

CIRCLES OF TRUST: A RELATIONAL APPROACH TO SOCIAL DESIRABILITY BIAS IN POLITICAL POLLING

Círculos de confianza: un enfoque relacional del sesgo de deseabilidad social en encuestas políticas

Círculos de confiança: uma abordagem relacional ao viés de deseabilidade social em pesquisas políticas

LUIS HERRERO-CORONA  lherrero@lac.com.mx ¹

¹ LAC Intermarketing

Submission: 2025-10-15

Accepted: 2026-04-22

Publication: 2026-05-27

Keywords:

social desirability bias; political polling; survey methodology; trust; Mexico

Abstract

Traditional approaches treat social desirability bias (SDB) in political polling as a stable individual trait. This study challenges this assumption by introducing the “circles of trust” framework, which conceptualizes SDB as a contextual variable dependent on the relational distance between respondents and their interlocutors. Using data from a national telephone survey in Mexico (n = 1,008), we measured respondents’ comfort in expressing political opinions across five social contexts: partners, family, friends, work colleagues, and survey interviewers. The results reveal an unexpected pattern: while comfort decreases from intimate relationships (partners: 80.5 %, family: 80 %) to broader social contexts (friends: 73 %, colleagues: 63 %), survey interviewers occupy a distinct position (78.5 %) characterized by higher comfort than friends or colleagues. This “pollster advantage” appears to be linked to perceived professional neutrality and anonymity. Crucially, institutional trust in elections strongly predicts both comfort with interviewers and self-reported response authenticity, suggesting that deficits in democratic legitimacy may systematically bias polling data. These findings have methodological implications for survey design in polarized democracies.

Palabras clave:
*sesgo de
deseabilidad
social; encuestas
políticas;
metodología
de encuestas;
confianza
institucional;
México*

Resumen

Los enfoques tradicionales tratan el sesgo de deseabilidad social (SDS) en encuestas políticas como un rasgo individual estable. Este estudio cuestiona dicha premisa mediante el marco de “círculos de confianza”, que conceptualiza el SDS como contextualmente variable, dependiente de la distancia relacional entre encuestados e interlocutores. A partir de una encuesta telefónica nacional en México (n = 1,008), medimos la comodidad de los encuestados para expresar opiniones políticas en cinco contextos sociales: pareja, familia, amigos, colegas de trabajo y encuestadores. Los resultados revelan un patrón inesperado: mientras la comodidad disminuye desde las relaciones íntimas (pareja: 80.5 %, familia: 80 %) hacia contextos sociales más amplios (amigos: 73 %, colegas: 63 %), los encuestadores ocupan una posición diferenciada (78.5 %), con mayor comodidad que amigos o colegas. Esta “ventaja del encuestador” parece vinculada a la percepción de neutralidad profesional y anonimato. De manera significativa, la confianza institucional en las elecciones predice tanto la comodidad con los encuestadores como la autenticidad autorreportada de las respuestas, lo que sugiere que los déficits de legitimidad democrática pueden sesgar sistemáticamente los datos de las encuestas. Estos hallazgos tienen implicaciones metodológicas para el diseño de encuestas en democracias polarizadas.

Palavras-chave:
*viés de
desejabilidade
social; pesquisas
políticas;
metodologia
de pesquisa;
confiança
institucional;
México*

Resumo

As abordagens tradicionais tratam o viés de desejabilidade social (VDS) em pesquisas políticas como um traço individual estável. Este estudo questiona essa premissa por meio do referencial dos “círculos de confiança”, que conceitua o VDS como uma variável contextual dependente da distância relacional entre respondentes e seus interlocutores. A partir de uma pesquisa telefônica nacional realizada no México (n = 1.008), medimos o conforto dos respondentes para expressar opiniões políticas em cinco contextos sociais: parceiros, família, amigos, colegas de trabalho e entrevistadores de pesquisa. Os resultados revelam um padrão inesperado: embora o conforto diminua das relações íntimas (parceiros: 80,5 %; família: 80 %) para contextos sociais mais amplos (amigos: 73 %; colegas: 63 %), os entrevistadores ocupam uma posição diferenciada (78,5 %), com maior conforto do que amigos ou colegas. Essa “vantagem do entrevistador” parece estar vinculada à percepção de neutralidade profissional e ao anonimato. De forma relevante, a confiança institucional nas eleições prediz tanto o conforto com os entrevistadores quanto a autenticidade autorrelatada das respostas, o que sugere que déficits de legitimidade democrática podem enviesar sistematicamente os dados das pesquisas. Esses achados têm implicações metodológicas para o desenho de pesquisas em democracias polarizadas.

1. INTRODUCTION

Political polling faces a persistent methodological challenge: respondents do not always report their true preferences. Social desirability bias (SDB)—the tendency to alter responses to align with perceived social expectations—has been

documented across electoral contexts, from underreporting support for controversial candidates to inflated turnout intentions (Claassen & Ryan, 2024; Holbrook & Krosnick, 2010). However, research typically treats SDB as a stable individual characteristic, overlooking a fundamental insight: the same person may respond authentically in one context while concealing preferences in another.

This study asks: “How does the social context of political communication shape respondents’ willingness to express authentic opinions?”. We propose that SDB in survey research is better understood through a relational lens, one that recognizes the interviewer–respondent interaction as just one node in a broader network of social relationships characterized by varying levels of trust, intimacy, and perceived risk.

To address this question, we introduce the circles of trust framework. Drawing on Fukuyama’s (2000) concept of the “radius of trust” and Delhey *et al.*’s (2011) cross-national operationalization, we conceptualize political expression as embedded within concentric social circles—from intimate partners and family to friends, colleagues, and institutional representatives such as survey interviewers. Each circle entails different social dynamics, power relationships, and potential consequences of political disclosure. By mapping comfort levels across these circles, we can identify where authentic expression becomes constrained and, critically, where the survey interview fits within respondents’ trust hierarchies.

Mexico provides an ideal case study to test this framework. As a democracy characterized by significant polarization, contested institutional legitimacy, and documented challenges with polling accuracy (Blanco, 2013; Frisbie, 2024), Mexico exemplifies the conditions under which relational trust dynamics are likely to shape survey validity. The 2024 electoral context, marked by debates over judicial reform and heightened political tensions, further amplifies these dynamics.

Our findings contribute to survey methodology in three ways. First, we document that comfort with survey interviewers occupies a distinct and surprisingly favorable position in respondents’ trust hierarchies, suggesting that professional neutrality creates conditions for disclosure that differ from everyday social interactions. Second, we demonstrate that institutional trust in democratic processes predicts authentic political expression, not only with interviewers but across all social circles, indicating that legitimacy deficits cascade through social relationships. Third, we identify systematic demographic variations in comfort patterns that conventional weighting procedures may not adequately address.

2. LITERATURE REVIEW

2.1. *Social Desirability Bias in Political Polling*

Social desirability bias represents one of the most persistent challenges in political survey research, particularly in contexts in which political preferences may be stigmatized or controversial. The foundational work by Edwards (1957) and later developments by Paulhus (1984) established a theoretical framework for understanding how respondents alter their responses to conform to perceived social norms. In the political domain, this phenomenon has manifested in various forms, from the underreporting of support for controversial candidates to the overestimation of socially approved political behaviors, such as voting turnout (Holbrook & Krosnick, 2010).

Recent studies have documented how social desirability bias can systematically distort polling results in polarized political environments. Claassen and Ryan (2024) demonstrated that survey respondents experience measurable pressure when revealing their political preferences, with this pressure varying significantly based on the perceived social acceptability of their choices. Their experimental research, using best-impression and worst-impression techniques, revealed that traditional polling methods may systematically underestimate support for socially stigmatized candidates or positions.

The interaction between respondent perceptions of survey sponsorship and response authenticity was documented in Bischooping and Schuman's (1992) seminal analysis of the 1990 Nicaraguan preelection polls. Their study revealed that in highly polarized contexts, respondents' willingness to express true voting intentions depended critically on their perception of the poll's partisan alignment. Polls perceived as neutral or government-affiliated systematically underestimated opposition support, whereas those perceived as opposition-aligned produced more accurate predictions. This paradoxical finding—that apparent interviewer bias was necessary to reduce respondent bias—underscores how social desirability pressures in polarized democracies operate through relational trust dynamics rather than as stable individual traits. The Nicaraguan case provides an important precedent for understanding political polling in contemporary Mexico, where similar dynamics of institutional distrust and political polarization may shape respondents' decisions about authentic self-expression in survey contexts.

The "Bradley effect" exemplifies how social desirability bias operates in racially charged electoral contexts (Streb *et al.*, 2008), with similar patterns observed across gender, religion, and ideological contexts, representing a fundamental challenge to democratic representation.

Larson (2019) provided comprehensive evidence that failing to account for social desirability bias can lead to misleading conclusions about the relationships

between demographic variables and political attitudes. His research demonstrated that including measures of social desirability bias in regression analyses can change the statistical significance of demographic variables by as much as 100 %, fundamentally altering the interpretation of survey results. This finding has profound implications for political polling, where an accurate understanding of demographic patterns is essential for electoral predictions and campaign strategy.

Recent research has further demonstrated that polling errors are not merely random occurrences but follow systematic patterns that can be predicted based on electoral and contextual factors. Chen *et al.* (2023) analyzed electoral predictors of polling errors and found that systematic biases in survey responses are associated with specific political and social conditions that vary across elections and contexts. Their findings suggest that social desirability bias and related response patterns may be particularly pronounced during certain types of electoral campaigns or in specific political environments, indicating that the challenges documented in our Mexican context may reflect broader patterns of how political polarization and institutional trust interact to shape survey accuracy in democratic societies.

2.2. Circles of Trust and Social Communication

The role of interpersonal trust in political communication has received increasing attention, as scholars recognize that political expression is fundamentally social and relational. Granovetter's (1973) seminal work on social network theory distinguished between "strong ties" characterized by frequent interaction, emotional closeness, and mutual trust, and "weak ties" that connect individuals across different social groups. This distinction has important implications for political communication, as individuals may express different levels of authenticity depending on the strength and nature of their social ties.

Research on political discussion networks has consistently demonstrated that the composition and characteristics of social networks significantly influence political attitudes and behaviors (Huckfeldt & Sprague, 1995). Individuals embedded in politically homogeneous networks tend to express more extreme political views and demonstrate greater confidence in their political judgments, whereas those in diverse networks show more moderate positions and greater uncertainty. These findings suggest that the social context of political expression is crucial for understanding both the formation and expression of political preferences.

Recent research by Belardinelli and Gili (2022) has highlighted how circles of trust operate at different levels of intimacy and formality, ranging from close family relationships characterized by "familiarity and mutual trust" to more distant relationships based on "weaker and more open ties". This conceptualization parallels our framework of circles of trust, suggesting that individuals navigate multiple

concentric layers of social relationships, each characterized by different expectations for authentic self-expression.

2.3. Institutional Trust in Latin America and Mexico

Institutional trust in Latin America has been characterized by chronically low levels and significant volatility, reflecting the region's complex political history and ongoing challenges with democratic consolidation. According to data from the Latin American Public Opinion Project (LAPOP), trust in key democratic institutions such as the police, judiciary, and electoral authorities remains consistently below the levels observed in established democracies (De Micheli & Taylor, 2024).

Mexico presents a complex case of institutional trust shaped by its democratic transition and challenges with corruption and violence. Research using LAPOP data has demonstrated that insecurity and crime victimization significantly reduce trust in Mexican institutions (Blanco, 2013), with implications for survey methodology, as skepticism about institutional legitimacy may compromise authentic responses.

The United Nations Development Programme (2024) reported that trust in Latin American governments declined significantly between 2009 and 2018, reaching levels where only one in five people expressed trust in their government. However, the period from 2018 to 2020 saw some improvements, particularly in countries that elected outsiders from traditional party systems, suggesting that institutional trust may be responsive to political change and renewal. In Mexico, this period coincided with the election of Andrés Manuel López Obrador, whose populist appeals and promises to combat corruption resonated with voters seeking alternatives to traditional political elites.

Maldonado and Crow's (2013) experimental research on institutional trust measurement in Mexico revealed important methodological insights relevant to our study. Their comparison of 4-point and 7-point trust scales demonstrated that measurement instruments can significantly affect reported levels of institutional trust, with implications for how different demographic groups evaluate political institutions. This finding underscores the importance of considering methodological factors when interpreting survey data on institutional trust and political attitudes.

2.4. The Role of Social Context in Political Expression

Recent research has increasingly recognized that political expression is not merely an individual choice but is fundamentally shaped by social context and interpersonal relationships. Yuen and Lee's (2025) study of self-censorship under democratic backsliding provides crucial insights into how social influence affects

citizens' willingness to express political opinions. Their research in Hong Kong demonstrated that individuals are more likely to engage in self-censorship when they receive cautionary reminders from family members and peers or when they observe self-censoring behavior among their acquaintances.

Miller and Mitamura's (2003) analysis of trust measurements across different cultural contexts highlights how social and cultural factors can influence the validity of survey responses. Their research suggests that ambiguous wording in survey questions about trust can lead to misinterpretations, particularly in cross-cultural contexts where social norms around political expression may vary significantly. This finding has important implications for political polling in diverse societies, such as Mexico, where regional, ethnic, and socioeconomic differences may create varying expectations for political communication.

Research on survey participation itself provides additional insights into how trust and social context influence political expression. Silber *et al.* (2022) demonstrated that democratic engagement and trust in institutions significantly affect citizens' willingness to participate in surveys, with implications for the representativeness of polling data. Their findings suggest that certain demographic groups may systematically underparticipate in political surveys because of trust-related factors, creating potential biases in the resulting data.

2.5. Theoretical Framework: Circles of Trust

Building on this literature, we propose the "circles of trust" framework as a comprehensive approach to understanding how social context shapes political expression in survey research. This framework recognizes that individuals operate within multiple concentric social networks characterized by different levels of interpersonal trust, intimacy, and perceived consequences for authentic political expression.

The circles of trust framework builds upon foundational work in social capital theory that distinguishes between the *level* and *radius* of trust. Fukuyama (2000) conceptualized the radius of trust as "the circle of people among whom cooperative norms are operative" (p. 4), noting that societies can be understood as consisting of concentric and overlapping radii of trust ranging from intimate friends to formal institutions. This distinction is critical: a society may exhibit high trust levels within narrow circles (such as family) while maintaining low trust toward outsiders, resulting in limited generalized cooperation.

Delhey *et al.* (2011) operationalized this concept empirically, demonstrating that the radius of trust varies substantially across countries and cultural contexts. Their analysis of 51 countries revealed that in Latin America, including Mexico, the trust radius tends to be narrower than that in Protestant or wealthy nations, with

social capital residing largely within families and close personal networks. This finding has direct implications for survey methodology: when respondents imagine “most people” in trust questions, their mental reference may be constrained to in-group members rather than extending to unfamiliar others such as survey interviewers. Our circles of trust framework extends this insight by explicitly measuring comfort levels across distinct social contexts, allowing us to map the boundaries where authentic political expression becomes constrained.

The innermost circle consists of intimate relationships, such as romantic partners and close family members, where high levels of trust and emotional closeness create conditions conducive to authentic political expression. Subsequent circles include friends, work colleagues, and formal relationships with institutional representatives such as survey interviewers. Each circle is characterized by different social dynamics, power relationships, and potential risks associated with political disclosure.

This framework addresses critical limitations in existing approaches to social desirability bias by recognizing the relational and contextual nature of political expression. Rather than treating social desirability as a stable individual trait, the circles of trust approach acknowledges that the same individual may demonstrate different levels of authenticity, depending on their audience and the specific social context of political communication.

Furthermore, this framework is particularly relevant for understanding political polling in polarized democracies where institutional trust is contested or declining. In such contexts, the outer circles of trust—particularly interactions with institutional representatives, such as survey interviewers—may be especially susceptible to social desirability pressures and self-censorship. By mapping the patterns of trust and comfort across different social circles, we can better understand how broader political and social contexts shape individual decisions regarding political communication and survey participation.

3. METHODS

3.1. *Research Design*

This study employed a cross-sectional survey design to examine the relationship between circles of trust and social desirability bias in political polling. The research used a national telephone survey conducted in Mexico in November 2024, following standard protocols for representative public opinion research. The study was designed to test the circles of trust framework by measuring respondents’ comfort levels in expressing political opinions across different social contexts while controlling for institutional trust and demographic factors.

The research design incorporated multiple measurement approaches to capture the complexity of trust relationships and their impact on authentic political expressions. Primary measures included direct assessments of comfort levels across different social circles, self-reported authenticity in survey responses, perceptions of social desirability pressures, and institutional trust indicators. This multidimensional approach allowed for a comprehensive analysis of how interpersonal and institutional trust factors interact to influence political communication.

3.2. Sample and Sampling Procedures

The study employed Random Digit Dialing (RDD) methodology to survey Mexican adults aged 18 and older via landline telephones. The sampling frame was constructed using Mexico's telephone dialing codes provided by the Federal Telecommunications Institute (IFT), with the RDD list generated to reflect the distribution of phone lines across states, which closely mirrors the population distribution across Mexico's territories.

The target sample size of 1,000 respondents was exceeded, and 1,008 completed interviews were achieved. Post-stratification weights were applied to adjust the sample distribution to match Mexican census parameters for sex and age groups, ensuring national representativeness. All analyses reported in this study used weighted data, unless otherwise specified.

While Random Digit Dialing provides a quasi-probability sampling frame, the use of predictive dialing to filter inactive numbers introduces minor deviations from strict probability sampling. Nevertheless, the distribution of telephone lines across geographical areas provides a reasonable representation of Mexico's demographic structure.

The interviews were conducted by 40 trained professional interviewers using CATI technology. Training was conducted via video conference with supervisor monitoring throughout fieldwork. The Spanish-language survey averaged 10 minutes per interview. Post-fieldwork procedures included data file preparation with appropriate labeling.

3.3. Measures and Instrumentation

Circles of Trust Assessment

The primary dependent variables measured comfort levels for expressing political opinions across five distinct social circles. Respondents were asked: "En general, ¿qué tan cómodo se siente compartiendo sus opiniones políticas con...?"

("In general, how comfortable do you feel sharing your political opinions with...?") in the following contexts: Su pareja (Your partner), Su familia (Your family), Sus amigos (Your friends), Compañeros de trabajo (Work colleagues), and Un(a) encuestador(a) (A pollster/interviewer).

Responses were measured on a 4-point Likert scale: (1) Muy cómodo (Very comfortable), (2) Algo cómodo (Somewhat comfortable), (3) Poco cómodo (A little comfortable), and (4) Nada cómodo (Not at all comfortable). For analytical purposes, responses were analyzed both as continuous variables (using the original 4-point scale) and dichotomized categories combining 'very comfortable' and 'somewhat comfortable' responses versus less comfortable responses for descriptive analysis.

Respondents could indicate "does not apply" for circles not relevant to their situation (e.g., respondents without partners or not currently employed), which accounts for the varying sample sizes across circles reported in Table 1.

Social Desirability and Response Authenticity Measures

Several questions assessed social desirability pressure and response authenticity. Perceptions of social desirability bias among others were assessed through a battery of questions asking: "¿Qué tan frecuentemente cree que otras personas cambian sus respuestas cuando contestan encuestas políticas por...?" ("How frequently do you think other people change their answers when responding to political surveys because of...?") followed by four specific motivations: Quedar bien con el encuestador (Making a good impression on the interviewer), Evitar problemas (Avoiding problems), No ser juzgados por sus opiniones (Not being judged for their opinions), and Proteger su privacidad (Protecting their privacy). Each item used a 4-point frequency scale: (1) Muy frecuentemente (Very frequently) to (4) Nunca (Never).

Response authenticity was measured by asking: "Cuando le preguntan su opinión sobre temas políticos en una encuesta, ¿qué tanto considera usted que sus respuestas reflejan su verdadera opinión?" ("When you are asked about your political views in a survey, how much do you consider that your responses reflect your true opinion?"). Response options ranged from (1) Totalmente (Completely) to (4) Nada (Not at all).

Institutional Trust Indicators

Institutional trust was measured through multiple indicators to capture both general democratic legitimacy and specific trust in electoral processes. Two items assessed democratic legitimacy using a 4-point agreement scale (strongly agree to

strongly disagree): “Mi país está gobernado por la voluntad del pueblo” (“My country is governed by the will of the people”) and “En general, las elecciones en mi país son libres y justas” (“In general, elections in my country are free and fair”).

Trust in electoral polling was measured with the question: “¿Qué tanta confianza tiene usted en las encuestas electorales?” (“How much confidence/trust do you have in electoral polls?”) using a 4-point scale from (1) Mucha confianza (Much confidence) to (4) Nada de confianza (No confidence). These measures were designed for the specific objectives of this study and differ from standardized institutional trust batteries used in regional surveys, such as LAPOP or Latinobarómetro. Our items emphasize perceived democratic legitimacy and trust in electoral processes rather than confidence in specific institutions, which aligns with the study’s interest in how trust relates to authentic political expression in survey contexts.

Demographic and Control Variables

The survey collected comprehensive demographic information, including age (grouped into six categories: 18-24, 25-34, 35-44, 45-54, 55-64, 65+), gender (male/female), education level (five categories from no education to graduate school), household income (five quintiles), employment status (six categories including full-time work, part-time work, unemployed, student, homemaker, and retired/disabled), and religious affiliation.

Additional variables measured in the survey included ethnic trust, perceived impact of U.S. foreign policy on Mexico, life orientation (worry a lot vs. take life as it comes), and perceived impact of mobile technology on quality of life. These variables are not analyzed in the present study, which focuses specifically on comfort with political expression across circles of trust and its relationship to institutional trust and perceived social desirability bias.

Data Collection Procedures

Data collection was conducted in November 2024 by a professional survey research organization with extensive experience in Mexican public opinion research. Interviewers underwent specific training on the survey instrument, with particular attention paid to the sensitive nature of political opinion questions and the importance of creating a comfortable interview environment to encourage honest responses.

The telephone survey protocol included multiple callback attempts to minimize non-response bias, with interviews scheduled at different times of day and days of the week to maximize reach across different demographic groups. Quality

assurance procedures included real-time monitoring by supervisors, verification callbacks for a subset of completed interviews, and systematic data cleaning and validation procedures.

The survey followed standard protocols for telephone public opinion research in Mexico. Interviewers identified themselves at the beginning of each call, stated the general purpose, and obtained verbal consent. Participation was voluntary: respondents could decline or terminate the interview at any point without consequences. Average interview duration was approximately 10 minutes.

3.4. *Statistical Analysis*

Data analysis employed three primary approaches: descriptive statistics (frequencies, cross-tabulations, means), Analysis of Variance (ANOVA) with Duncan's post-hoc tests for group comparisons, and Pearson correlations for bivariate associations.

Descriptive analyses examined the distribution of responses across the five circles of trust and demographic subgroups, with mean comfort levels calculated for each circle overall and by demographic categories. ANOVA was used to test for significant differences in comfort levels across demographic groups for each circle of trust; when significant main effects were detected, post-hoc comparisons were performed using Duncan's multiple range test to identify which groups differed. Pearson product-moment correlations examined the relationships between comfort levels across circles of trust, as well as associations with institutional trust measures, response authenticity indicators, and perceptions of social desirability bias. A chi-square test of independence examined the relationship between trust in electoral polls and self-reported response authenticity.

All analyses were conducted using weighted data to account for post-stratification adjustments, with unweighted sample sizes (n) reported for transparency. Statistical significance was evaluated at the $p < .05$ level, with effect sizes calculated where appropriate. Analyses were performed using SPSS version 28.0, with missing data handled through pairwise deletion (excluding cases analysis by analysis) to maximize the use of available data.

The analysis strategy recognized the hierarchical nature of the circles of trust framework by examining both individual circle characteristics and overall patterns across all five circles. Given the exploratory nature of this study, the analyses focused on establishing bivariate patterns and relationships. Future research should employ multivariate models to examine the independent contributions of demographic and institutional factors while controlling for potential confounders.

4. RESULTS

4.1. *Sample Characteristics*

The final unweighted sample consisted of 1,008 Mexican adults who were reached through landline telephone interviews. Post-stratification weights were applied based on the most recent national census data to adjust for differential coverage by sex and age groups; all percentages reported below reflect weighted estimates, unless otherwise noted.

After weighting, the estimated population distribution was 51.2 % females and 48.8 % males, with ages ranging from 18 to 65+ years. Educational attainment varied from no formal education (3.1 %) to a graduate degree (8.7 %), with the majority having completed secondary education (28.4 %) or higher. Employment status was distributed among full-time workers (41.2 %), part-time workers (15.3 %), students (12.8 %), homemakers (18.4 %), and retirees or individuals with disabilities (6.9 %). Household income was relatively evenly distributed across quintiles.

4.2. *Trust Gradients Across Social Circles*

Our analysis of the comfort levels for expressing political opinions revealed a clear and intriguing hierarchy, as shown in Table 1. As expected, respondents were most comfortable sharing their views with their intimate partners (80.5 %) and family (80.0 %), confirming the foundational role of close personal relationships in political communication. The comfort levels then began to decline, with friends (73.0 %) and work colleagues (63.0 %).

A central and unexpected finding emerges from this ordering: comfort with survey interviewers (78.5 %) exceeds that reported for both work colleagues (63.0 %) and friends (73.0 %). Rather than occupying the expected position at the outer edge of trust circles, interviewers appear to benefit from perceived professional neutrality and anonymity, thus creating distinct conditions for political disclosure. This unexpected result challenges the conventional wisdom that formal, external interactions inherently generate less comfort than informal, social ones. The data suggest that perceived anonymity and professional neutrality of the survey context may create a unique space for respondents to express opinions that they might otherwise withhold due to social pressures.

Table 1. Comfort Levels for Political Expression Across Five Circles of Trust

Trust Circle	n	M	SD	95 % CI	% Comf.	Very Comf.	Somewhat	A Little	Not At All
Partner	850	1.69	1.00	[1.63, 1.76]	80.5	60.0	20.5	10.0	9.5
Family	968	1.72	0.97	[1.66, 1.78]	80.0	56.0	24.0	12.0	8.0
Friends	934	1.95	1.03	[1.88, 2.02]	73.0	43.0	30.0	15.0	12.0
Work Colleagues	807	2.25	1.08	[2.17, 2.32]	63.0	31.0	32.0	19.0	18.0
Survey Interviewer	964	1.83	0.95	[1.77, 1.89]	78.5	46.0	32.5	13.0	8.5

Note. Trust circles are ordered from intimate to institutional according to the theoretical framework. Weighted data, 2024 National Telephone Survey of Mexican adults. Unweighted *n* values are reported. Question: “In general, how comfortable do you feel sharing your political opinions with the following people or groups?” Scale: 1 = Very comfortable, 2 = Somewhat comfortable, 3 = A little comfortable, 4 = Not at all comfortable. Lower means indicate greater comfort. Percentages are based on weighted data; “% Comf.” combines “Very comfortable” and “Somewhat comfortable” categories. CI = confidence interval.

4.3. Demographic Variations in Trust Patterns

Age effects demonstrate clear generational patterns, with younger adults (18-24) showing significantly higher comfort levels than older adults (65+) in partner relationships (81 % vs. 71 %) and friendships (83 % vs. 65 %), reflecting broader generational differences in political communication patterns ($F(5, 864) = 3.44, p = .004$ for partners; $F(5, 937) = 3.22, p = .007$ for friends). Educational gradients revealed substantial disparities between graduate-educated and non-educated respondents, particularly in partner relationships (89 % vs. 69 %) and survey contexts (86 % vs. 60 %), indicating that higher education facilitates political expression comfort in specific circles. Gender differences, while modest, showed consistent patterns with higher comfort among male respondents across all trust circles, with the most pronounced gap appearing in survey interviewer comfort ($t(967) = -2.99, p = .003$).

These demographic patterns suggest that traditional weighting procedures may inadequately address systematic biases in political polling. Standard post-stratification adjusts for demographic composition but cannot correct for within-group variation in the willingness to express authentic opinions. If lower-income respondents within a given age-gender cell are systematically less comfortable

than their higher-income counterparts, weighting by age and gender alone will not eliminate the resulting bias.

Table 2 presents the comprehensive ANOVA results for all three demographic factors across the five trust circles, revealing systematic patterns in how age, education, and income shape political expression comfort. These demographic effects demonstrate that trust circles are not uniform across the population but are significantly influenced by social position and life circumstances.

Table 2. Age, Education, and Income Effects on Comfort Across Trust Circles

Trust Circle	Age F (df)	p	Education F (df)	p	Income F (df)	p
Partner	3.44 (5, 864)	.004**	7.99 (4, 852)	< .001***	5.16 (4, 721)	< .001***
Family	0.83 (5, 968)	.532	1.46 (4, 955)	.211	2.79 (4, 795)	.026*
Friends	3.22 (5, 937)	.007**	3.08 (4, 924)	.015*	4.16 (4, 775)	.002**
Work Colleagues	0.95 (5, 847)	.450	0.82 (4, 836)	.513	1.97 (4, 713)	.098
Survey Interviewer	0.54 (5, 962)	.746	3.67 (4, 951)	.006**	8.78 (4, 791)	< .001***

Note. N = 1,008. Weighted data, 2024 National Telephone Survey of Mexican adults. One-way ANOVA with post-hoc Duncan tests. Comfort question: “In general, how comfortable do you feel sharing your political opinions with the following people or groups?” F-values with degrees of freedom in parentheses. * $p < .05$. ** $p < .01$. *** $p < .001$.

Age Effects

Age emerged as a significant predictor of comfort levels in some circles of trust, with generational variation evident for partners and friends. Post-hoc comparisons revealed clear patterns across demographic groups. For intimate relationships, younger adults aged 18-24 showed the highest comfort levels (81 % comfortable), while comfort decreased with age, reaching its lowest point among respondents aged 65+ (71 % comfortable). Similar age effects emerged for friends (83 % vs. 65 % for the youngest and oldest groups, respectively). In contrast, age did not significantly predict comfort with family ($F(5, 968) = 0.83, p = .532$), work colleagues ($F(5, 847) = 0.95, p = .450$), or survey interviewers ($F(5, 962) = 0.54, p = .746$), suggesting that formal polling contexts may minimize age-related variations in political expression comfort.

Educational Attainment Effects

Educational effects were strongest for Partner ($F(4, 852) = 7.99, p < .001$), Friends ($F(4, 924) = 3.08, p = .015$), and Survey Interviewer ($F(4, 951) = 3.67, p = .006$). The gradient was particularly pronounced for intimate relationships, where comfort with partners increased from 69 % among those with no formal education to 89 % among respondents with graduate education. Graduate-educated respondents also showed 86 % comfort with survey interviewers, compared to 60 % among those with no formal education. Family and work colleague contexts did not show significant educational gradients (Family: $F(4, 955) = 1.46, p = .211$; Colleagues: $F(4, 836) = 0.82, p = .513$), suggesting that educational effects concentrate in intimate relationships and formal survey contexts rather than being uniformly distributed across all social circles.

Income and Social Class Effects

Household income demonstrated significant associations with political expression comfort in most circles. Analysis of variance revealed significant income effects for partners ($F(4, 721) = 5.16, p < .001$), friends ($F(4, 775) = 4.16, p = .002$), family ($F(4, 795) = 2.79, p = .026$), and survey interviewers ($F(4, 791) = 8.78, p < .001$). Income effects for work colleagues did not reach statistical significance ($F(4, 713) = 1.97, p = .098$), suggesting that workplace political comfort may be influenced by factors beyond income alone. The clearest socioeconomic gradient emerged for survey interviewers, showing comfort levels ranging from the highest-income respondents ($M = 1.51$) to the lowest-income groups ($M = 2.04$).

4.4. Socioeconomic Status and Response Authenticity

Income effects extended beyond comfort levels to response authenticity patterns, revealing important implications for polling accuracy across different socioeconomic segments. Higher-income respondents reported significantly greater authenticity in their survey responses across socioeconomic strata ($F(4, 805) = 11.43, p < .001$), suggesting that economic resources may provide greater freedom for authentic political expression in formal survey contexts. This authenticity gradient indicates that polling data may systematically underrepresent the true political preferences of lower-income populations, who may feel greater pressure to provide socially desirable rather than authentic responses.

Employment status was significantly associated with interviewer comfort, with substantial variation across employment categories. As shown in Table 3, unemployed respondents showed markedly lower comfort levels (60 %) than

students (88 %), full-time workers (83 %), and retirees (84 %), further supporting the relationship between economic security and political expression comfort. Gender effects consistently emerged across multiple trust contexts and institutional measures. Women reported significantly lower comfort with survey interviewers ($t = -2.99, p = .003, d = 0.19$) and lower response authenticity ($t = -2.36, p = .018, d = 0.15$), while showing marginally lower trust in electoral polls ($t = -1.70, p = .090$). These patterns suggest that gender dynamics continue to influence political expression in both interpersonal and institutional contexts.

Table 3 presents these results in three sections: the upper section reports gender effects across key trust and authenticity measures, the middle section shows employment status variation in survey interviewer comfort, and the lower section displays the employment $\times$ gender interaction pattern for work colleague contexts.

While socioeconomic factors demonstrate clear patterns in trust-based political expression, the intersection of class and other demographic characteristics may reveal additional complexities. Gender represents a particularly important dimension, as women’s political participation has historically been shaped by distinct social expectations and power dynamics that may influence comfort with political expression across different social contexts.

Table 3. Gender and Employment Effects on Political Expression Comfort Across Trust Circles

Circle/Context	Variable	Male M (SD)	Female M (SD)	t	df	p	Cohen's d
Survey Interviewer	Political Expression Comfort	1.74 (0.90)	1.92 (0.98)	-2.99	967	.003**	0.19
Institutional Trust	Trust in Electoral Polls	2.47 (0.97)	2.58 (0.97)	-1.70	984	.090	0.11
Response Quality	Response Authenticity	1.99 (0.92)	2.12 (0.89)	-2.36	990	.018*	0.15
Democratic Beliefs	Electoral FairnessBelief	2.29 (0.85)	2.42 (0.86)	-2.38	980	.017*	0.15
Employment Status		Comfort with Survey Interviewers (%)					
Full-time Employment		83 %					
Part-time Employment		78 %					
Unemployed		60 %					

Employment Status	Comfort with Survey Interviewers (%)
Students	88 %
Homemakers	73 %
Retired/Disabled	84 %
Employment × Gender Interaction	Work Colleagues Comfort (%)
Male Full-time	65 %
Female Full-time	63 %
Male Unemployed	54 %
Female Unemployed	63 %

Note. $N = 1,008$. Weighted data, 2024 National Telephone Survey of Mexican adults. Gender effects: independent samples t -tests with Cohen's d . Employment effects: descriptive percentages. Comfort question: "In general, how comfortable do you feel sharing your political opinions with the following people or groups?" Lower means indicate greater comfort and authenticity. * $p < .05$. ** $p < .01$.

Gender and Political Expression Across Trust Circles

Gender differences revealed consistent but modest patterns, with higher comfort among male respondents across all trust circles. Men demonstrated higher comfort levels in most contexts: partners (82 % vs. 79 %), family (81 % vs. 79 %), friends (75 % vs. 71 %), and work colleagues (65 % vs. 61 %). The gender gap was most pronounced with survey interviewers, where women reported significantly lower comfort ($M = 1.92$ vs. 1.74 , $t = -2.99$, $p = .003$) and lower response authenticity ($M = 2.12$ vs. 1.99 , $t = -2.36$, $p = .018$). Women also expressed less agreement that elections are free and fair ($M = 2.42$ vs. 2.29 , $t = -2.38$, $p = .017$), suggesting that gender differences in political trust may compound effects on authentic political expression.

4.5. Institutional Trust and Authentic Response Patterns

Correlations With Democratic Legitimacy

Institutional trust measures showed significant positive correlations with comfort levels across all trust circles (see Figure 1c). Belief that "my country is

governed by the will of the people” correlated positively with comfort levels across all trust circles, with the strongest associations emerging for survey interviewers ($r = .184, p < .01$) and work colleagues ($r = .155, p < .01$). Similarly, trust in electoral fairness showed consistent positive associations across all circles, ranging from family relationships ($r = .134, p < .01$) to survey interviewers ($r = .218, p < .01$).

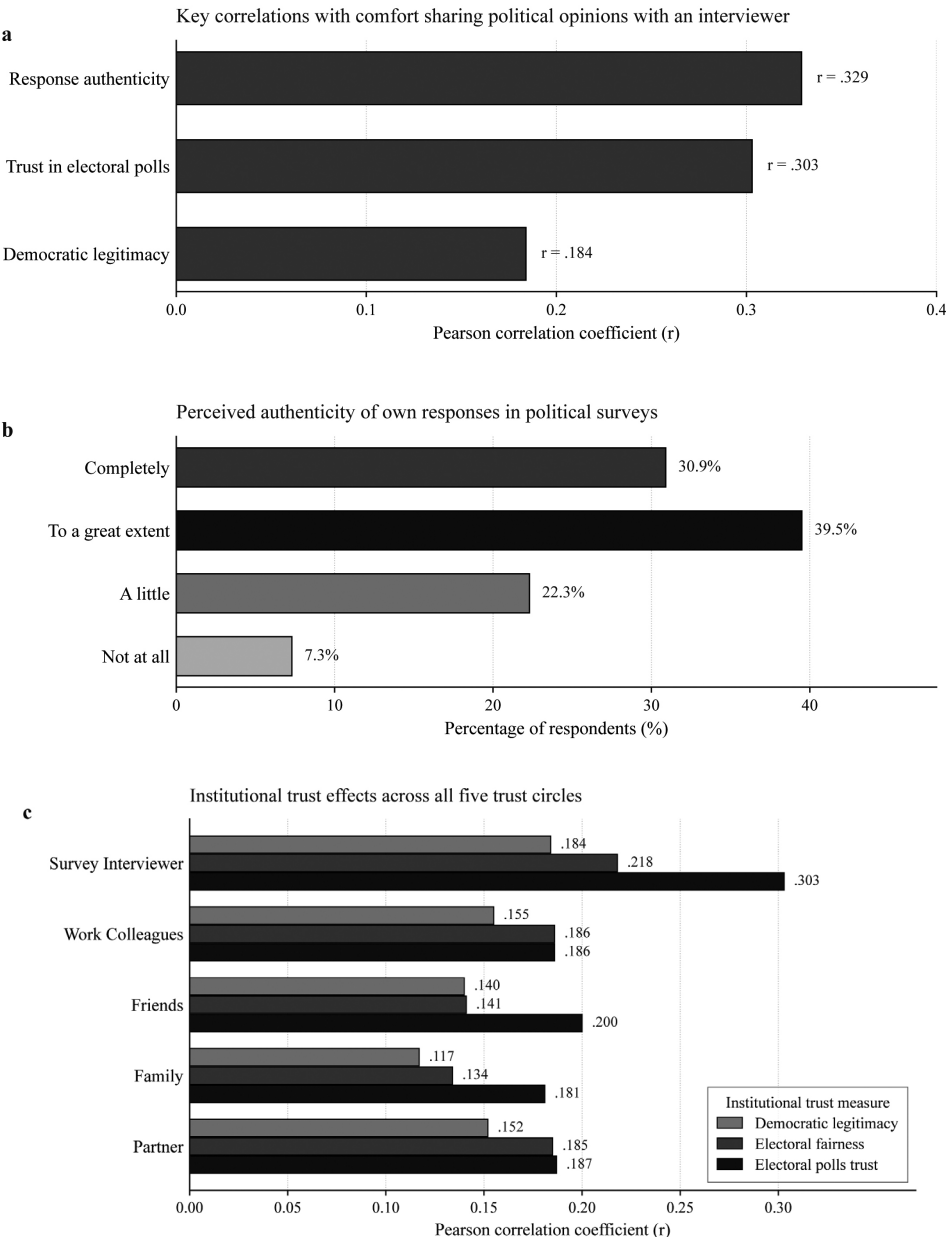
Trust in electoral polls demonstrated the strongest correlation with survey interviewer comfort ($r = .303, p < .01$), suggesting that institutional trust in polling directly affects willingness to engage authentically with survey research. Response authenticity showed particularly strong correlations with survey interviewer comfort ($r = .329, p < .01$), indicating that citizens who reported more authentic responses also felt more comfortable expressing political opinions in formal polling contexts.

Response Authenticity and Social Desirability Perceptions

Self-reported authenticity in survey responses showed meaningful associations with comfort levels across different social circles. Respondents who reported that their survey answers “completely” reflected their true opinions showed higher comfort levels with survey interviewers ($r = .329, p < .01$) and other formal contexts.

Figure 1 brings together the correlational and distributional evidence linking institutional trust and response authenticity to comfort across the five trust circles. Panel 1a shows the three correlations that are highest with survey interviewer comfort: response authenticity ($r = .329, p < .01$), trust in electoral polls ($r = .303, p < .01$), and democratic legitimacy ($r = .184, p < .01$), indicating that institutional trust in polling directly affects willingness to engage authentically with survey research. Panel 1b shows the distribution of self-reported response authenticity: 70.4 % of respondents reported that their survey answers “completely” (30.9 %) or “to a great extent” (39.5 %) reflected their true opinions, while 29.6 % indicated some degree of response modification (22.3 % “a little”, 7.3 % “not at all”). Panel 1c extends the analysis across all five trust circles, plotting the correlations of the three institutional trust measures (democratic legitimacy, electoral fairness, and trust in electoral polls) with comfort in each circle; the strongest association across all three measures appears for the survey interviewer. Critically, respondents with higher trust in electoral polls were significantly more likely to report authentic responses, $\chi^2(3) = 71.08, p < .001, \Phi = .27, N = 974$. Among those reporting “a lot of confidence” in electoral polls, 85.3 % gave authentic responses, compared with only 51.8 % among those with “no confidence at all”—a gradient of 33 percentage points. This finding provides direct evidence that institutional trust in electoral processes serves as a foundation for authentic political expression in survey contexts.

Figure 1. Institutional Trust and Response Authenticity: Correlations With Political Expression Comfort Across Trust Circles



Note. Three-panel summary. Panel a: Pearson correlations between three key institutional-trust and response-authenticity measures and comfort sharing political opinions with a survey interviewer. Panel b: weighted distribution of self-reported response authenticity (“When you are asked about your political views in a survey, how much do you consider that your responses reflect your true opinion?”). Panel c: Pearson correlations between democratic legitimacy, electoral fairness, and trust in electoral polls, and comfort in each of the five trust circles. $N = 1,008$; weighted data, 2024 National Telephone Survey of Mexican adults. Chi-square test of independence (trust in electoral polls $\times$ authentic responses): $\chi^2(3) = 71.08, p < .001, \Phi = .27, N = 974$.

Perceptions of social desirability bias among other respondents revealed patterns distinct from those observed for personal comfort. Respondents who believed that others frequently modify their responses to “make a good impression on the interviewer” showed no significant correlation with personal comfort levels with survey interviewers ($r = -.033, n.s.$). However, these perceptions did correlate meaningfully with institutional trust measures, as discussed below in the Social Desirability Bias Indicators section.

Trust Circle Interconnections

Comfort levels across different trust circles showed strong positive intercorrelations, supporting the theoretical coherence of the circles of trust framework (see Table 4). The intercorrelation matrix reveals systematic patterns of trust circle organization with consistent moderate to strong associations among all circles.

Table 4. Correlations Between Socio-Demographic Variables and Comfort Levels Across Circles of Trust

Measure	1	2	3	4	5	6	7
Socio-Demographics							
1. Age	—						
2. Education Level	-.277***	—					
Circles of Trust							
3. Partner	.014	-.135***	—				
4. Family	-.015	-.049	.495***	—			
5. Friends	.114***	-.110***	.518***	.530***	—		

Measure	1	2	3	4	5	6	7
6. Colleagues	-.045	-.036	.428***	.509***	.586***	—	
7. Survey Interviewer	.032	-.114***	.381***	.407***	.413***	.386***	—

Note. $N = 1,008$. Pearson correlations, weighted data, 2024 National Telephone Survey of Mexican adults. Comfort question: “In general, how comfortable do you feel sharing your political opinions with the following people or groups?” Comfort scale: 1 = Very comfortable to 4 = Not at all comfortable; lower scores indicate greater comfort. Positive correlations with age therefore indicate that older respondents report less comfort, while negative correlations with education indicate that more educated respondents report greater comfort. * $p < .05$. ** $p < .01$. *** $p < .001$.

The strongest correlations occurred between adjacent circles in terms of social intimacy and proximity. Friends and work colleagues showed the highest intercorrelation ($r = .586, p < .01$), followed by family and friends ($r = .530, p < .01$), and family and colleagues ($r = .509, p < .01$). These adjacency patterns suggest that trust circles operate in overlapping but theoretically distinct spheres of social relationships, with comfort readily transferring between socially proximate circles.

Survey interviewer comfort showed moderate but consistent correlations with all interpersonal circles ($r = .381$ to $.413$, all $p < .01$), indicating that formal polling comfort is related to, but empirically distinct from, interpersonal trust patterns. This finding supports the conceptual distinction between interpersonal and institutional trust domains, while confirming their empirical relationships.

4.6. Social Desirability Bias Indicators

Motivations for Response Modification

The analysis of perceived motivations for response modification revealed a clear ordering, as illustrated in Figure 2. The most frequently perceived reason cited by respondents was “protecting their privacy” ($M = 1.99, SD = 1.02, 95\% \text{ CI } [1.93, 2.05]$), indicating a primary concern about the personal implications of their answers. This was followed by “avoiding problems” ($M = 2.14, SD = 1.04, 95\% \text{ CI } [2.08, 2.21]$), which suggests a pragmatic motivation related to potential negative outcomes. The third-ranked motivation was “making a good impression on the interviewer” ($M = 2.21, SD = 1.02, 95\% \text{ CI } [2.14, 2.27]$), which reflects a degree of social conformity. The least frequently perceived motivation was “not wanting to be judged for their opinions” ($M = 2.27, SD = 1.05, 95\% \text{ CI } [2.20, 2.33]$), suggesting a lower overall concern with social judgment in this context. Despite the ordering above, the 95% confidence intervals for “avoiding problems” and “making a good

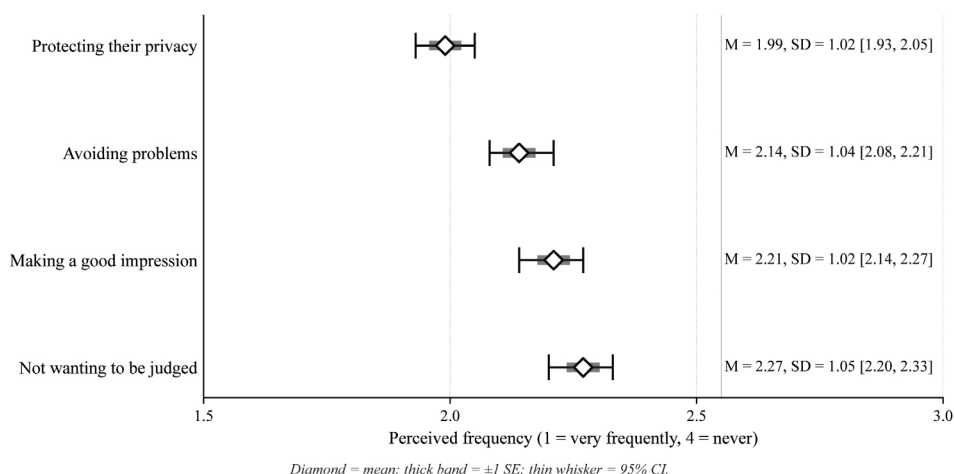
impression” overlap, indicating that the difference between these two motivations cannot be reliably distinguished.

Correlational Patterns

Correlational analysis revealed that respondents’ perceptions of why *others* modify survey responses were not significantly associated with their own reported comfort levels across any of the five trust circles. All bivariate correlations between perceived motivations for response modification (privacy, avoidance, good impression, judgment) and personal comfort levels fell below $|r| = .05$ and were non-significant (all $p > .05$). This pattern suggests that awareness of social desirability pressures in the abstract does not directly translate into constrained personal expression in specific social contexts—a distinction that has implications for how researchers interpret survey-based measures of perceived bias in polling.

In contrast, perceptions of response modification motivations did correlate meaningfully with institutional trust measures. Respondents who perceived others as more prone to modifying their answers to make a good impression on the interviewer or to avoid problems reported lower trust that their country is governed by the will of the people ($r = -.122, p < .01$ and $r = -.126, p < .01$, respectively) and lower agreement that elections are free and fair ($r = -.081, p < .05$ and $r = -.094, p < .01$). This pattern suggests that perceptions of widespread response modification are linked to broader concerns about democratic legitimacy rather than to personal reticence across social circles.

Figure 2. Perceived Motivations for Modifying Survey Responses



Note. Dot-and-interval plot showing the weighted mean (diamond marker), the ± 1 standard error band (thick horizontal band), and the 95 % confidence interval (thin whisker) for each of the four perceived motivations for response modification. Items are displayed in ascending order of mean (top to bottom). Scale: 1 = very frequently, 4 = never; lower values indicate a higher perceived frequency. The x-axis is plotted over the range 1.5 to 3.0 to reflect the location of the confidence intervals and avoid a misleading truncated axis. Item-level weighted *n*: privacy 976, avoiding problems 973, good impression 959, not being judged 983. Weighted data, 2024 National Telephone Survey of Mexican adults; *N* = 1,008.

Individual Differences in Response Authenticity

Self-reported response authenticity varied considerably across the sample, with 70.4 % reporting that their survey answers “completely” or “to a great extent” reflected their true opinions, while 29.6 % indicated some degree of response modification (see Table 5). The distribution of social desirability motivations revealed that privacy protection was perceived as the most common reason for response modification ($M = 1.99$, $SD = 1.02$), while not wanting to be judged for their opinions was perceived as the least common ($M = 2.27$, $SD = 1.05$).

Respondents with higher trust in electoral polls were significantly more likely to report authentic responses, $\chi^2(3) = 71.08$, $p < .001$, $\Phi = .27$, $N = 974$, providing direct evidence for the relationship between institutional trust and survey response validity.

Table 5 summarizes these findings in four sections: the upper section reports the response authenticity distribution, the second section ranks the four perceived motivations for response modification, the third section shows their correlations with comfort across trust circles, and the lower section reports the chi-square test linking trust in electoral polls to authentic response patterns.

Table 5. Response Authenticity and Social Desirability Motivations in Political Survey Context

Response Authenticity Category	Weighted Percentage	Cumulative Percentage
Completely reflects true opinion	30.9 %	30.9 %
To a great extent reflects true opinion	39.5 %	70.4 %
A little reflects true opinion	22.3 %	92.7 %
Does not reflect true opinion	7.3 %	100.0 %

Correlations with Trust Circles	Good Impression	Avoiding Problems	Not Being Judged	Protecting Privacy
Social Desirability Motivation		M	SD	Rank
Protecting privacy		1.99	1.02	1 (Most common)
Avoiding problems		2.14	1.04	2
Making good impression on interviewer		2.21	1.02	3
Not wanting to be judged for opinions		2.27	1.05	4 (Least common)
Correlations with Trust Circles	Good Impression	Avoiding Problems	Not Being Judged	Protecting Privacy
Partner comfort	-.039	.004	.014	.022
Family comfort	-.011	-.018	.002	.039
Friends comfort	.023	.014	.028	.039
Work colleagues comfort	-.041	-.041	-.044	.001
Survey interviewer comfort	-.033	.014	-.003	.019
Trust in Electoral Polls × Authenticity		$\chi^2(3)$	p	Effect Size
Higher trust in electoral polls, authentic responses		71.08	< .001**	Phi = .27 (moderate)

Note. N = 1,008. Weighted data, 2024 National Telephone Survey of Mexican adults. Social desirability motivations: 4-point scale; lower scores indicate higher perceived frequency. Chi-square test N = 974 (reflecting valid cases with non-missing values on both variables). None of the correlations between SDB motivations and trust circle comfort reached statistical significance (all $p > .05$). ** $p < .01$.

5. DISCUSSION

5.1. *Interpreting the Trust Gradient Phenomenon*

The results provide evidence for systematic trust gradients that shape political expression. Comfort proceeds from intimate partners (80.5 %) and family (80.0 %) but then reveals a surprising inversion: respondents reported higher comfort with survey interviewers (78.5 %) than with friends (73.0 %), followed by work colleagues (63.0 %). Political communication thus operates within distinct social spheres characterized by different levels of interpersonal trust and perceived social risks.

Perhaps most striking is the positioning of survey interviewers within this trust hierarchy. While conventional wisdom might suggest that formal polling contexts would rank lowest in terms of comfort due to their institutional nature, our findings reveal that respondents actually express greater comfort with survey interviewers than with work colleagues. This pattern suggests that the perceived anonymity and professional neutrality of survey research may create a unique social space that falls between intimate personal relationships and potentially contentious workplace interactions.

Circles of Trust and Democratic Legitimacy in Polarized Contexts

The strong correlations between institutional trust and authentic political expression across all trust circles have profound implications for understanding democratic legitimacy in polarized societies. Our findings demonstrate that respondents who express greater trust in democratic institutions and electoral processes are significantly more likely to report authentic responses across all trust circles. This pattern suggests that institutional trust serves as a foundation for broader social and political communication, extending beyond formal institutional interactions to shape interpersonal political discourse.

In Mexico's evolving democracy, institutional trust remains contested across social groups. While recent data show that 54 % of Mexicans report moderate-to-high trust in the federal government—above the OECD average—this trust varies substantially by perceptions of political voice and representation (OECD, 2024). Our results suggest that these trust deficits may create cascading effects where certain population segments systematically withdraw from authentic political expression.

5.2. *Limitations and Methodological Considerations*

A central methodological limitation concerns the potential endogeneity between the survey mode and key variables of interest. This study employed telephone interviews with live interviewers to measure respondents' comfort in expressing political opinions to pollsters and their self-reported response authenticity. Because social desirability bias is inherently influenced by interpersonal dynamics, the presence of an interviewer may have affected how honestly respondents reported their trust in pollsters and the degree to which they modified their survey answers. Specifically, the measurement of comfort with survey interviewers was itself conducted through a survey interview, potentially inflating reported comfort levels and underestimating the prevalence of response modification. This limitation is particularly critical given that interviewer-respondent comfort lies at the core of the circles of trust framework. While we cannot address this concern empirically with the current data, future research should consider mixed-mode designs comparing telephone, online, and self-administered questionnaire responses to disentangle mode effects from the trust dynamics under investigation. Therefore, the patterns documented here should be interpreted as potentially conservative estimates of authentic discomfort with political expression in survey contexts.

While this study provides valuable insights into the relationship between circles of trust and political expression, several other limitations should be acknowledged. First, the cross-sectional design prevents us from establishing definitive causal relationships between institutional trust and patterns of political expression. While our theoretical framework suggests that institutional trust influences political communication comfort, reverse causation or third-variable explanations cannot be ruled out without longitudinal data.

Second, the study relied on self-reported measures of comfort and authenticity, which may themselves be subject to social desirability pressures. Respondents who are generally prone to socially desirable responses may over-report their comfort with political expression or their authenticity in survey responses. Future research should incorporate behavioral or experimental measures to validate these self-report findings.

Third, the focus on Mexico limits the generalizability of the findings to other national contexts. While Mexico provides an excellent case for studying these dynamics in an emerging democracy, the specific patterns of trust gradients and their demographic correlates may vary in societies with different political histories, institutional arrangements, and social structures.

Fourth, the exclusive use of landline telephones for data collection introduces a potential coverage bias. In Mexico, landline penetration is concentrated among urban households with higher socioeconomic status and older residents. While post-stratification weights adjusted for age and gender, they cannot fully correct

for attitudinal differences between populations with and without landline access. This limitation may partially explain the observed associations between income, education, and comfort levels and suggests caution when generalizing findings to populations excluded from the sampling frame.

Fifth, the survey was conducted in November 2024, a politically charged moment following Mexico's controversial judicial reform and amid heightened institutional polarization. This timing may have influenced both institutional trust levels and respondents' comfort with political expression. The documented patterns of distrust toward democratic institutions and constrained political communication may reflect this specific political climate rather than stable dispositions. Conversely, if polarization typically suppresses authentic political expression, the comfort levels observed here may underestimate what would be found in less contentious periods. Future research should examine the stability of trust gradients across different political contexts.

Sixth, the analysis did not include partisan identification or other political variables that may moderate trust patterns, a notable omission given Mexico's polarized political environment. Additionally, when rating comfort with survey interviewers, respondents may have imagined different types of pollsters (academic, commercial, or political), potentially introducing measurement heterogeneity. Finally, interviewer characteristics such as gender, age, or perceived social status may influence respondent comfort in ways that this study could not assess. Information on interviewer demographics was not available for this analysis; investigating potential matching effects (e.g., same-gender interviewer-respondent dyads) would require experimental designs or detailed interviewer-level data that were beyond the scope of this study.

5.3. *Directions for Future Research*

The circles of trust framework opens several promising avenues for future research on political communication and survey methodology. First, longitudinal studies could examine how circles of trust evolve over time and in response to political events. Understanding whether trust gradients remain stable or shift in response to changing political circumstances would provide important insights into the dynamic nature of political communication patterns.

Second, experimental research could test the causal effects of different survey contexts and interviewer characteristics on response authenticity. Randomizing features such as interviewer demographics, question framing, or privacy assurances could help identify specific factors that enhance or diminish authentic political expression in survey contexts.

Third, comparative research across different national contexts could examine how political systems, cultural norms, and institutional arrangements shape trust gradients and their demographic correlates. Such research could help identify universal principles of trust and political expression while recognizing context-specific factors that influence these relationships.

Fourth, research should examine how digital technologies and social media platforms may create new circles of trust or alter existing trust relationships. The growing importance of online political communication and digital political participation may require expanding the trust circle framework to include virtual social contexts and digital political relationships.

Finally, intervention research could test whether specific training, institutional reforms, or communication strategies can enhance trust and authentic political expression. Such research could provide evidence-based recommendations for improving both survey methodology and broader democratic communication practices in polarized societies.

6. CONCLUSION

This study addressed a methodological question: How does social context shape respondents' willingness to express authentic political opinions in surveys? By introducing the circles of trust framework, we propose that social desirability bias operates relationally rather than as a stable individual trait. Our empirical analysis of Mexican respondents supports this reconceptualization and reveals patterns with implications for survey design in polarized democracies.

Our findings from Mexico reveal three patterns with implications for survey methodology and democratic theory. First, political expression comfort follows a gradient from intimate relationships (partners, family) through broader social contexts (friends, colleagues) to institutional representatives (survey interviewers) but with a notable inversion: respondents reported higher comfort with pollsters than with friends or colleagues. This unexpected finding suggests that professional neutrality and perceived anonymity may create distinct conditions for political disclosure that differ from those of everyday social interactions.

Second, institutional trust serves as a foundation for authentic expression across all trust circles. Citizens who distrust democratic institutions report constrained political communication not only in formal contexts but also within their personal networks, suggesting that democratic legitimacy deficits may cascade through social relationships to undermine both survey validity and informal political dialogue.

Third, systematic demographic variations in comfort levels—with younger, more educated, and higher-income respondents reporting greater ease with

political expression—indicate that conventional demographic weighting may not adequately address trust-related biases in polling data.

As democratic societies are confronted with increasing polarization and institutional distrust, understanding the social foundations of authentic political communication becomes urgent. The circles of trust framework provides both theoretical insight and practical tools for improving survey methodology while recognizing that polling accuracy ultimately depends on the broader social conditions that enable citizens to express their preferences without fear of judgment or reprisal.

7. REFERENCES

- Belardinelli, S., & Gili, G. (2022). Credibility, trust, and suspicion: Social relationships put to the test by Covid-19. *Church, Communication and Culture*, 7(1), 91–113. <https://doi.org/10.1080/23753234.2022.2039075>
- Bischoping, K., & Schuman, H. (1992). Pens and polls in Nicaragua: An analysis of the 1990 preelection surveys. *American Journal of Political Science*, 36(2), 331–350. <https://doi.org/10.2307/2111480>
- Blanco, L. R. (2013). The impact of crime on trust in institutions in Mexico. *European Journal of Political Economy*, 32, 38–55. <https://doi.org/10.1016/j.ejpoleco.2013.06.004>
- Chen, S., Körtner, J., Selb, P., & Wiederspohn, J. (2023). *Electoral predictors of polling errors*. <https://doi.org/10.33774/apsa-2023-t1vh8-v2>
- Claassen, R. L., & Ryan, J. B. (2024). Biased polls: Investigating the pressures survey respondents feel. *Acta Politica*. <https://doi.org/10.1057/s41269-024-00356-4>
- De Micheli, D., & Taylor, W. K. (2024). Public trust in Latin America's courts: Do institutions matter? *Government and Opposition*, 59(1), 146–167. <https://doi.org/10.1017/gov.2022.6>
- Delhey, J., Newton, K., & Welzel, C. (2011). How general is trust in “most people”? Solving the radius of trust problem. *American Sociological Review*, 76(5), 786–807. <https://doi.org/10.1177/0003122411420817>
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Dryden Press.
- Frisbie, S. (2024, March 21). Global elections in 2024 could open a Pandora's box of political polarization. *Morning Consult*. <https://pro.morningconsult.com/analysis/tracking-political-polarization-2024-global-elections>
- Fukuyama, F. (2000). *Social capital and civil society* (IMF Working Paper No. WP/00/74). International Monetary Fund.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380. <https://doi.org/10.1086/225469>
- Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74(1), 37–67. <https://doi.org/10.1093/poq/nfp065>

- Huckfeldt, R. R., & Sprague, J. (1995). *Citizens, politics and social communication: Information and influence in an election campaign*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511664113>
- Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, 61(5), 534–547. <https://doi.org/10.1177/1470785318805305>
- Maldonado Hernández, G. de J., & Crow, D. (2013). *Are we measuring institutional trust right? Evidence from a Mexican survey experiment* (Working paper). <http://repositorio-digital.cide.edu/handle/11651/884>
- Miller, A. S., & Mitamura, T. (2003). Are surveys on trust trustworthy? *Social Psychology Quarterly*, 66(1), 62–70. <https://doi.org/10.2307/3090141>
- OECD. (2024). *OECD survey on drivers of trust in public institutions 2024 results: Mexico*. Organisation for Economic Co-operation and Development. https://www.oecd.org/en/publications/oecd-survey-on-drivers-of-trust-in-public-institutions-2024-results-country-notes_a8004759-en/mexico_14d321b0-en.html
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598–609. <https://doi.org/10.1037/0022-3514.46.3.598>
- Silber, H., Moy, P., Johnson, T. P., Neumann, R., Stadtmüller, S., & Repke, L. (2022). Survey participation as a function of democratic engagement, trust in institutions, and perceptions of surveys. *Social Science Quarterly*, 103(7), 1619–1632. <https://doi.org/10.1111/ssqu.13218>
- Streb, M. J., Burrell, B., Frederick, B., & Genovese, M. A. (2008). Social desirability effects and support for a female American president. *Public Opinion Quarterly*, 72(1), 76–89. <https://doi.org/10.1093/poq/nfm035>
- United Nations Development Programme. (2024, January 23). *In whom do we trust? Less in institutions and more in communities in LAC*. <https://www.undp.org/latin-america/blog/whom-do-we-trust-less-institutions-and-more-communities-lac>
- Yuen, S., & Lee, F. L. F. (2025). Echoes of silence: How social influence fosters self-censorship under democratic backsliding. *Democratization*, 32(5), 1213–1238. <https://doi.org/10.1080/13510347.2025.2455127>