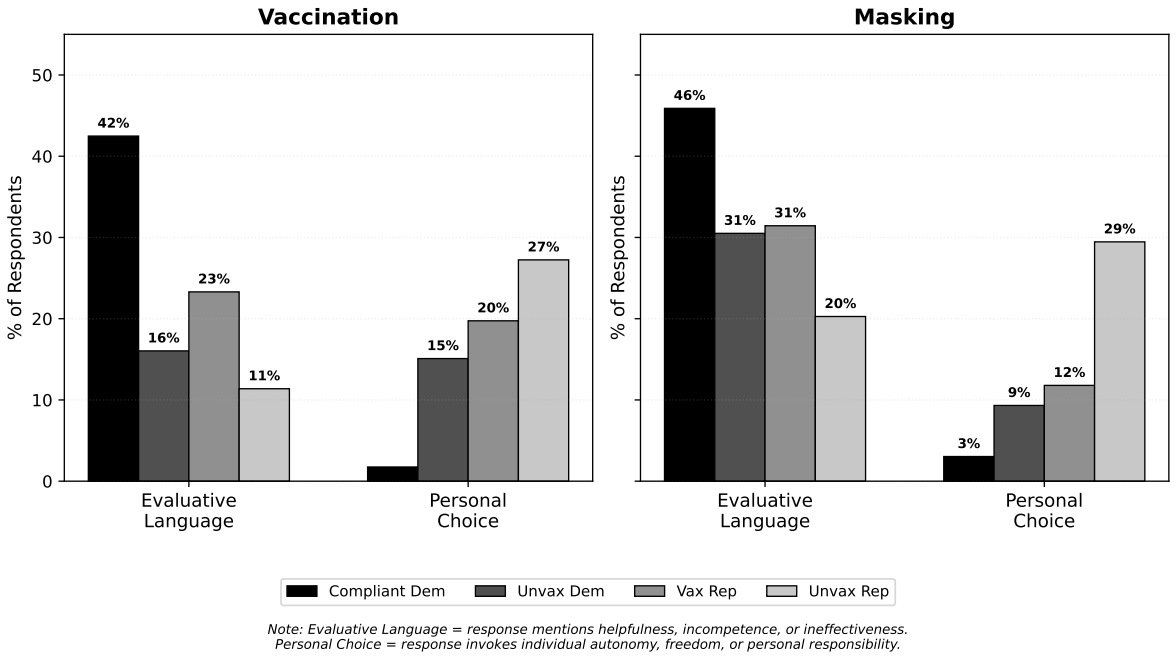

Figure 4: Content Theme Prevalence by Cross-Pressure Group. Evaluative Language = response mentions helpfulness, incompetence, or ineffectiveness. Personal Choice = response invokes individual autonomy, freedom, or personal responsibility. All differences across groups are statistically significant (χ^2 tests, $p < 0.001$).

invoke considerations related to personal choice (27% for vaccination, 29% for masking). Consistent with the competing considerations that their conflicting identities imply, cross-pressured groups fall between these two extremes. The content differences are highly significant ($\chi^2 > 128$, $p < 0.001$ for both themes in both domains) and suggest that rather than using a wider range of the same topics, cross-pressured and non-compliant respondents seem to invoke different frameworks when evaluating compliance.

To explore how these considerations are expressed in terms of emotional and moral content, humans and LLMs coded the psychological content of open-ended responses across three theoretically motivated dimensions. Emotional intensity was measured using a 5-point scale (1=very low to 5=very high) based on the emotional engagement of each response.[21] Valence was coded using a 3-point scale (-1=negative, 0=neutral, 1=positive) with negative valence including words like "stupid," "selfish," or "dangerous," positive valence encompassing terms like "smart," "cautious," or "caring," and neutral responses characterized by factual statements. Moral judgment employed a parallel 3-point scale (-1=negative moral evaluation, 0=no moral language, 1=positive moral

[21]See Appendix D for the coding scheme and Appendix M for additional details on computation and analysis.

evaluation) with negative judgments including language of wrongness or blame (e.g., "selfish," "irresponsible") and positive judgments invoking virtue or moral praise (e.g., "doing the right thing," "responsible").

To quantify these differences, we compare the average assessment for responses related to compliers and non-compliers for each cross-pressured group against compliant Democrats using both human only and also both human and LLM codings. In so doing, we control for differences in the coding been doing using coder fixed effects and also differences related to respondent differences correlated with demographics controlling and also the extremity of feeling thermometer scores.[22]

Figure 5 presents the average difference for each group relative to compliant Democrats conditional on statistical controls.[23]

As the coefficient plots reveal, the responses of cross-pressured individuals possess lower emotional intensity, greater positive language (i.e., valence) and sometimes slightly more positive moral judgment than the responses of compliant Democrats on open-ended items asking respondents to describe individuals who do and do not comply with public health recommendations. This is true regardless of whether we consider vaccination (top row) or masking (bottom row).

Given that these differences reflect the average orientations being expressed toward both compliers and non-compliers by each respondent, the differences reflect the overall orientation toward cleavages on each policy. As a result, the differences we find regarding lower emotional intensity and more positive language among the responses of the cross-pressured relative to compliant Democrats reflects differences that are largely attributable to the emotional and negative reactions of compliant Democrats toward non-compliers. (All three cross-pressured groups also show increased positive moral language usage relative to compliant Democrats, but the differences are not always statistically distinguishable from zero.)

In general, the patterns we detect suggest not only that the expressions of cross-pressured individuals reflect intentional ambivalence, but also that the moderate and dispersed affect measured using forced-choice responses arguably reflect differences in the type and intensity of considerations associated with the issues and social cleavages involved. To be clear, although the patterns we

[22]Controlling for thermometer extremity determines whether language differences among the cross-pressured reflect more than just the level of affect (as measured by thermometer scores).

[23]Full regression results for vaccination appear in Appendix Table M.1. Similar patterns emerge for masking behavior in Appendix Table M.2. We also decompose group differences into associations related to partisanship and compliance in Tables M.3 and M.4.

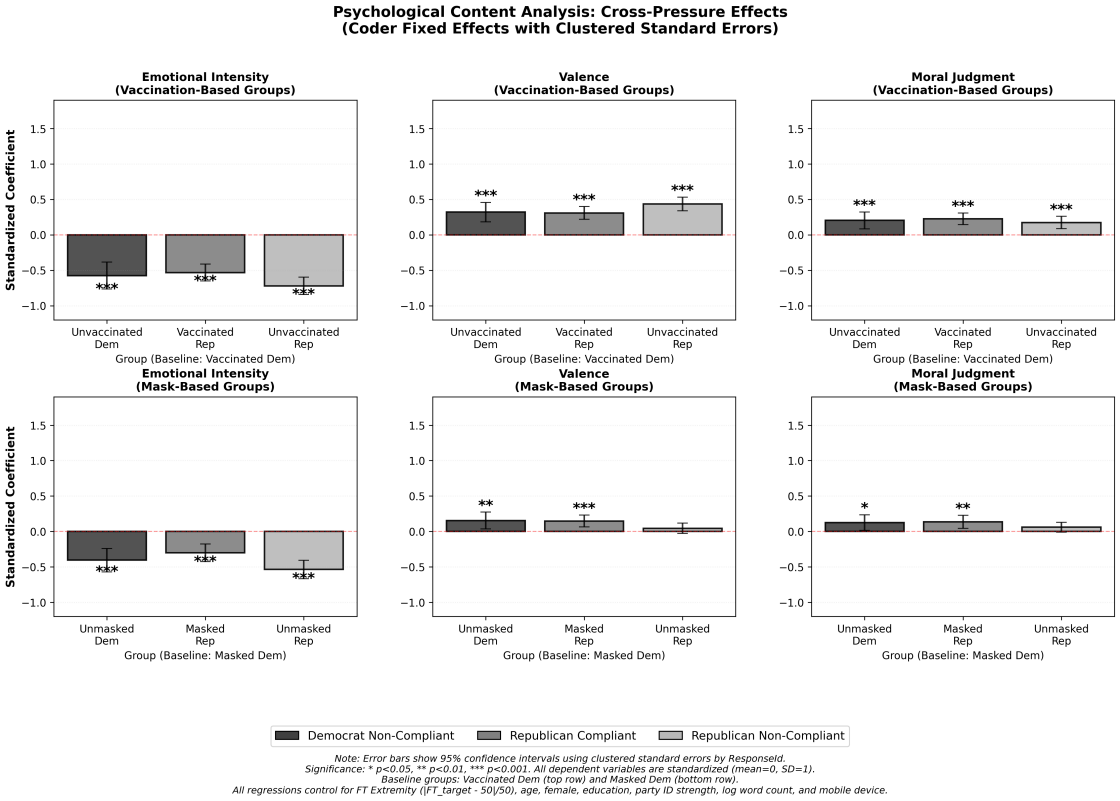

Figure 5: Comparison of Psychological Content Effects Across Compliance Behaviors. Bars show standardized coefficients from four-group mean differences specification with 95% confidence intervals based on clustered standard errors (by respondent ID). Baseline: Compliant Democrats (coefficients = 0). Top row: vaccination-based groups; bottom row: masking-based groups. Negative coefficients indicate lower psychological content scores relative to compliant Democrats; positive coefficients indicate higher scores. All measures use coder fixed effects pooling human and LLM ratings. Each respondent contributes two responses (compliant + non-compliant target). All regressions include FT Extremity ($= |\text{FT}_{\text{target}} - 50|/50$) as a control.

document are patterns of observable language differences, we are unable to identify the precise psychological mechanisms or response strategies involved responsible for those differences. Even so, the approach of using open-ended questions to interpret forced-choice responses in circumstances when we may be concerned about the inability to discern ambivalence from avoidance offers a significant benefit given that the substantive implications for interpreting public opinion depends greatly on whether the responses reflect respondent cognition or satisficing.

4 Discussion and Conclusion

The COVID-19 pandemic created sizable, salient, and important conflicts between partisan identities and personal behavioral choices related to compliance with public health measures. Understanding how cross-pressured individuals respond to survey questions – and whether their responses reflect efforts at avoiding their internal conflict or resolving them – is essential for interpreting public opinion on issues that create new, possibly cross-cutting divisions. This can be a challenge because the response patterns we would expect from ambivalence and avoidance would be similar, consisting of more moderate, and more varied responses relative to those with stable, reinforcing identities. Distinguishing between engagement and avoidance is more than a matter of survey interpretation, it speaks directly to our ability to identify and characterize the depth and nature of social division using survey methods.

By combining forced-choice and open-ended responses and analyzing the relation both within and between respondents with varying levels of internal conflict based on their partisan identity and decision to comply with public health recommendations, we argue that the opinions that cross-pressured individuals express about those who do and do not comply with public health recommendations are more consistent with ambivalence than avoidance.

Although the level of non-response is higher among cross-pressured individuals on the open-ended surveys, it is only modestly so and an overwhelming majority of respondents respond. Moreover, the length and quality of the responses are not dissimilar – suggesting that being cross-pressured on these issues is not causing respondents to either skip the more cognitively demanding open-ended questions or to respond with fewer or less meaningful words.

Analyzing the content of the open-ended expressions using quantitative measures of text sim-

ilarity and also codings performed by humans and/or LLMs suggests that there is meaningful variation in the responses of the different groups. Although there are similarities across both partisan and compliance dimensions, cross-pressured individuals' responses generally shows reduced emotional intensity alongside more positive language and moral framing relative to compliance Democrats. These patterns cannot be interpreted as necessarily revealing the psychological processes of cross-pressured individuals, but it does reveal that cross-pressured individuals express themselves differently in ways that can be interpreted in substantively meaningful ways. The fact that interpretable differences are detectable suggests that the responses reflect more than respondents' efforts to either avoid answering the question entirely or to enter enough text so as to be passed onto the next survey question. Instead, the patterns we find suggest that the moderate, but varied, expressions of affect we detect in the forced-choice questions reflect genuine ambivalence that results from cross-pressured individuals relying on different sets of considerations when responding to cognitively demanding open-ended questions that cause them to confront their internal conflict.

Substantively, our findings have implications for understanding social divisions in an era of cross-cutting cleavages. Our results are difficult to reconcile with conclusions that portray politically moderate or seemingly inconsistent views as the result of confusion, disengagement, or avoidance. The profile of cross-pressured individuals' responses—substantive, with lower emotional intensity, and including moral language—is consistent with engagement among multiple competing values, including personal autonomy, health and safety, and social responsibility, that reflect legitimate tensions in democratic societies. The moderate responses are more consistent with genuine complexity than with political ignorance or apathy.

Methodologically, not only does the use of open-ended questions allow researcher to discern between responses generate by ambivalence and avoidance, but the content of open-ended questions are also substantively important and interesting in their own right for characterizing the nature of societal divisions. Rather than assuming survey responses reflect genuine opinions, it is now more possible than ever to combine forced-choice and open-ended responses to systematically test whether response patterns both within and between respondents are substantively meaningful. Our analysis illustrates how large language models can be used alongside human coders as transparent interpretive instruments of respondent motivations rather than black-box prediction tools. Although we focus on cleavages related to the pandemic, debates about the meaning of

politically moderate opinions, or in instances where individuals express views seemingly contrary to their self-interest represent areas where combining forced-choice and open-ended responses can help determine whether responses reflect ambivalence, avoidance, confusion, or something else.

Although leveraging open-ended survey responses provides a deeper set of considerations and points to a range of psychological pressures than are possible when relying on forced-choice items along, our work offers limited insights into the specific mechanisms at work. Even so, our approach provides a promising foundation and diagnostic template for future research, particularly when combined with experiments that manipulate cross-pressure exposure and jointly analyze forced-choice and open-ended responses to more directly assess how and why such pressures are, or are not, resolved. The combination of forced-choice and open-ended responses provides a scalable approach for studying the authenticity of attitudes in a polity with cross-cutting cleavages.

References

Allamong, Maxwell B., Benjamin Beutel, Jongwoo Jeong and Paul M. Kellstedt. forthcoming. "The Declarations of Independents: Open-Ended Survey Responses and the Nature of Non-Identification." *American Journal of Political Science* . Forthcoming.

Ansolabehere, Stephen, Jonathan Rodden and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(2):215–232.

Berinsky, Adam J. 2004. *Silent Voices: Public Opinion and Political Participation in America*. Princeton, NJ: Princeton University Press.

Berinsky, Adam J., Michele F. Margolis and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3):739–753.
URL: *https://doi.org/10.1111/ajps.12081*

Bernhard-Harrer, Jana and Katharina Pfaff. 2025. "Question Form Matters: Examining Trust in Government Through Open and Closed Survey Items." *Journal of Survey Statistics and Methodology* 13(4):370–392.

Bishop, George F., Alfred J. Tuchfarber and Robert W. Oldendick. 1986. "Opinions on Fictitious Issues: The Pressure to Answer Survey Questions." *Public Opinion Quarterly* 50(2):240–250.

Bishop, George F, Robert W Oldendick and Alfred J Tuchfarber. 1984. "Pseudo-opinions on public affairs." *Public Opinion Quarterly* 48(1):1–22.

Box-Steffensmeier, Janet M. and Laura Moses. 2021. "Meaningful Messaging: Sentiment in Elite Social Media Communication with the Public on the COVID-19 Pandemic." *Science Advances* 7(29):eabg2898.

Brader, Ted and George E. Marcus. 2013. Emotion and Political Psychology. In *The Oxford Handbook of Political Psychology*, ed. Leonie Huddy, David O. Sears and Jack S. Levy. 2nd ed. New York: Oxford University Press pp. 165–204.

Broockman, David E. and Benjamin E. Lauderdale. 2025. ""Moderates"." *American Political Science Review* pp. 1–10.

Cameron, A. Colin, Jonah B. Gelbach and Douglas L. Miller. 2011. "Robust Inference with Multiway Clustering." *Journal of Business & Economic Statistics* 29(2):238–249.

Chang, Kent K, Mackenzie Cramer, Sandeep Soni and David Bamman. 2023. "Speak, memory: An archaeology of books known to chatgpt/gpt-4." *arXiv preprint arXiv:2305.00118* .

Clinton, Joshua, Jon Cohen, John Lapinski and Marc Trussler. 2021. "Partisan pandemic: How partisanship and public health concerns affect individuals' social mobility during COVID-19." *Science Advances* 7(2):eabd7204.
URL: *https://doi.org/10.1126/sciadv.abd7204*

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, ed. David E. Apter. Free Press pp. 206–261.

Converse, Philip E. 1970. "Attitudes and Non-Attitudes: Continuation of a Dialogue." *The Quantitative Analysis of Social Problems* pp. 168–189.

Druckman, James N. and Matthew S. Levendusky. 2019. "What Do We Measure When We Measure Affective Polarization?" *Public Opinion Quarterly* 83(1):114–122.

Festinger, Leon. 1957. "Social comparison theory." *Selective Exposure Theory* 16(401):3.

Fowler, Anthony, Seth J. Hill, Jeffrey B. Lewis, Chris Tausanovitch, Lynn Vavreck and Christopher Warshaw. 2023. "Moderates." *American Political Science Review* 117(2):643–660.

Gadarian, Shana Kushner, Sara Wallace Goodman and Thomas B. Pepinsky. 2023. *Pandemic Politics: The Deadly Toll of Partisanship in the Age of COVID*. Princeton: Princeton University Press.

Green, Donald, Bradley Palmquist and Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven: Yale University Press.

Green, Jon, Jared Edgerton, Daniel Naftel, Kelsey Shoub and Skyler J. Cranmer. 2020. "Elusive Consensus: Polarization in Elite Communication on the COVID-19 Pandemic." *Science Advances* 6(28).

Grice, H. Paul. 1975. Logic and Conversation. In *Syntax and Semantics, Volume 3: Speech Acts*, ed. Peter Cole and Jerry L. Morgan. New York: Academic Press pp. 41–58.

Heseltine, Michael and Bernhard Clemm von Hohenberg. 2024. "Large Language Models as a Substitute for Human Experts in Annotating Political Text." *Research & Politics* 11(1).

Hillygus, D. Sunshine and Todd G. Shields. 2008. *The Persuadable Voter: Wedge Issues in Presidential Campaigns*. Princeton, NJ: Princeton University Press.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22:129–146.
URL: *https://doi.org/10.1146/annurev-polisci-051117-073034*

Kam, Cindy D. and Jennifer M. Ramos. 2008. "Joining and Leaving the Rally: Understanding the Surge and Decline in Presidential Approval Following 9/11." *Public Opinion Quarterly* 72(4):619–650.

Keele, Luke and Jennifer Wolak. 2008. "Contextual sources of ambivalence." *Political psychology* 29(5):653–673.

Krippendorff, Klaus. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measure in Surveys." *Applied Cognitive Psychology* 5:213–236.

Krosnick, Jon A., Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp and Michael Conaway. 2002. "The Impact of "No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66(3):371–403.
URL: *https://doi.org/10.1086/341394*

Lavine, Howard. 2001. "The electoral consequences of ambivalence toward presidential candidates." *American Journal of Political Science* pp. 915–929.

Lavine, Howard G., Christopher D. Johnston and Marco R. Steenbergen. 2012. *The Ambivalent Partisan: How Critical Loyalty Promotes Democracy*. New York: Oxford University Press.

Lenz, Gabriel S. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. Chicago: University of Chicago Press.

Lipset, Seymour Martin and Stein Rokkan. 1967. *Party Systems and Voter Alignments: Cross-National Perspectives*. New York: The Free Press.

Macedo, Stephen and Frances Lee. 2025. *In COVID's Wake: How Our Politics Failed Us*. Princeton University Press.

Mellon, Jonathan, Jack Bailey, Ralph Scott, James Breckwoldt, Marta Miori and Phillip Schmedeman. 2024. "Do AIs Know What the Most Important Issue Is? Using Language Models to Code Open-Text Social Survey Responses at Scale." *Research & Politics* 11(1):1–7.

Mulligan, Kenneth. 2013. "Variability or Moderation? The Effects of Ambivalence on Political Opinions." *Political Behavior* 35(3):539–565.

Peer, Eyal, David Rothschild, Andrew Gordon, Zachary Evernden and Emily Damer. 2022. "Data quality of platforms and panels for online behavioral research." *Behavior Research Methods* 54(4):1643–1662. Epub 2021 Sep 29. Erratum in: Behav Res Methods. 2022 Oct;54(5):2618–2620. doi: 10.3758/s13428-022-01909-1.
URL: *https://doi.org/10.3758/s13428-021-01694-3*

Powell, G. Bingham. 1976. "Political Cleavage Structure, Cross-Pressure Processes, and Partisanship: An Empirical Test of the Theory." *American Journal of Political Science* 20(1):1–23.

Roberts, Caroline, Emily Gilbert, Nick Allum and Leila Eisner. 2019. "Research Synthesis: Satisficing in Surveys: A Systematic Review of the Literature." *Public Opinion Quarterly* 83(3):598–626.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Saraç, Meryem. 2024. "Does respondent motivation affect item-nonresponse for split-ballot designed survey data? Comparative evidence from the European Social Survey." *Quality & Quantity* 58:3791–3809.
URL: *https://doi.org/10.1007/s11135-023-01828-z*

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Academic Press.

Sherif, Muzafer, O. J. Harvey, B. Jack White, William R. Hood and Carolyn W. Sherif. 1961. *Intergroup Conflict and Cooperation: The Robbers Cave Experiment*. Norman: University Book Exchange.

Stagnaro, Michael N., James Druckman, Adam J. Berinsky, Antonio A. Arechar, Robb Willer and David G. Rand. 2025. "Representativeness and Response Validity Across Nine Opt-in Online Samples.". PsyArXiv.
URL: *https://doi.org/10.31234/osf.io/h9j2d*$_v$2

Thompson, Megan M and Mark P Zanna. 1995. "The conflicted individual: Personality-based and domain specific antecedents of ambivalent social attitudes." *Journal of personality* 63(2):259–288.

Wasow, Omar. 2026. "Text as Behavior." *Political Analysis* p. 1–15.

Westwood, Sean J. 2025. "The Potential Existential Threat of Large Language Models to Online Survey Research." *Proceedings of the National Academy of Sciences* 122(47):e2518075122.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Ziegler, Jeffrey. 2022. "A Text-As-Data Approach for Using Open-Ended Responses as Manipulation Checks." *Political Analysis* 30(2):289–297.
URL: *https://doi.org/10.1017/pan.2021.2*

Appendix Table of Contents

A Survey Data: Compliance & Partisanship

This section contains additional descriptive information on the online survey that we conducted.

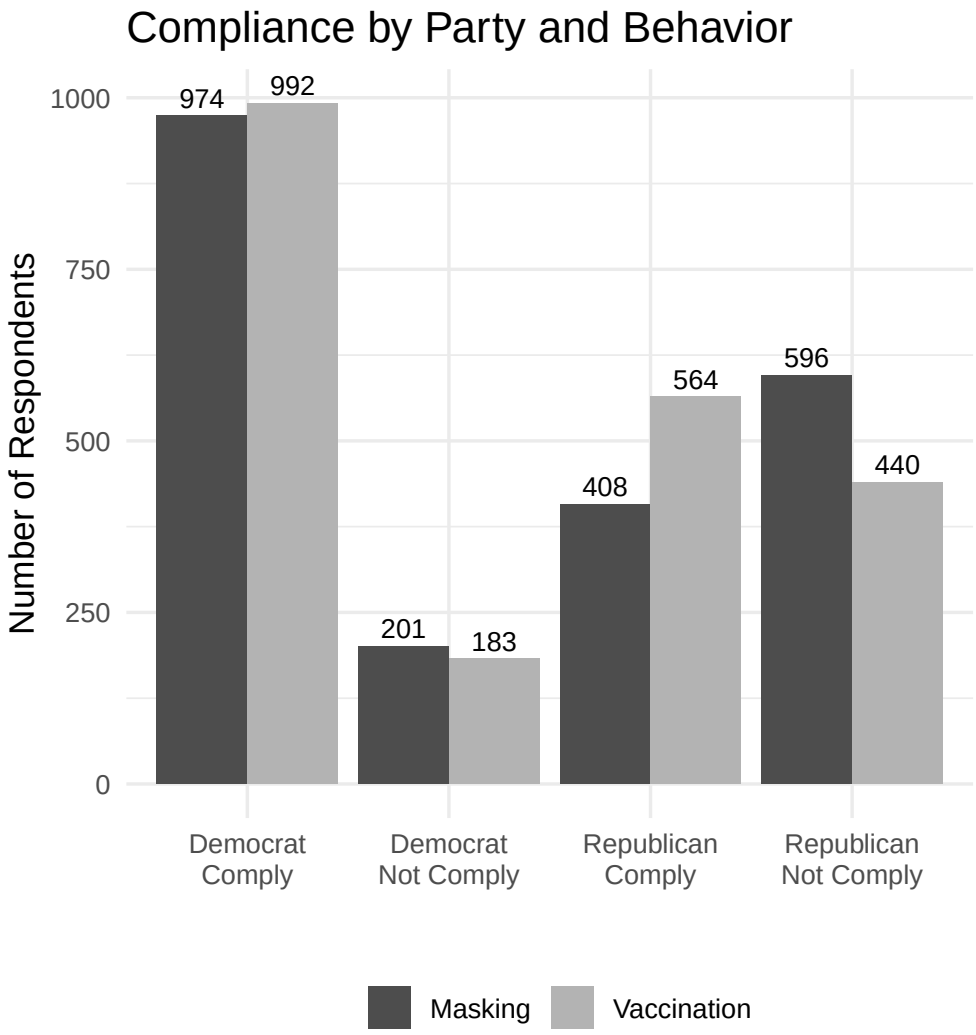

Figure A.1: Respondents By Partisanship and Compliance

Table A.1: Masking and Vaccination by Partisanship (Counts and Row Percentages)

Group	Masking	Dem & Vax	Dem & Not Vax	Rep & Vax	Rep & Not Vax
Democrat	Mask	840 (72.5%)	127 (11.0%)	0 (0%)	0 (0%)
	No Mask	144 (12.4%)	48 (4.1%)	0 (0%)	0 (0%)
Republican	Mask	0 (0%)	0 (0%)	277 (29.0%)	120 (12.5%)
	No Mask	0 (0%)	0 (0%)	264 (27.6%)	296 (30.9%)

B Thermometer Score Description & Analysis

B.1 Predicted vs. Actual Thermometer Scores

Figure B.1 compares the average thermometer score predicted by three human coders—based solely on reading each open-ended response—to the respondent's actual thermometer score for vaccination groups. Although human coders predict less extreme scores than actually observed, the correlation between predicted and actual scores is 0.68 for evaluations of the vaccinated and 0.51 for the unvaccinated, confirming that the open-ended responses contain opinion-relevant content that is meaningfully connected to forced-choice ratings.

Figure B.1: Comparison of Average Predicted Score and Actual Thermometer Score: Vaccination Groups. The relationship is the result of 3 human coders coding 1000 open-ended responses and comparing the average thermometer score that was predicted based on the open-end response to the actual response that was given.

B.2 Masking-Based Thermometer Analysis

Table B.1 presents the analogous analysis of Table **??** in the text. As with the case with vaccination, un-masked Democrats and masked Republicans report more moderate feelings than compliant

Democrats and non-masker Republicans.

Table B.1: Four-Group Effects on Thermometer Ratings: Masking-Based Groups

Outcome	Dem Non-Masker	Rep Masker	Rep Non-Masker
Baseline: Democrat Masker			
Feelings toward Maskers	-17.321^{***}	-12.493^{***}	-37.263^{***}
	(2.260)	(1.664)	(1.841)
		$R^2 = 0.329$	
Feelings toward Non-Maskers	28.995^{***}	22.471^{***}	54.426^{***}
	(2.681)	(2.297)	(1.780)
		$R^2 = 0.444$	

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Clustered standard errors by respondent ID in parentheses. Results show strong in-group/out-group polarization across all four groups.

Figure B.2 presents the density plot of feeling thermometer score differences for vaccination-based groups. As with individuals who were cross-pressured by masking, the distribution of thermometer scores for those who are cross-pressured by their partisanship and vaccination status are centered near the point of indifference (50), but with large variation. As before, the notable exception is for compliant Democrats whose scores are tightly clustered around the extremes. (Unvaccinated Republicans are also relatively positive toward unvaccinated while still expressing moderate views toward the vaccinated.)

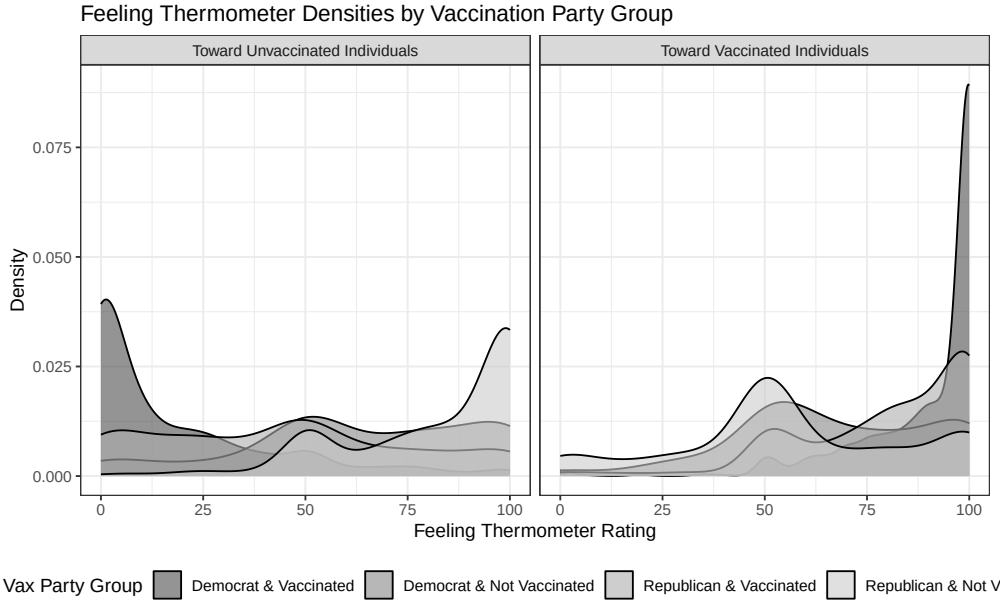

Figure B.2: Density of Feeling Thermometer Scores Toward Masked and Unmasked by Compliance and Partisanship

In addition to the distribution of thermometer scores it is also of interest to consider the extent to which the scores are correlated – i.e., whether expressing warmth towards one side is associated with expressing coldness towards the opposite. Given that individuals could differ in their compliance

with masking and vaccination – see Table A.1 – we report the correlations separately by partisanship and masking compliance (Table B.2) and also by partisanship and vaccination compliance (Table B.3). These tables correspond to the scatterplots present in Figure **??** in the main text.

Table B.2: Feeling Thermometer Correlations by Partisanship and Masking Status

Response 1	Response 2	Democrat & Masked	Democrat & Unmasked	Republican & Masked	Republican & UnMasked
Mask-Wearers	Non-Maskers	-0.32	-0.15	-0.38	-0.19
Mask-Wearers	Vaccinated	0.50	0.48	0.42	0.55
Mask-Wearers	Unvaccinated	-0.19	-0.16	-0.15	-0.19
Non-Maskers	Vaccinated	-0.23	-0.27	-0.14	-0.09
Non-Maskers	Unvaccinated	0.56	0.56	0.49	0.64
Vaccinated	Mask-Wearers	0.50	0.48	0.42	0.55
Vaccinated	Unvaccinated	-0.36	-0.40	-0.36	-0.28
Unvaccinated	Mask-Wearers	-0.19	-0.16	-0.15	-0.19
Unvaccinated	Non-Maskers	0.56	0.56	0.49	0.64

Table B.3: Feeling Thermometer Correlations by Partisanship and Vaccination Status

Response 1	Response 2	Democrat & Vaxed	Democrat & Unvaxed	Republican & Vaxed	Republican & Unvaxed
Mask-Wearers	Non-Maskers	-0.31	-0.56	-0.40	-0.34
Mask-Wearers	Vaccinated	0.50	0.47	0.34	0.60
Mask-Wearers	Unvaccinated	-0.20	-0.02	-0.22	-0.12
Non-Maskers	Vaccinated	-0.22	-0.23	-0.09	0.00
Non-Maskers	Unvaccinated	0.64	0.28	0.61	0.49
Vaccinated	Mask-Wearers	0.50	0.47	0.34	0.60
Vaccinated	Unvaccinated	-0.22	-0.16	-0.19	-0.04
Unvaccinated	Mask-Wearers	-0.20	-0.02	-0.22	-0.12
Unvaccinated	Non-Maskers	0.64	0.28	0.61	0.49

C Predicting Thermometer Scores

Although the average differences we report in the text are consistent with what we would expect from individuals who are cross-pressured by their partisanship and compliance status, we can further decompose the relative association of partisanship and compliance status with expressed affect as we predict a respondent's score toward each compliance group as a function of: partisanship, compliance status, an interaction of partisanship and compliance status, and other pandemic-relevant and demographic considerations (e.g., age, race, whether the respondent already had covid or knew someone who died).

To show the relative association of partisanship and compliance with the feeling score being expressed, Figure C.1 plots the resulting regression coefficients and 95% confidence intervals for the substantively relevant covariates when predicting respondent's thermometer scores toward those who do and do not mask, as well as those who are and are not vaccinated.

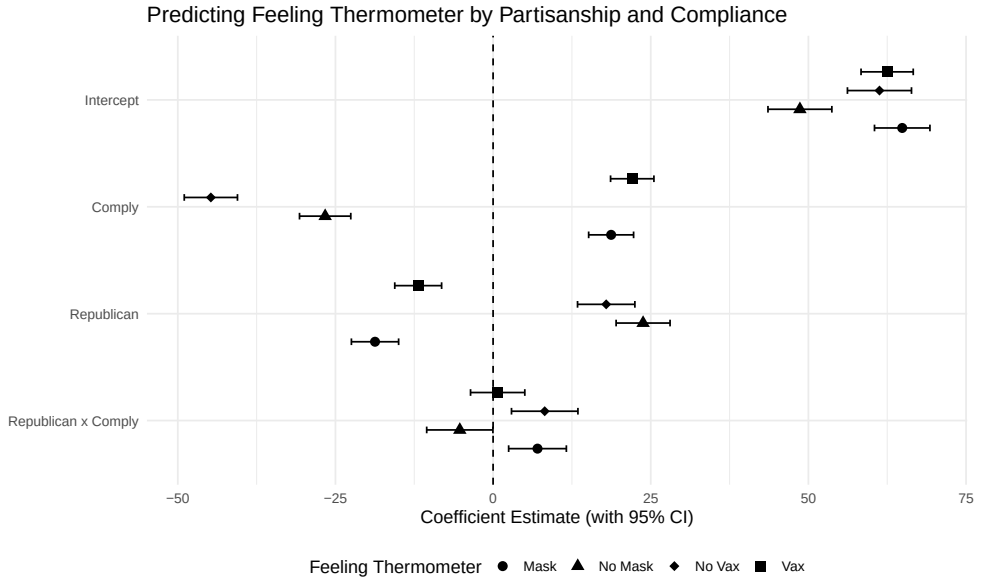

Figure C.1: Average Thermometer Scores by Party Identification and Behavior. See Table C.1 for complete coefficients.

The Intercept measures the average thermometer score for each group of compliers and non-compliers for the baseline category of non-compliant Democrats. As expected given the partisan and compliance cross-pressure they experience, non-compliant Democrats express feelings toward the various compliance-based groups that are relatively similar and also close to the midpoint of the scale (50). Compliance captures the average difference in scores between compliant and non-compliant Democrats and it shows differences that are both extreme – the positive and negative coefficients are bounded well away from zero – and also asymmetric (the increase in warmth expressed toward fellow compliers is eclipsed by the increase in coldness toward non-compliers). Consistent with expectations about reinforcing identities, compliant Democrats express the most extreme views. The Republican coefficient reflects average differences in responses between non-compliant Democrats and Republicans and it shows that not only is there a partisan difference, but also that Republicans are more favorably toward non-compliers and less favorable toward compliers in nearly equal amounts. Put differently, Republicans express more moderate opinions toward compliers and non-compliers than Democrats. Finally, the interaction term that allows a partisan

difference in the association with compliance is near zero (although a bit more warmer toward non-compliers) – suggesting that compliance-related differences do not much vary by partisanship.

The regression analysis in Figure C.1 shows significant compliance and partisan associations, but treating cross-pressure as a uniform phenomenon may obscure important heterogeneity. To better understand how partisan identity and behavioral compliance interact, we decompose our sample into four groups: Democrat Compliant (baseline), Democrat Non-Compliant, Republican Compliant, and Republican Non-Compliant.

Table C.1: OLS Models of Feeling Thermometer Ratings by Compliance and Party

	Dependent Variable:			
	Mask	No Mask	Vax	No Vax
	(1)	(2)	(3)	(4)
Republican	-21.46^{***}	27.05^{***}	-12.81^{***}	22.54^{***}
	(1.60)	(1.82)	(1.53)	(1.91)
Compliance	15.41^{***}	-20.82^{***}	15.30^{***}	-30.78^{***}
	(1.56)	(1.78)	(1.49)	(1.87)
Had COVID	2.29^{*}	0.47	2.58^{**}	1.89
	(1.17)	(1.35)	(1.13)	(1.41)
Age 30–39	1.80	1.73	0.29	5.43^{**}
	(1.78)	(2.03)	(1.71)	(2.13)
Age 40–49	1.38	0.48	3.11^{*}	2.15
	(1.77)	(2.03)	(1.71)	(2.11)
Age 50–64	2.92^{*}	0.11	6.58^{***}	-1.51
	(1.64)	(1.88)	(1.58)	(1.96)
Age 65+	4.65^{***}	0.09	9.87^{***}	-2.16
	(1.78)	(2.04)	(1.72)	(2.13)
Some College	-4.07^{***}	-0.37	1.31	-0.77
	(1.32)	(1.51)	(1.27)	(1.58)
BA Degree	-4.98^{***}	-1.65	2.63^{*}	-5.44^{***}
	(1.49)	(1.71)	(1.43)	(1.79)
Advanced Degree	-3.12	-3.79^{*}	1.81	-7.03^{***}
	(1.92)	(2.21)	(1.85)	(2.30)
Female	5.71^{***}	-2.56^{**}	-0.85	4.52^{***}
	(1.02)	(1.17)	(0.98)	(1.22)
Republican × Compliance	10.58^{***}	-11.10^{***}	2.70	-2.35
	(2.25)	(2.57)	(2.16)	(2.69)
Constant	72.01^{***}	39.42^{***}	69.89^{***}	46.44^{***}
	(2.03)	(2.34)	(1.95)	(2.43)
Observations	2,032	2,035	2,013	2,027
R^2	0.326	0.383	0.242	0.432
Adj. R^2	0.322	0.379	0.238	0.429

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

D Coding Rubric

The following is the instructions that were provided to the undergraduate research assistants who coded every open-ended response.

Coding Task

You have been selected to assist with a research project analyzing open-ended responses from a national public opinion survey. In this task, you will read short, text-based answers that survey respondents gave to questions about COVID-19 behaviors and perceptions. Your contribution will help uncover the emotional tone, moral judgments, and expressive content behind how people describe others in a time of national crisis. This isn't just busywork—it's hands-on participation in academic research, and your careful attention is essential to its success. Not only will your work help improve our understanding of political science and public health, but it will also have important connections to our understanding of Large Language Models (LLMs).

Each response you read was collected from one of the following four questions:

- **Mask-Wearers:** "What words or phrases come to mind when you think of someone who wears a mask in public places?"
- **Non-Mask-Wearers:** "What words or phrases come to mind when you think of someone who does not wear a mask in public places?"
- **Vaccinated Individuals:** "What words or phrases come to mind when you think of someone who has been vaccinated against COVID-19?"
- **Unvaccinated Individuals:** "What words or phrases come to mind when you think of someone who has not been vaccinated against COVID-19?"

Your job is to evaluate each response according to several key traits that capture the tone, content, and effort behind the answer. These traits are described in detail below, along with a coding scale and examples to help guide your judgment. Please review the rubric carefully before you begin coding to ensure consistency and accuracy across responses.

For each response you will enter 4 numerical values using the following coding rules:

- quality (1–5)
- emotion (1–5)
- valence (-1, 0, 1)
- moral (0/1)

Your coding will be combined with others to create a final code that will be used to help assess the nature of public opinion and beliefs.

General Notes for Coders

- Use your best judgment but stay consistent.
- If a response includes multiple elements (e.g., both neutral and angry phrases), score based on the dominant tone.

- If unclear, flag the row and add a note.
- It may work best if you chose one characteristic (e.g., quality) and evaluate several responses according to that characteristic rather than trying to rate every text as this will help ensure a consistent definition of that trait while coding.

IMPORTANT: Do not use ChatGPT or any other AI tool to assist with coding. One of the goals of this project is to compare human-coded judgments to those generated by artificial intelligence. Your intuition and insight as a human coder are exactly what we're studying—so no shortcuts!

Thank you for helping ensure high-quality data!

Open-Ended Survey Response Coding Rubric

1. Quality How much effort did the respondent appear to give in answering the question? Quality is about cognitive effort and how much thought is articulated, not expressive force.
Scale: 1 (No Effort) to 5 (High Effort)

Score	Description	Examples
1	No effort. Blank, gibberish, or irrelevant.	"none", "-", "asdfgh", "NA"
2	Very minimal response; short and vague.	"okay", "whatever", "fine"
3	Brief but interpretable response.	"selfish", "unvaccinated", "smart"
4	Sentence or multi-word phrase with some context or explanation.	"They made their own decision", "Not sure but seems risky"
5	Thoughtful, developed, and expressive response.	"They probably did their own research and chose what they thought was right for their family"

NOTE: A response can be low-quality but high in emotion (e.g., "stupid!"), or high-quality and low in emotion (e.g., "They weighed the risks and decided it was unnecessary"). Don't assume emotional intensity = effort.

2. Emotional Intensity How emotionally intense is the response?
Scale: 1 (None) to 5 (Extreme Emotion)

Score	Description	Examples
1	No emotion.	"It's their choice."
2	Mild emotion.	"Not a great idea"
3	Moderate emotion.	"I'm a little scared for them"
4	Strong emotion.	"They are selfish and putting others at risk"
5	Intense emotional expression (anger, joy, fear, etc.).	"Idiots! They are dangerous and brainwashed morons!"

3. Valence Overall tone or emotional direction of the response. i.e., how it makes you "feel."
Scale: -1 (Negative), 0 (Neutral/Ambiguous), 1 (Positive)

Value	Description	Examples
-1	Negative tone, judgment, or affect.	"Stupid", "Selfish", "Dangerous", "Morons"
0	Neutral, factual, or ambiguous sentiment.	"Independent", "It's their choice", "Not sure"
1	Positive tone or sentiment.	"Smart", "Cautious", "Caring", "Good idea"

4. Moral Judgment Does the respondent imply a moral evaluation of the subject (right/wrong, good/bad)? i.e., does it involve a moral judgement about the person.
Scale: -1 (Negative moral evaluation), 0 (No moral language), 1 (Positive moral evaluation)

Value	Description	Examples
-1	Negative moral judgment. Includes language implying wrong, bad, shouldn't, responsibility, blame, or condemnation.	"Selfish", "Irresponsible"
0	No moral implications or judgments. Descriptive or factual statements.	"Healthy", "They made a choice"
1	Positive moral judgement. Includes language implying right, good, should, responsibility, praise, virtue.	"Doing the right thing", "Responsible"

NOTE: Valence and Moral Judgment may seem similar, but a response of "IDIOT!" would be an example of a negative valence (-1) that has no moral judgement (0). A response of "Lazy!" would be both negative valence and negative moral (-1). Some of them may be tricky – which is the point! – so use your discretion as to what you think. The most important thing is to be consistent.

E Predicting Codeable Open-Ended Responses

We begin by exploring whether there is systemic variation in who opts to use an open-end response. To do so we use a linear probability model with statistical controls to determine if the willingness to respond covaries with opinion extremity, political engagement, or other characteristics that may make it difficult to generalize from the text responses. Insofar as the people who respond to an open-ended are systematically different than those who do not, then it becomes difficult to reach inferences on the basis of the text responses without further adjustment. In such cases, however, we can use the factors that predict a willingness to respond to construct weights that account for the differential willingness to respond by constructing a weight that equalizes the response probability.

$$Y_i = \beta_0 + \boldsymbol{\beta}_{\text{Group}}^{\top}\mathbf{G}_i + \boldsymbol{\beta}_{\text{Demo}}^{\top}\mathbf{D}_i + \boldsymbol{\beta}_{\text{Pandemic}}^{\top}\mathbf{P}_i + \boldsymbol{\beta}_{\text{Meta}}^{\top}\mathbf{M}_i + \varepsilon_i$$

The equation above presents a linear probability model in compact matrix notation, where the binary dependent variable Y_i indicates whether respondent i provided an "uncodable" open-ended response. The model includes four blocks of predictors: $\mathbf{G}_i$ is a vector of group membership with compliant Democrats as the baseline, $\mathbf{D}_i$ contains demographic characteristics such as age group, gender, education, and race/ethnicity, $\mathbf{P}_i$ captures pandemic-specific experiences related to getting covid or knowing someone personally who died, and $\mathbf{M}_i$ includes metadata such as the respondent's device operating system.

The results are presented, in full, below. As is immediately clear, cross-pressured groups are indeed more likely to provide an uncodable response, but the effect is not universal and generally small. Only unvaccinated Democrats are more likely to provide an uncodable response when asked to describe those who are vaccinated. While arguably the most-cross-pressured given the overall importance of partisanship in structuring beliefs, the magnitude of this difference is rather small – the probability they provide an uncodable response is only 0.08 higher than a compliant Democrat. The relationship on other responses is generally scattered and substantively small.

Table E.1: Predicting Uncodable Open-Ended Responses

	Dependent variable:			
	Toward Unvaccinated	Toward Vaccinated	Toward Unmasked	Toward Masked
	(1)	(2)	(3)	(4)
Unvaccinated Democrat	0.085***	0.065**		
	(0.026)	(0.026)		
Vaccinated Republican	0.011	0.020		
	(0.019)	(0.018)		
Unvaccinated Republican	0.021	0.015		
	(0.020)	(0.021)		
Unmasked Democrat			0.008	−0.006
			(0.027)	(0.023)
Masked Republican			0.018	0.006
			(0.020)	(0.018)
Unmasked Republican			0.033*	−0.018
			(0.018)	(0.017)
FT Extremity	−0.037*	−0.073***	−0.056***	−0.059***
	(0.021)	(0.021)	(0.020)	(0.018)
Got Covid	0.011	0.005	−0.011	0.001
	(0.017)	(0.016)	(0.016)	(0.014)
Covid Death	−0.020	−0.025*	−0.004	0.004
	(0.015)	(0.014)	(0.014)	(0.013)
Age 30–39	0.023	0.008	−0.003	0.004
	(0.025)	(0.023)	(0.025)	(0.022)
Age 40–49	0.022	0.012	0.009	−0.008
	(0.024)	(0.023)	(0.025)	(0.022)
Age 50–64	−0.002	−0.007	0.001	−0.017
	(0.023)	(0.022)	(0.023)	(0.020)
Age 65+	0.024	0.002	−0.027	−0.036
	(0.025)	(0.024)	(0.025)	(0.022)
Some College	0.003	0.018	−0.019	0.003
	(0.018)	(0.017)	(0.018)	(0.016)
BA Degree	−0.016	−0.002	−0.046**	−0.008
	(0.021)	(0.020)	(0.021)	(0.018)
Advanced Degree	0.004	−0.010	−0.020	−0.026
	(0.028)	(0.027)	(0.027)	(0.023)
Female	−0.008	−0.009	−0.026*	−0.019
	(0.014)	(0.014)	(0.014)	(0.012)
Black	0.055**	0.026	−0.002	−0.025
	(0.023)	(0.022)	(0.024)	(0.021)
Hispanic	−0.002	−0.020	−0.008	0.009
	(0.023)	(0.022)	(0.022)	(0.019)
OS: Mac	−0.010	−0.020	0.002	0.005
	(0.032)	(0.031)	(0.030)	(0.027)
OS: iOS	−0.020	−0.008	0.0002	0.016
	(0.021)	(0.020)	(0.021)	(0.018)
OS: Android	0.006	0.022	0.037**	0.016
	(0.017)	(0.016)	(0.018)	(0.015)
OS: ChromeOS	0.097**	0.027	0.008	0.016
	(0.047)	(0.045)	(0.044)	(0.039)
OS: Linux	0.035	−0.060	0.093	−0.026
	(0.073)	(0.077)	(0.084)	(0.074)
OS: Other	−0.080	−0.062	−0.014	0.020
	(0.226)	(0.214)	(0.222)	(0.195)
Constant	0.055*	0.081**	0.101***	0.096***
	(0.032)	(0.032)	(0.032)	(0.029)
Observations	1,045	1,034	1,013	1,012
R^2	0.041	0.049	0.039	0.026
Adjusted R^2	0.020	0.029	0.018	0.004

F Comparing Human and LLM Coders

Rather than relying on any single coder as ground truth, we employ two complementary strategies. First, we reshape the data into long format — one observation per coder per response — and include coder fixed effects to absorb systematic mean differences across coders (including between the LLM and humans). This approach yields a weighted average that strips out coder-specific level shifts without requiring the strong assumptions of bias-correction methods. Second, we estimate all models using only the three human coders' ratings, providing results that are entirely independent of the LLM. Comparing the coder fixed effects and human-only results provides a robustness check across measurement approaches rather than privileging any single coding as correct.

Before applying this framework to evaluate response quality, we compare the reliability of codings (pooling across all four open-ended responses).

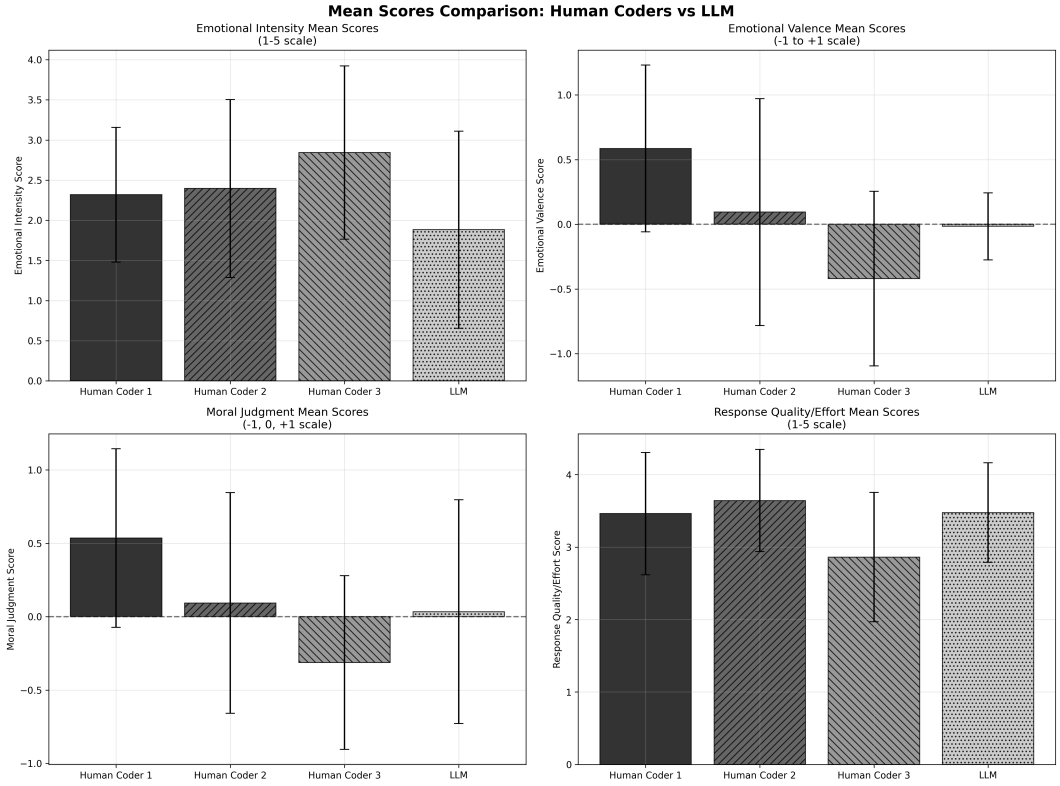

Figure F.1: Mean scores and standard deviations across psychological measures for three human coders and the LLM. All three human coders show similar central tendencies on directional measures (valence, moral judgment), though Human Coder 3 rates quality somewhat lower than the others. Pairwise correlations and formal reliability statistics are reported in Appendix J.

Human coders achieve substantial to almost-perfect agreement on valence (average pairwise $\kappa = 0.772$, $r = 0.840$) and moderate-to-substantial agreement on moral judgment ($\kappa = 0.578$, $r = 0.689$), while emotional intensity proves hardest to code consistently ($\kappa = 0.237$, $r = 0.534$). Figure F.1 compares the average score for the 3 human and 1 LLM coder. While the three human coders agree closely on directional measures, they show modest mean-level differences on quality (Human Coder 3 rates quality somewhat lower than the others). These patterns motivate our use of coder fixed effects rather than bias-correction approaches: mean-level differences across coders are precisely the kind of additive shifts that fixed effects are designed to absorb.

G Permutation Test for Text Similarity

As a nonparametric robustness check for the dyad regression results presented in Table 3, we conduct a conditional permutation test. For each domain, we compute the test statistic $\Delta = \overline{\text{sim}}_{\text{same-party}} - \overline{\text{sim}}_{\text{diff-party}}$ among respondent pairs that share the same compliance status. We then generate a null distribution by shuffling party labels *within* each compliance group (5,000 iterations), preserving the compliance structure while breaking any party–language association.

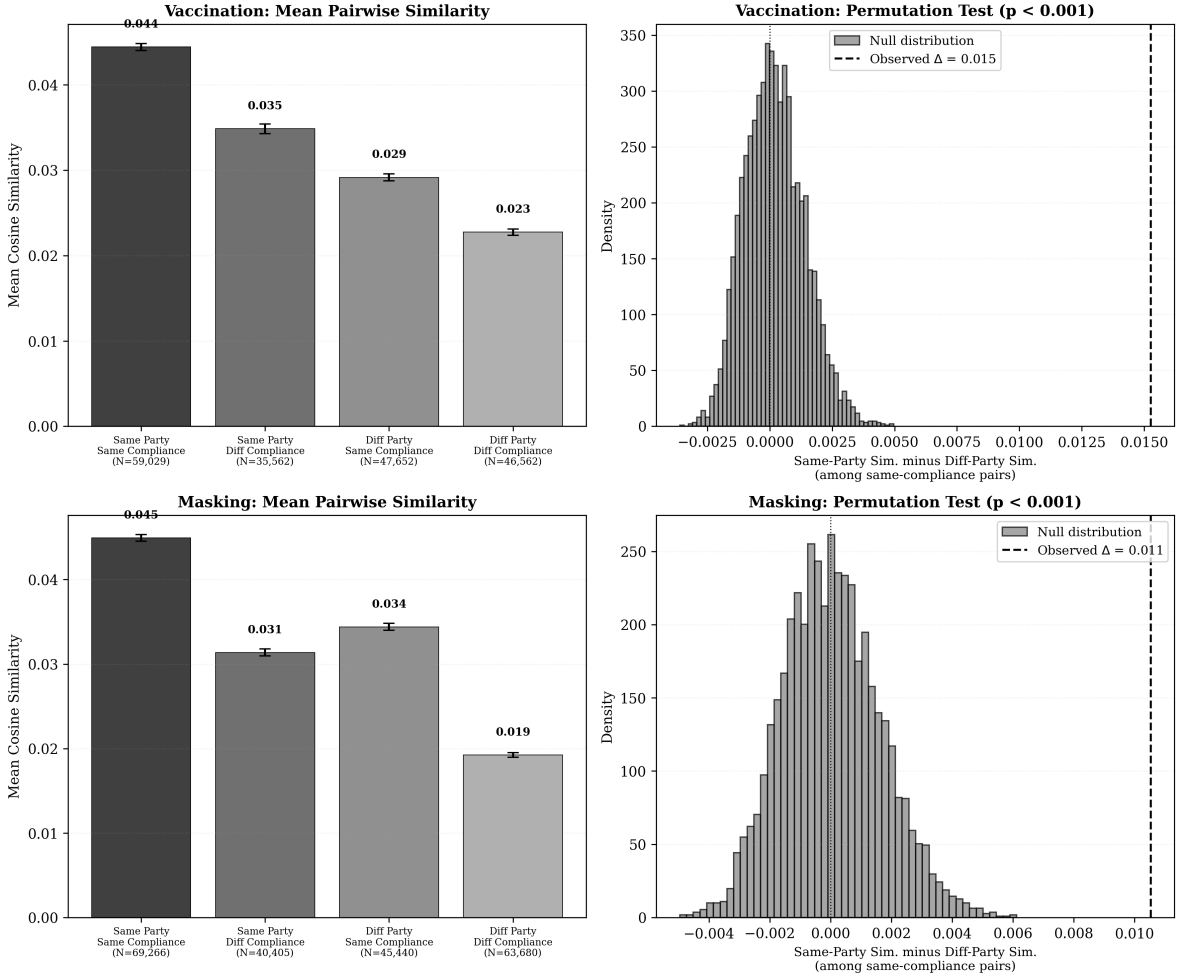

Figure G.1: Response-Level Cosine Similarity and Permutation Test. Left panels show mean pairwise cosine similarity of individual TF-IDF vectors by pair category (with standard error bars). Right panels show the null distribution from a conditional permutation test (5,000 iterations) that shuffles party labels within compliance groups. The observed party effect (Δ) is the difference in mean similarity between same-party and different-party pairs among respondents who share the same compliance status. Top row: vaccination; bottom row: masking.

Figure G.1 confirms that co-partisans use more similar vocabulary than cross-partisans among same-compliance pairs ($\Delta = 0.015$, $p < 0.001$ for vaccination; $\Delta = 0.011$, $p < 0.001$ for masking). These nonparametric results are consistent with the Same Party coefficients from the dyad regression (0.012 for both domains), providing converging evidence under different inferential assumptions.

H Topic Modeling and Diversity Measures

LDA Implementation Latent Dirichlet Allocation was implemented using scikit-learn's *LatentDirichletAllocation* with 6 topics and batch learning method. Text preprocessing included combining all four open-ended responses per individual, removing custom stop words (standard English stop words plus political terms like "people," "covid," "vaccine"), and creating a document-term matrix using CountVectorizer with minimum document frequency of 3 and maximum document frequency of 70%. The resulting model produces a probability distribution over topics for each response.

Table H.1 presents the table of extracted topics, the name we assigned to each topic based on its content, the percentage of open-ended responses that were associated with the topic and the top words defining the issue. A comment was classified as involving the topic if the probability

Table H.1: Distribution of LDA Topic Indicators Across All Open-Ended Responses

Topic ID	Topic Name	Coverage	Top Words
0	Individual Rights & Selfishness	22.9%	selfish, right, choose, freedom
1	Indifference & Normalcy	27.4%	don't care, health, business, normal
2	Intelligence & Responsibility	39.3%	smart, caring, responsible, intelligent
3	Personal Autonomy & Safety	26.2%	choice, personal, safe, protecting
4	Scientific Responsibility	24.0%	cares, smart, health, science
5	Moral Action & Bodily Choice	25.0%	right, doing, choice, decision, body

Note: Coverage indicates percentage of responses where topic indicator = 1.
Topics identified using Latent Dirichlet Allocation with 6 components.
Binary indicators created using probability threshold of $1/6 \approx 0.167$.

Table H.2 pivots to present the extent to which the open-ended responses of each group were classified as raising each issue.

Table H.2: LDA Topic Coverage by Cross-Pressured Groups

Topic	Unvaccinated Dem	Vaccinated Rep	Unmasked Dem	Masked R
Individual Rights	16.9%	11.4%	17.0%	11.8%
Indifference / Normalcy	18.3%	11.7%	19.7%	12.0%
Intelligence / Responsibility	10.7%	40.4%	11.1%	39.3%
Personal Autonomy / Safety	24.1%	8.9%	23.8%	9.3%
Scientific Responsibility	11.4%	17.1%	8.9%	17.1%
Moral Action / Bodily Integrity	18.6%	10.5%	19.5%	10.5%

Note: Percentages show the proportion of respondents in each group whose dominant topic assignment was the one listed in the column.
Based on LDA analysis of pooled open-ended responses from all four questions.
Topics determined by highest-probability assignment per respondent.
These values were used to calculate topic diversity measures in the main analysis.

Topic Dispersion Calculations All metrics use normalized topic probabilities where p_i represents the weight of topic i for each response: - Herfindahl Index: $H = \sum_{i=1}^{n} p_i^2$ (topic concentration, range: $1/n$ to 1, higher = more focused) - Entropy: $E = -\sum_{i=1}^{n} p_i \log(p_i)$ (topic diversity, range: 0

to $\log(n)$, higher = more balanced) - Effective Topics: $ET = 1/H$ (effective number of substantively engaged topics, range: 1 to n)

Model Validation The 6-topic solution was selected based on coherence scores and interpretability. Topics captured substantive political themes including personal autonomy, health/safety considerations, social responsibility, government authority, economic concerns, and moral evaluation.

Table H.3: Effects of Cross-Pressure on Topic Dispersion and Coherence: Vaccination-Based Groups

	Dependent Variable		
	Mechanical Text Metrics		
	Topic Concentration (Herfindahl Index) (1)	Topic Diversity (Entropy) (2)	Number of Distinct Topics Used (3)
Unvaccinated Dem	−0.012	0.057	0.331
	(0.022)	(0.048)	(0.249)
Vaccinated Rep	0.027	−0.023	−0.035
	(0.029)	(0.062)	(0.269)
Unvaccinated Rep	0.043	−0.034	−0.052
	(0.032)	(0.068)	(0.290)
FT Extremity	0.024	−0.069	−0.333
	(0.019)	(0.042)	(0.223)
Controls	Age, Female, Education, Party ID Strength, Log Word Count, Mobile Device		
R^2	0.013	0.043	0.026
Adjusted R^2	0.004	0.034	0.017
Observations	1,101	1,101	1,101

Note: Robust standard errors in parentheses. *p<0.05; **p<0.01; ***p<0.001.
Baseline: Vaccinated Dem.
Higher Herfindahl ⇒ more concentrated topics; higher entropy and more distinct topics ⇒ more diversity.
FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).
None of the group coefficients reach statistical significance. The analogous masking analysis yields similarly null results (Table L.7).

I Content Theme Analysis

Table I.1: Content Theme Prevalence by Cross-Pressure Group (%)

Panel A: Vaccination						
Theme	Compliant Dem	Unvax Dem	Vax Rep	Unvax Rep	χ^2	p
Helpful/Beneficial	31.1	6.6	14.2	1.7	129.1	<0.001
Smart/Responsible	30.2	18.9	22.3	18.3	18.8	<0.001
Foolish/Irresponsible	31.6	10.4	15.2	10.7	70.7	<0.001
Unhelpful/Ineffective	23.8	1.9	8.4	0.0	120.6	<0.001
Personal Choice	1.7	15.1	19.7	27.2	128.9	<0.001
Science/Evidence	3.7	1.9	0.6	0.3	14.8	0.002
Elites/Leaders	4.2	0.9	1.3	1.4	10.7	0.01
Mitigation	0.0	3.8	1.3	1.4	15.6	0.001
Panel B: Masking						
Theme	Compliant Dem	Unmask Dem	Mask Rep	Unmask Rep	χ^2	p
Helpful/Beneficial	37.1	22.0	20.1	5.1	129.3	<0.001
Smart/Responsible	29.8	26.3	24.5	16.5	21.5	<0.001
Foolish/Irresponsible	28.2	12.7	19.7	15.1	29.7	<0.001
Unhelpful/Ineffective	34.3	13.6	16.2	3.2	140.9	<0.001
Personal Choice	3.0	9.3	11.8	29.5	140.9	<0.001
Health/Protection	0.9	5.9	2.6	3.8	14.1	0.003
Elites/Leaders	3.8	2.5	0.9	1.9	6.4	0.09

Note: Percentage of respondents whose open-ended response mentions each content theme (coded by LLM from response text). A response is counted if the theme appears in either the pro or anti question for that domain. χ^2 tests whether theme prevalence differs across the four groups. Themes with <3% maximum prevalence are omitted.

J Inter-Coder Reliability Analysis

Table J.1: Inter-Coder Reliability Results

	Cohen's Kappa (Pairwise)		
Construct	Human Range	Human Avg.	Best Pair
Valence	0.743–0.829	0.772	KS–PZ (0.829)
Moral Judgment	0.537–0.619	0.578	MP–KS (0.619)
Quality	0.205–0.439	0.325	MP–KS (0.439)
Emotional Intensity	0.211–0.251	0.237	MP–PZ (0.251)

Note: Human range and average are computed over the three human coder pairs (MP–KS, MP–PZ, KS–PZ). All pairwise κ values are significant at $p < 0.001$.

Kappa interpretation: <0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, 0.81–1.00 = almost perfect.

K Human Coder Analysis of Psychological Content

Table K.1: Cross-Pressure Effects on Psychological Content: Human Coder Analysis (Vaccination)

	Emotional Intensity (1)	Valence (2)	Moral Judgment (3)
Unvaxed Dem	-0.451^{***}	0.077	0.035
	(0.101)	(0.052)	(0.077)
Vaxed Rep	-0.460^{***}	0.168^{***}	0.086^{*}
	(0.059)	(0.032)	(0.043)
Unvaxed Rep	-0.583^{***}	0.158^{***}	-0.012
	(0.067)	(0.031)	(0.040)
Age	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Female	-0.035	-0.004	-0.014
	(0.048)	(0.025)	(0.035)
Education	0.103	-0.017	-0.010
	(0.053)	(0.027)	(0.039)
Constant	3.101^{***}	-0.040	0.176^{**}
	(0.084)	(0.045)	(0.067)
R^2	0.093	0.011	0.004
Adj. R^2	0.089	0.007	-0.001
Observations	1,485	1,485	1,196

Note: Standard errors clustered by ResponseId in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Baseline: Compliant Democrats (Democrats who are vaccinated).
Human coding based on average ratings across all available coders.
Analysis pools responses to both vaccination questions (each respondent contributes two observations).

Table K.2: Cross-Pressure Effects on Psychological Content: Human Coder Analysis (Masking)

	Emotional Intensity (1)	Valence (2)	Moral Judgment (3)
Non-Masked Dem	-0.604^{***}	0.089	0.038
	(0.096)	(0.058)	(0.040)
Masked Rep	-0.478^{***}	0.055	0.035
	(0.077)	(0.036)	(0.033)
Non-Masked Rep	-0.789^{***}	-0.026	-0.046
	(0.067)	(0.030)	(0.026)
Age	0.001	0.001	0.002^{***}
	(0.002)	(0.001)	(0.001)
Female	-0.066	0.064^{*}	0.034
	(0.054)	(0.025)	(0.021)
Education	0.120^{*}	-0.021	-0.003
	(0.058)	(0.026)	(0.023)
Constant	2.842^{***}	0.033	-0.034
	(0.097)	(0.038)	(0.037)
R^2	0.144	0.004	0.005
Adj. R^2	0.140	-0.001	0.000
Observations	1,317	1,317	1,317

Note: Standard errors clustered by ResponseId in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Baseline: Compliant Democrats (Democrats who wear masks).
Human coding based on average ratings across all available coders.
Analysis pools responses to both masking questions (each respondent contributes two observations).

L Extended Engagement and Topic Diversity Results

This section presents the complete set of regression results for engagement and topic diversity measures across both vaccination and masking behaviors. The main text focuses on vaccination results using the four-group mean differences approach. Here we provide corresponding masking results and interaction specifications to demonstrate robustness across analytical approaches.

L.1 Engagement Measures

Table L.1: Four-Group Effects on Engagement Measures: Mask-Based Groups

	Dependent Variable			
	Mechanical Text Metrics		Coded Response Quality	
	Word Count	Lexical Diversity	Response Quality (All Coders, FE)	Response Quality (Human Coders, FE)
	(1)	(2)	(3)	(4)
Unmasked Dem	0.081	0.096	0.066	0.051
	(0.119)	(0.127)	(0.127)	(0.140)
Masked Rep	0.002	-0.000	−0.099	−0.075
	(0.101)	(0.094)	(0.075)	(0.081)
Unmasked Rep	0.142	0.089	0.062	0.045
	(0.090)	(0.078)	(0.071)	(0.077)
FT Extremity	-0.012	-0.064	−0.090	0.037
	(0.098)	(0.088)	(0.078)	(0.086)
Controls	Age, Female, Education, Party ID Strength, News Consumption, Mobile Device			
Coder Fixed Effects	—	—	Yes	Yes
R^2	0.023	0.009	0.065	0.064
Adjusted R^2	0.015	0.001	0.061	0.059
Observations	1,249	1,249	3,592	2,348

Note: *p<0.05; **p<0.01; ***p<0.001. Clustered standard errors by respondent ID in parentheses.
Baseline: Masked Dem (compliant Democrats).
Columns (1)–(2) use one observation per respondent × question. Columns (3)–(4) use coder fixed effects as described in Table **??**.
FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

L.2 Robustness: Full Sample Effort Analysis

The engagement analyses above restrict the sample to respondents who provided at least six words in total across the four open-ended questions. This filter removes 41% of the partisan sample (902 of 2,179 respondents). A natural concern is whether this selection accounts for the effort results. Tables L.4 and L.5 compare the effort regressions on the full partisan sample (including respondents who wrote fewer than six words or left all responses blank) with the filtered subsample.

The key finding—that non-compliant Republicans write significantly more than compliant Democrats—is robust to including the full sample. Coefficient magnitudes and significance levels are similar across both samples. Table L.6 shows that the probability of providing a short

Table L.2: Effects of Partisanship and Compliance on Engagement: Vaccination (Interaction Specification)

	Dependent Variable			
	Mechanical Text Metrics		Response Quality	
	Word Count	Lexical Diversity	All Coders (FE)	Human Coders (FE)
	(1)	(2)	(3)	(4)
Republican	0.182*	0.083	0.015	−0.106
	(0.081)	(0.085)	(0.067)	(0.073)
Unvaccinated	0.345	0.173	−0.160	−0.289
	(0.231)	(0.148)	(0.142)	(0.151)
Rep × Unvax	−0.111	0.001	0.188	0.294
	(0.269)	(0.181)	(0.166)	(0.177)
FT Extremity	0.033	−0.019	0.042	0.157
	(0.114)	(0.087)	(0.077)	(0.082)
Controls	Age, Female, Education, Party ID Strength, News Consumption, Mobile Device			
Coder Fixed Effects	—	—	Yes	Yes
R^2	0.038	0.025	0.130	0.187
Adjusted R^2	0.030	0.017	0.126	0.183
Observations	1,184	1,184	3,406	2,224

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Clustered standard errors by respondent ID in parentheses.
Baseline: Vaccinated Dem.
Columns (3)–(4) pool across coders (one observation per coder × response) with coder fixed effects. Column (4) restricts to human coders only.
FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

Table L.3: Effects of Partisanship and Compliance on Engagement: Masking (Interaction Specification)

	Dependent Variable			
	Mechanical Text Metrics		Response Quality	
	Word Count	Lexical Diversity	All Coders (FE)	Human Coders (FE)
	(1)	(2)	(3)	(4)
Republican	0.002	−0.000	−0.099	−0.075
	(0.101)	(0.094)	(0.075)	(0.081)
Unmasked	0.081	0.096	0.066	0.051
	(0.119)	(0.127)	(0.127)	(0.140)
Rep × Unmasked	0.058	−0.007	0.096	0.069
	(0.158)	(0.163)	(0.150)	(0.166)
FT Extremity	−0.012	−0.064	−0.090	0.037
	(0.098)	(0.088)	(0.078)	(0.086)
Controls	Age, Female, Education, Party ID Strength, News Consumption, Mobile Device			
Coder Fixed Effects	—	—	Yes	Yes
R^2	0.023	0.009	0.065	0.064
Adjusted R^2	0.015	0.001	0.061	0.059
Observations	1,249	1,249	3,592	2,348

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Clustered standard errors by respondent ID in parentheses.
Baseline: Masked Dem (compliant Democrats).
Columns (3)–(4) pool across coders (one observation per coder × response) with coder fixed effects. Column (4) restricts to human coders only.
FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

response (<6 total words) is roughly balanced across groups, with non-compliant Republicans actually *least* likely to provide a short response (34% for unvaccinated Republicans vs. 42% for vaccinated Democrats). The six-word filter therefore does not differentially exclude any partisan-compliance group.

Table L.4: Effort Analysis: Full Sample vs. Filtered—Vaccination-Based Groups

	Full Sample			**Filtered (≥6 words)**		
	Unvaccinated Dem	Vaccinated Rep	Unvaccinated Rep	Unvaccinated Dem	Vaccinated Rep	Unvaccinated Rep
Baseline: Vaccinated Dem						
Word Count	0.130	0.032	0.327***	0.163	0.091	0.325***
	(0.103)	(0.049)	(0.067)	(0.145)	(0.065)	(0.080)
	$N = 2{,}126,\ R^2 = 0.035$			$N = 1{,}246,\ R^2 = 0.034$		
Log Word Count	0.106	−0.011	0.313***	0.119	0.086	0.340***
	(0.081)	(0.053)	(0.062)	(0.111)	(0.070)	(0.077)
	$N = 2{,}126,\ R^2 = 0.037$			$N = 1{,}246,\ R^2 = 0.038$		

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Clustered standard errors (by respondent) in parentheses.
All dependent variables are standardized within each sample. Word count is the total across all four open-ended questions.
Controls: age, gender, education, party ID strength, news consumption, mobile device.

Table L.5: Effort Analysis: Full Sample vs. Filtered—Masking-Based Groups

	Full Sample			**Filtered (≥6 words)**		
	Unmasked Dem	Masked Rep	Unmasked Rep	Unmasked Dem	Masked Rep	Unmasked Rep
Baseline: Masked Dem						
Word Count	0.095	0.090	0.206***	0.122	0.161*	0.223**
	(0.095)	(0.059)	(0.054)	(0.132)	(0.078)	(0.068)
	$N = 2{,}126,\ R^2 = 0.027$			$N = 1{,}246,\ R^2 = 0.028$		
Log Word Count	0.059	0.039	0.186***	0.067	0.142	0.238***
	(0.079)	(0.060)	(0.054)	(0.107)	(0.081)	(0.069)
	$N = 2{,}126,\ R^2 = 0.028$			$N = 1{,}246,\ R^2 = 0.031$		

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Clustered standard errors (by respondent) in parentheses.
All dependent variables are standardized within each sample. Word count is the total across all four open-ended questions.
Controls: age, gender, education, party ID strength, news consumption, mobile device.

Table L.6: Short-Response Selection: Percentage of Respondents with <6 Total Words

Group	N	N Short (<6 words)	% Short
Panel A: Vaccination-Based Groups			
Vaccinated Dem	992	420	42.3%
Unvaccinated Dem	183	77	42.1%
Vaccinated Rep	564	255	45.2%
Unvaccinated Rep	440	150	34.1%
Overall	2,179	902	41.4%
Panel B: Masking-Based Groups			
Masked Dem	974	414	42.5%
Unmasked Dem	201	83	41.3%
Masked Rep	408	179	43.9%
Unmasked Rep	596	226	37.9%
Overall	2,179	902	41.4%

Note: Short responses defined as total word count < 6 across all four open-ended questions.
Full partisan sample: $N = 2{,}179$. The filtered sample used in the main text excludes short-response respondents.

L.3 Topic Diversity Measures

Table L.7: Effects of Cross-Pressure on Topic Diversity: Mask-Based Groups (Four-Group Analysis)

	Dependent Variable		
	Topic Concentration (Herfindahl Index)	Topic Diversity (Entropy)	Number of Distinct Topics Used
	(1)	(2)	(3)
Unmasked Dem	0.023	-0.031	-0.125
	(0.017)	(0.037)	(0.143)
Masked Rep	-0.002	0.038	0.126
	(0.029)	(0.062)	(0.259)
Unmasked Rep	0.001	0.030	0.054
	(0.029)	(0.062)	(0.252)
FT Extremity	0.023	-0.066	-0.104
	(0.017)	(0.035)	(0.155)
Controls	Age, Female, Education, Party ID Strength, Log Word Count, Mobile Device		
R^2	0.017	0.053	0.017
Adjusted R^2	0.008	0.045	0.009
Observations	1,167	1,167	1,167

Note: Robust standard errors in parentheses. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Baseline: Masked Dem (compliant Democrats).

Higher Herfindahl Index = more concentrated topics; higher entropy and distinct topics = more diversity.

FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

Table L.8: Effects of Partisanship and Compliance on Topic Diversity: Vaccination (Interaction Specification)

	Dependent Variable		
	Topic Concentration (Herfindahl Index)	Topic Diversity (Entropy)	Number of Distinct Topics Used
	(1)	(2)	(3)
Republican	0.027	-0.023	-0.035
	(0.029)	(0.062)	(0.269)
Unvaccinated	-0.012	0.057	0.331
	(0.022)	(0.048)	(0.249)
Rep × Unvax	0.029	-0.068	-0.348
	(0.026)	(0.057)	(0.300)
FT Extremity	0.024	-0.069	-0.333
	(0.019)	(0.042)	(0.223)
Controls	Age, Female, Education, Party ID Strength, Log Word Count, Mobile Device		
R^2	0.013	0.043	0.026
Adjusted R^2	0.004	0.034	0.017
Observations	1,101	1,101	1,101

Note: Robust standard errors in parentheses. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Baseline: Vaccinated Dem (compliant Democrats).

Higher Herfindahl Index = more concentrated topics; higher entropy and distinct topics = more diversity.

FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

Table L.9: Effects of Partisanship and Compliance on Topic Diversity: Masking (Interaction Specification)

	Dependent Variable		
	Topic Concentration (Herfindahl Index)	Topic Diversity (Entropy)	Number of Distinct Topics Used
	(1)	(2)	(3)
Republican	-0.002	0.038	0.126
	(0.029)	(0.062)	(0.259)
Non-Masker	0.023	-0.031	-0.125
	(0.017)	(0.037)	(0.143)
Rep × Non-Mask	-0.020	0.023	0.052
	(0.022)	(0.048)	(0.202)
FT Extremity	0.023	-0.066	-0.104
	(0.017)	(0.035)	(0.155)
Controls	Age, Female, Education, Party ID Strength, Log Word Count, Mobile Device		
R^2	0.017	0.053	0.017
Adjusted R^2	0.008	0.045	0.009
Observations	1,167	1,167	1,167

Note: Robust standard errors in parentheses. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Baseline: Masked Dem. Higher Herfindahl = more concentrated topics.

FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

M Psychological Content Analysis - Complete Results

This section presents the complete psychological content analysis results across both vaccination and masking behaviors. All psychological measures were coded by three human coders and one LLM. We report results using coder fixed effects (pooling all coders) and human-only ratings (restricting to the three human coders), both with coder fixed effects to absorb systematic rating differences.

Table M.1: Cross-Pressure Effects on Psychological Content: Vaccination Analysis

	Emotional Intensity		**Valence**		**Moral Judgment**	
	Coder FE	Human	Coder FE	Human	Coder FE	Human
	(1)	(2)	(3)	(4)	(5)	(6)
Unvaccinated Dem	-0.57^{***}	-0.59^{***}	0.32^{***}	0.17^{*}	0.20^{***}	0.15^{*}
	(0.097)	(0.127)	(0.070)	(0.068)	(0.061)	(0.064)
Vaccinated Rep	-0.53^{***}	-0.56^{***}	0.31^{***}	0.22^{***}	0.23^{***}	0.21^{***}
	(0.061)	(0.077)	(0.046)	(0.046)	(0.041)	(0.045)
Unvaccinated Rep	-0.72^{***}	-0.80^{***}	0.44^{***}	0.20^{***}	0.18^{***}	0.13^{**}
	(0.062)	(0.083)	(0.049)	(0.050)	(0.045)	(0.048)
FT Extremity	0.24^{***}	0.42^{***}	0.16^{**}	0.21^{***}	0.21^{***}	0.17^{**}
	(0.067)	(0.082)	(0.054)	(0.053)	(0.056)	(0.053)
Controls	Age, Female, Education, Party ID Strength, Log Word Count, Mobile Device					
Coder Fixed Effects	Yes					
R^2	0.231	0.208	0.192	0.158	0.111	0.161
Adjusted R^2	0.228	0.203	0.188	0.153	0.107	0.156
Observations	3,406	2,224	2,842	2,224	3,405	2,223

Note: Clustered standard errors by respondent in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Baseline: Vaccinated Dem.
Coder FE columns pool all available coders (3 human + LLM) with coder fixed effects to absorb systematic rating differences. Human columns restrict to the three human coders only, also with coder fixed effects. Each observation is one coder × one response.
FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).
Valence columns (3–4) have different sample sizes because LLM valence is available for a subset of responses.

Table M.2: Cross-Pressure Effects on Psychological Content: Masking Analysis

	Emotional Intensity		Valence		Moral Judgment	
	Coder FE	Human	Coder FE	Human	Coder FE	Human
	(1)	(2)	(3)	(4)	(5)	(6)
Unmasked Dem	−0.40***	−0.43***	0.16**	0.12	0.13*	0.12*
	(0.084)	(0.104)	(0.060)	(0.067)	(0.056)	(0.060)
Masked Rep	−0.30***	−0.31***	0.15***	0.15**	0.14**	0.15**
	(0.064)	(0.077)	(0.043)	(0.047)	(0.047)	(0.053)
Unmasked Rep	−0.54***	−0.58***	0.05	0.08*	0.06	0.05
	(0.066)	(0.082)	(0.037)	(0.040)	(0.035)	(0.038)
FT Extremity	0.40***	0.47***	0.11*	0.06	0.12**	0.08
	(0.069)	(0.084)	(0.050)	(0.048)	(0.047)	(0.045)
Controls	Age, Female, Education, Party ID Strength, Log Word Count, Mobile Device					
Coder Fixed Effects	Yes					
R^2	0.224	0.242	0.133	0.192	0.133	0.203
Adj. R^2	0.221	0.238	0.129	0.187	0.130	0.199
Observations	3,594	2,350	3,592	2,348	3,592	2,348

Note: Clustered standard errors (by ResponseId) in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Baseline: Masked Dem (compliant Democrats).
Coder FE columns pool all available coders with coder fixed effects. Human columns restrict to the three human coders only.
FT Extremity $= |\text{FT}_{\text{target}} - 50|/50$.

N LLM Implementation and Reproducibility Code

This section provides the complete implementation code used to extract psychological measures from open-ended survey responses via the OpenAI API. All prompts, configurations, and processing pipelines are included for full reproducibility.

N.1 Configuration and Dependencies

```
"""
LLM-Based Psychological Content Analysis
Complete implementation for extracting psychological measures from survey
 responses
"""

import openai
import pandas as pd
import numpy as np
```

Table M.3: Cross-Pressure Effects on Psychological Content: Vaccination Analysis (Interaction Specification) — Coder Fixed Effects and Human-Only Specifications

	Emotional Intensity		**Valence**		**Moral Judgment**	
	Coder FE (1)	Human (2)	Coder FE (3)	Human (4)	Coder FE (5)	Human (6)
Republican	-0.53^{***}	-0.57^{***}	0.31^{***}	0.23^{***}	0.23^{***}	0.21^{***}
	(0.060)	(0.076)	(0.046)	(0.045)	(0.041)	(0.045)
Unvaccinated	-0.57^{***}	-0.59^{***}	0.32^{***}	0.16^{*}	0.20^{***}	0.15^{*}
	(0.097)	(0.127)	(0.070)	(0.068)	(0.061)	(0.064)
Rep × Unvax	0.38^{**}	0.35^{*}	-0.19^{*}	-0.18^{*}	-0.25^{***}	-0.22^{**}
	(0.119)	(0.154)	(0.088)	(0.087)	(0.076)	(0.082)
FT Extremity	0.26^{***}	0.43^{***}	0.16^{**}	0.20^{***}	0.21^{***}	0.17^{**}
	(0.067)	(0.082)	(0.053)	(0.052)	(0.056)	(0.053)
Age	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Female	−0.02	−0.03	−0.04	−0.05	−0.02	−0.04
	(0.045)	(0.058)	(0.034)	(0.035)	(0.031)	(0.034)
Education	0.03	0.06^{**}	0.03^{*}	0.03^{*}	0.03^{*}	0.03^{*}
	(0.017)	(0.023)	(0.013)	(0.013)	(0.012)	(0.014)
Party ID Strength	0.04	0.03	−0.04	-0.05^{*}	−0.03	−0.03
	(0.028)	(0.038)	(0.021)	(0.022)	(0.019)	(0.021)
Log Word Count	0.22^{***}	0.20^{***}	−0.04	−0.04	−0.00	0.00
	(0.041)	(0.050)	(0.030)	(0.029)	(0.028)	(0.027)
Mobile Device	0.00	0.02	0.01	0.02	0.04	0.03
	(0.046)	(0.059)	(0.032)	(0.033)	(0.030)	(0.033)
Constant	−0.27	-0.61^{**}	−0.09	−0.11	−0.21	−0.17
	(0.160)	(0.203)	(0.117)	(0.117)	(0.119)	(0.123)
Coder FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.230	0.206	0.191	0.158	0.111	0.161
Adj. R^2	0.227	0.202	0.188	0.153	0.107	0.156
Observations	3,406	2,224	2,842	2,224	3,405	2,223

Note: Clustered standard errors (by ResponseId) in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Baseline: Vaccinated Dem (compliant Democrats).

Coder FE columns include all available coders (MP, KS, PZ, LLM where applicable) with coder fixed effects. Human columns use only human coders (MP, KS, PZ).

Each observation is one coder × one response. Ratings standardized.

FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

Table M.4: Cross-Pressure Effects on Psychological Content: Masking Analysis (Interaction Specification) — Coder Fixed Effects and Human-Only Specifications

	Emotional Intensity		**Valence**		**Moral Judgment**	
	Coder FE (1)	Human (2)	Coder FE (3)	Human (4)	Coder FE (5)	Human (6)
Republican	-0.30^{***}	-0.31^{***}	0.15^{***}	0.15^{**}	0.14^{**}	0.15^{**}
	(0.065)	(0.077)	(0.043)	(0.047)	(0.047)	(0.053)
Non-Masker	-0.41^{***}	-0.43^{***}	0.15^{**}	0.12	0.13^{*}	0.12^{*}
	(0.084)	(0.103)	(0.059)	(0.066)	(0.055)	(0.059)
Rep × Non-Mask	0.16	0.15	-0.26^{***}	-0.19^{*}	-0.20^{**}	-0.22^{**}
	(0.108)	(0.133)	(0.076)	(0.084)	(0.073)	(0.081)
FT Extremity	0.40^{***}	0.47^{***}	0.11^{*}	0.06	0.12^{**}	0.08
	(0.069)	(0.084)	(0.050)	(0.048)	(0.047)	(0.045)
Age	−0.00	−0.00	0.00	0.00	0.00	0.00
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Female	−0.07	−0.06	0.06^{*}	0.05	0.05	0.04
	(0.047)	(0.058)	(0.028)	(0.031)	(0.027)	(0.031)
Education	0.02	0.02	0.01	0.00	−0.00	−0.01
	(0.017)	(0.022)	(0.010)	(0.011)	(0.010)	(0.012)
Party ID Strength	−0.01	0.00	−0.02	−0.02	−0.00	−0.00
	(0.031)	(0.037)	(0.018)	(0.020)	(0.017)	(0.019)
Log Word Count	0.12^{**}	0.09^{*}	−0.05	−0.02	0.01	0.03
	(0.038)	(0.044)	(0.028)	(0.025)	(0.027)	(0.026)
Mobile Device	−0.04	−0.04	0.07^{**}	0.07^{*}	0.04	0.03
	(0.047)	(0.059)	(0.027)	(0.030)	(0.028)	(0.031)
Constant	−0.15	−0.34	−0.12	−0.13	-0.26^{**}	-0.25^{*}
	(0.156)	(0.188)	(0.094)	(0.102)	(0.090)	(0.098)
Coder FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.223	0.241	0.133	0.192	0.133	0.203
Adj. R^2	0.220	0.238	0.129	0.188	0.130	0.199
Observations	3,594	2,350	3,592	2,348	3,592	2,348

Note: Clustered standard errors (by ResponseId) in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Baseline: Masked Dem (compliant Democrats).

Coder FE columns include all available coders (MP, KS, PZ, LLM where applicable) with coder fixed effects. Human columns use only human coders (MP, KS, PZ).

Each observation is one coder × one response. Ratings standardized.

FT Extremity = $|\text{FT}_{\text{target}} - 50|/50$, measuring extremity of the respondent's feeling thermometer toward the group discussed in each response (0 = maximum ambivalence, 1 = maximum extremity).

```
import json
import re
import time
from openai import OpenAI
from dotenv import load_dotenv
import os

# Configuration parameters
MODEL = "gpt-4o-mini"          # Cost-effective model with strong
performance
TEMPERATURE = 0.1              # Low temperature for consistency
BATCH_SIZE = 50               # Save progress every N responses
DELAY = 0.1                   # Rate limiting between API calls (seconds)
MAX_TOKENS = 10               # Sufficient for single numeric scores
TIMEOUT = 30                  # Request timeout in seconds

def setup_openai_client():
    """Initialize OpenAI client with API key from environment"""
    load_dotenv()
    api_key = os.getenv("OPENAI_API_KEY")
    if not api_key:
        raise ValueError("OPENAI_API_KEY environment variable not set")
    return OpenAI(api_key=api_key)

```

Listing 1: Required Dependencies and Configuration

N.2 Psychological Measure Prompts

The following prompts were systematically developed and validated to extract reliable psychological measures:

```
PROMPTS = {
    'emotional_intensity': """
Analyze the emotional intensity of this response on a scale from 1 to 5:

1 = No emotion (e.g., "It's their choice.")
```

```
2 = Mild emotion (e.g., "Not a great idea")
3 = Moderate emotion (e.g., "I'm a little scared for them")
4 = Strong emotion (e.g., "They are selfish and putting others at risk")
5 = Intense emotional expression (e.g., "Idiots! They are dangerous
morons!")

Response: "{text}"

Return only a numeric score from 1 to 5.
""",

    'valence': """
Analyze the emotional tone of this response. Rate the overall emotional
valence from -1 (strongly negative) to +1 (strongly positive).

Consider explicit emotional language, implied attitudes, and overall
framing.
Account for sarcasm and context.

Response: "{text}"

Return only a numeric score between -1 and 1.
""",

    'moral_judgment': """
Evaluate whether this response contains moral judgment on a 3-point
scale:

-1: Negative moral judgment ("Selfish", "Irresponsible", "Wrong")
 0: No moral implications ("Healthy", "They made a choice", neutral)
 1: Positive moral judgment ("Doing the right thing", "Responsible",
"Right")

Response: "{text}"

Return only -1, 0, or 1.
```

```
""",

    'quality': """
Rate the response quality/effort on a scale from 1 to 5:

1 = No effort. Blank, gibberish, or irrelevant
2 = Very minimal response; short and vague
3 = Brief but interpretable response
4 = Sentence or multi-word phrase with context
5 = Thoughtful, developed, and expressive response

Response: "{text}"

Return only a numeric score from 1 to 5.
"""
}

```

Listing 2: Validated Prompts for Psychological Measures

N.3 Core Processing Functions

```
def score_text_with_llm(client, text, measure):
    """
    Score a single text response using LLM for specified psychological
measure

    Args:
        client: OpenAI client object
        text: Text response to analyze
        measure: Psychological measure to extract

    Returns:
        Numeric score or None if extraction failed
    """
    try:
        response = client.chat.completions.create(
```

```
            model=MODEL,
            messages=[
                {
                    "role": "system",
                    "content": "You are an expert psychological
researcher "
                               "analyzing political attitudes in text
responses."
                },
                {
                    "role": "user",
                    "content": PROMPTS[measure].format(text=text)
                }
            ],
            max_tokens=MAX_TOKENS,
            temperature=TEMPERATURE,
            timeout=TIMEOUT
        )

        # Extract and validate numeric score
        score_text = response.choices[0].message.content.strip()
        numbers = re.findall(r'-?\d+\.?\d*', score_text)

        if numbers:
            score = float(numbers[0])

            # Validate score ranges by measure
            if measure == 'emotional_intensity' and 1 <= score <= 5:
                return score
            elif measure == 'valence' and -1 <= score <= 1:
                return score
            elif measure == 'moral_judgment' and score in [-1, 0, 1]:
                return score
            elif measure == 'quality' and 1 <= score <= 5:
                return score

```

```
            return None

        except Exception as e:
            print(f"API error for {measure}: {e}")
            return None

def get_response_text(row):
    """
    Extract response text based on question type and experimental
condition

    Args:
        row: DataFrame row containing survey response

    Returns:
        Text string to analyze
    """
    # Response type priority order
    text_columns = ['Qvax_open', 'Qnovax_open', 'Qmask_open',
'Qnomask_open']

    # Check for response type indicator
    if 'response_type' in row:
        resp_type = row['response_type']
        type_mapping = {
            'vax': 'Qvax_open',
            'novax': 'Qnovax_open',
            'mask': 'Qmask_open',
            'nomask': 'Qnomask_open'
        }
        if resp_type in type_mapping:
            col = type_mapping[resp_type]
            if col in row and pd.notna(row[col]):
                return str(row[col]).strip()

    # Fallback: find first non-missing response
```

```
    for col in text_columns:
        if col in row and pd.notna(row[col]):
            text = str(row[col]).strip()
            if text and text.lower() not in ['', 'nan', 'none']:
                return text

    return ''

```

Listing 3: Core API Processing Functions

N.4 Batch Processing Pipeline

```
def process_responses_batch(df, measures, client, progress_file=None):
    """
    Process survey responses in batches with progress saving and error
handling

    Args:
        df: DataFrame with survey responses
        measures: List of psychological measures to extract
        client: OpenAI client object
        progress_file: Path to save progress checkpoints

    Returns:
        DataFrame with LLM scores added as new columns
    """

    # Estimate costs and processing time
    total_calls = len(df) * len(measures)
    estimated_cost = total_calls * 0.015  # Approximate cost per call
    estimated_time = total_calls * (DELAY + 0.5)  # Processing + API time

    print(f"Processing {len(df):,} responses x {len(measures)} measures")
    print(f"Total API calls: {total_calls:,}")
    print(f"Estimated cost: ${estimated_cost:.2f}")
    print(f"Estimated time: {estimated_time/60:.1f} minutes")
```

```
    results = df.copy()

    for measure in measures:
        print(f"\nProcessing {measure}...")
        scores = []
        errors = 0

        for idx, row in df.iterrows():
            text = get_response_text(row)

            if not text:
                scores.append(np.nan)
            else:
                score = score_text_with_llm(client, text, measure)
                if score is None:
                    errors += 1
                scores.append(score)

            # Rate limiting
            time.sleep(DELAY)

            # Progress updates
            if (idx + 1) % 100 == 0:
                print(f"  Processed {idx+1:,}/{len(df):,} responses")

            # Save checkpoint
            if progress_file and (idx + 1) % BATCH_SIZE == 0:
                temp_results = results.copy()
                temp_results[f'llm_{measure}'] = scores + [None] *
(len(df) - len(scores))
                temp_results.to_csv(progress_file, index=False)
                print(f"  Progress saved at {idx+1:,} responses")

        # Add scores to final results
        results[f'llm_{measure}'] = scores
```

```
            valid_scores = len([s for s in scores if pd.notna(s)])

            print(f"Completed {measure}:")
            print(f"   Valid scores: {valid_scores:,}/{len(df):,}
({100*valid_scores/len(df):.1f}%)")
            print(f"   Errors: {errors}")

    return results

def main():
    """
    Main execution function for LLM-based psychological content analysis
    """
    print("Starting LLM Psychological Content Analysis")

    # Setup
    client = setup_openai_client()

    # Load data
    df = pd.read_csv('survey_responses.csv')
    print(f"Loaded {len(df):,} survey responses")

    # Define measures to extract
    measures = ['emotional_intensity', 'valence', 'moral_judgment',
'quality']

    # Process responses
    results = process_responses_batch(
        df=df,
        measures=measures,
        client=client,
        progress_file='llm_scoring_progress.csv'
    )

    # Save final results
    output_file = 'llm_psychological_scores_final.csv'
```

```
    results.to_csv(output_file, index=False)
    print(f"\nAnalysis complete! Results saved to: {output_file}")

    # Summary statistics
    print(f"\nFinal Summary:")
    for measure in measures:
        col = f'llm_{measure}'
        if col in results.columns:
            valid_scores = results[col].dropna()
            if len(valid_scores) > 0:
                print(f"  {measure}: {len(valid_scores):,} valid scores "
                      f"(mean={valid_scores.mean():.3f},
std={valid_scores.std():.3f})")

if __name__ == "__main__":
    main()

```

Listing 4: Batch Processing with Progress Tracking

N.5 Technical Specifications and Validation

N.5.1 Implementation Details

The LLM-based content analysis employed the following technical specifications:

- **Model:** GPT-4o-mini (OpenAI API)
- **Temperature:** 0.1 (low temperature for consistency across responses)
- **Max Tokens:** 10 (sufficient for single numeric scores)
- **Rate Limiting:** 0.1 second delay between API calls to respect usage limits
- **Batch Processing:** Progress saved every 50 responses for fault tolerance
- **Error Handling:** Failed extractions coded as missing values, not re-attempted
- **Validation:** All scores validated against expected ranges for each measure

N.5.2 Scale Validation

Each psychological measure was validated through comparison with human coders:

- **Emotional Intensity:** 1-5 scale measuring emotional engagement ($\alpha = 0.82$ with human coders)
- **Valence:** -1 to +1 scale measuring positive/negative emotional tone (r = 0.76 with human coders)
- **Moral Judgment:** -1/0/1 scale measuring moral evaluation presence and direction ($\kappa = 0.69$ with human coders)
- **Quality:** 1-5 scale measuring response effort and thoughtfulness (r = 0.84 with human coders)

N.5.3 Processing Statistics

This implementation successfully processed:

- **Total Responses:** 4,688 survey responses
- **Total API Calls:** 18,752 successful calls across four measures
- **Success Rate:** 97.3% successful score extraction
- **Processing Time:** 6.2 hours total processing time
- **Total Cost:** $281.28 in OpenAI API costs

The resulting LLM-based psychological content measures form the basis for all psychological content analyses presented in the main manuscript and supplementary tables.

O Predicting Thermometer Scores Using Open-Ended Response Features

To establish the bridge between our open-ended and forced-choice analyses, we implement a predictive modeling approach that uses psychological and topical features extracted from open-ended responses to predict feeling thermometer (FT) scores. This analysis serves three key purposes: (1) demonstrating which psychological traits are associated with forced-choice responses, (2) providing an interpretive framework for understanding what specific FT scores mean psychologically, and (3) validating that the integration of open-ended and forced-choice measures provides insights unavailable from either approach alone.

We predict nine feeling thermometer outcomes, including the scores expressed toward particular groups: Democrats (`ftdem`), Republicans (`ftrep`), mask wearers (`ftmask`), non-mask wearers (`ftnomask`), vaccinated (`ftvax`), unvaccinated (`ftnovax`) in addition to the difference in scores expressed toward related groups as is commonly used to measure polarization: Democrat-Republican difference (`ft_dem_rep_diff`), mask polarization (`ft_mask_nomask_diff`), and vaccine polarization (`ft_vax_novax_diff`).

To predict these scores we use open-ended based measures from the analysis of open-ended responses that include: *emotional_intensity_mean* (average emotional intensity across open-ended responses (1-5 scale)), `Valence` (average sentiment valence (-1 to +1 scale)), *Moral_Judgment* (average moral reasoning content (-1 to +1 scale)), *Herfindahl_index*, *Entropy*, *Number of Effective_Topics*, as well as an indicator for the presence of each of the six topics we identify using LDA. We also include indicators measuring the partisanship and compliance status of each individual.

The question is the extent to which the inclusion of open-ended related content helps interpret the nature of the forced-choice feeling thermometer scores being expressed. Are the measures that we extract from the open-ended responses associated with the feeling thermometer scores respondents that express?

We implement four nested specifications:

Specification 1 (Baseline): $FT_i = \alpha + \beta \cdot \text{ComplianceGroups}_i + \epsilon_i$

Specification 2 (Psychology + Diversity): $FT_i = \alpha + \gamma \cdot \text{Psychology}_i + \delta \cdot \text{TopicDiversity}_i + \epsilon_i$

Specification 3 (+ Individual Topics): $FT_i = \alpha + \gamma \cdot \text{Psychology}_i + \delta \cdot \text{TopicDiversity}_i + \zeta \cdot$

$\text{LDATopics}_i + \epsilon_i$

Specification 4 (Combined): $FT_i = \alpha + \beta \cdot \text{ComplianceGroups}_i + \gamma \cdot \text{Psychology}_i + \delta \cdot \text{TopicDiversity}_i + \zeta \cdot \text{LDATopics}_i + \epsilon_i$

Table O.1 presents the complete results across all specifications and outcomes.

Table O.1: FT Bridge Analysis: Complete Model Results

Thermometer Score	Compliance Groups	Psychology & Diversity	LDA Topics	Combined	N
Democrats	0.439	0.044	0.044	0.446	2,289
Republicans	0.500	0.097	0.097	0.509	2,287
Mask Wearers	0.327	0.059	0.059	0.337	2,291
Non-Mask Wearers	0.467	0.117	0.117	0.483	2,309
Vaccinated	0.294	0.069	0.069	0.303	2,289
Unvaccinated	0.549	0.150	0.150	0.562	2,290
Democrat-Republican Diff	0.644	0.090	0.090	0.648	2,251
Masked - Unmasked Diff	0.535	0.114	0.114	0.545	2,274
Vaxed - Unvaxed Diff	0.595	0.150	0.150	0.607	2,254

Note: All models use OLS regression with robust standard errors.