

DOI: https://doi.org/10.48009/1_iis_2026_125

Digital engagement and identity online: Measuring deviance propensity through projective vignette methodology

Frank X. Hartle III, *Robert Morris University*, hartle@rmu.edu

Christopher Wydra, *Pennsylvania Western University*, wydra@pennwest.edu

Robert Ambrosini, *Robert Morris University*, ambrosini@rmu.edu

Abstract

Social desirability bias undermines self-report measures of online misconduct. This study used a projective vignette methodology to indirectly assess deviance propensity among 202 U.S. adults. Participants evaluated three scenarios (technical exploitation, anonymous misinformation, and professional identity fraud), providing moral judgments, peer prevalence estimates, and personal behavioral intentions. Higher peer estimates significantly predicted personal willingness to deviate across all scenarios, validating the indirect approach. The strongest predictor of deviance was the belief that one's online and offline identities are psychologically separate. Younger adults reported substantially higher deviance propensity. Findings support Suler's (2004) dissociative anonymity construct.

Keywords: online deviance, projective vignette, false consensus effect, social desirability bias, anonymity, dissociative anonymity

Introduction

People lie on surveys. This is more than an inconvenience; it constitutes a substantial threat to the reliability and validity of research on online misconduct. Social desirability bias leads respondents to underreport the very behaviors that researchers, law enforcement, and platform designers most need to understand (Krumpal, 2013). When a survey directly asks whether someone has spread misinformation, exploited a system vulnerability, or created a fake profile, the answer is almost always no, even when behavioral trace data suggests otherwise.

For two decades, psychology scholars have investigated how online environments alter moral decision-making. Suler's (2004) concept of the online disinhibition effect suggests that digital anonymity and the absence of face-to-face cues may cause a person to alter their behavior. While this disinhibition can foster positive social engagement, it may also manifest as acting out, ranging from social aggression to technical exploitation. Accurately measuring these behaviors, however, has remained a persistent challenge.

This study tested a method inspired by the false consensus effect (Ross et al., 1977): rather than asking people what they would do, participants assessed what everyone else would do, and those assessments were compared with their own stated behavioral intentions. Using three hypothetical scenarios representing distinct categories of online deviance (technical exploitation, anonymous misinformation, and professional identity fraud), this study examined whether indirect projective measures can capture the propensity to commit deviant acts that direct questions may miss, and whether the specific psychological mechanism of dissociative anonymity predicts deviance.

Background and Literature Review

The Nature of Online Deviance

Online deviance includes behaviors that violate what others see as appropriate in digital communities or platforms (Holt & Bossler, 2014). Unlike physical crime, digital misconduct exists on a continuum from ambiguously unethical (sharing a streaming password) to clearly criminal (hacking, fraud). Wall's (2001) typology of cybercrime, distinguishing cyber-trespass, cyber-deception, cyber-pornography, and cyber-violence, provides a framework, but most existing classifications focus on criminal behavior. The vast majority of online deviance may fall below the criminal threshold. This sub-criminal deviance may be far more prevalent and may normalize more serious misconduct.

Measuring the prevalence of deviance presents a persistent challenge. Direct self-reports are compromised by social desirability bias; respondents consistently underreport unethical acts to align with perceived social norms (Krumpal, 2013). The specific mechanisms by which this bias operates in digital contexts include uncertainty about legality, perception of researcher judgment, and self-monitoring through the survey instrument itself. As such, the survey data may underestimate the true prevalence of digital misconduct, establishing the need for indirect measurement approaches.

Theoretical Foundations

The online disinhibition effect

Suler (2004) identified six factors contributing to online disinhibition: dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, and minimization of status and authority. Dissociative anonymity is the psychological separation of one's online actions from one's offline identity and is most relevant to the present study. This effect creates the "online self" that operates in a consequence-free zone. Subsequent empirical work has tested this theory: Lapidot-Lefler and Barak (2012) demonstrated the relationship between anonymity and flaming behavior, while Zimmerman and Ybarra (2016) examined online aggression. Most of these tests relied on direct self-report, and may be compromised by social desirability bias.

The false consensus effect and projective measurement

The false consensus effect (Ross et al., 1977) describes a tendency for individuals who engage in behavior to overestimate its prevalence among peers. This is not just an error in reasoning; it serves a motivated function, normalizing one's behavior by attributing it to the majority (Marks & Miller, 1987). This principle provides the construct for projective measurement. If people who deviate from social norms project their behavior onto others, asking respondents to estimate peer behavior may be an indication of personal behavioral tendencies. The vignette methodology (Wason et al., 2002; Alexander & Becker, 1978) implements this approach through hypothetical scenarios that reduce defensiveness while maintaining validity.

Social norms theory

Cialdini and Goldstein (2004) distinguish between injunctive norms (what people believe others approve of) and descriptive norms (what people believe others actually do). The present study measures both: acceptability ratings capture injunctive norms, while projective estimates capture perceived descriptive norms. Literature demonstrates that people consistently overestimate the prevalence of risky behavior among peers (Perkins & Berkowitz, 1986; Borsari & Carey, 2003), and these misperceptions may predict personal behavior, providing a theoretical basis for our projective vignette approach.

Anonymity, Identity, and Moral Decision-Making

De-individuation research (Zimbardo, 1969; Postmes & Spears, 1998) asserts that anonymity reduces self-awareness and weakens the connection between personal values and behavior. In digital contexts, Christopherson (2007) showed that even partial technical anonymity creates the psychological conditions

for disinhibition. Turkle (1995) introduced the concept of online identity multiplicity. This is the idea that digital environments allow individuals to cycle through different selves. The argument of this study is that individuals who perceive a boundary between their online and offline identities may be more willing to engage in online behavior they would never perform in person. This may not be a function of anonymity; it may be a deeper psychological separation in which the online self operates under different rules.

Bandura's (1991, 2002) moral disengagement study identifies mechanisms relevant to digital misconduct: diffusion of responsibility ("everyone does it"), dehumanization of victims (reduced social cues online), and displacement of responsibility ("the platform allows it"). If individuals believe deviance is the majority behavior, this may activate diffusion of responsibility and weaken moral self-sanctions even among those who personally condemn the behavior.

Demographic Moderators

Age differences in online deviance are supported by developmental psychology research on moral reasoning maturation (Steinberg, 2008) and findings on cyberbullying (Hinduja & Patchin, 2008). Gender differences show mixed results: some studies find higher rates of technical deviance among males (Higgins et al., 2006), while other studies suggest that the direct effect of gender on online deviance may weaken once attitudinal and related factors are taken into account (Wang et al., 2023).

Linking Theory to the Research Questions

Each research question operationalizes one of these frameworks. RQ1 tests the false consensus effect (Ross et al., 1977): if projection serves a motivated normalization function, estimates of peer deviance should track personal behavioral intentions. RQ2 tests the dissociative anonymity component of the online disinhibition effect (Suler, 2004): if psychological separation of online and offline identity is the operative mechanism, identity separation beliefs should predict deviance propensity more strongly than general beliefs about anonymity's effects. RQ3 draws on social norms theory and typologies of online misconduct (Cialdini & Goldstein, 2004; Wall, 2001): if injunctive and descriptive norms vary across categories of misconduct, moral judgments and behavioral intentions should differ across technical, social, and identity deviance rather than reflecting a single undifferentiated propensity.

Research Questions

RQ1: When people believe that most other people would engage in deviant online behavior, are they themselves more likely to admit they would do the same?

RQ2: Do people who believe their online identity is separate from their real-life identity show a greater willingness to engage in online deviance?

RQ3: Do people judge and respond to different types of online deviance differently, specifically, do they treat exploiting a technical glitch differently than spreading misinformation or faking a professional profile?

Methodology

Research Design

This study utilized a quantitative, cross-sectional survey design using a projective vignette methodology to mitigate social desirability bias (Krumpal, 2013). The approach relies on the false consensus effect (Ross et al., 1977): participants estimated the prevalence of deviant behaviors among peers as an indirect proxy for their own behavioral tendencies (Wason et al., 2002).

Participants

Participants were 202 U.S. adults (56.9% female, 42.6% male) recruited through SurveyMonkey Audience. The modal age group was 35–44 (31.7%), with 71.3% in the 25–54 range. Master’s degree holders were the largest educational group (34.7%). Data collection occurred primarily on February 20, 2026. Nearly all respondents (98.5%) completed the survey on mobile devices. Completion times ranged from 0.5 to 75.9 minutes (*Mdn* = 2.4; *IQR* = 1.2–4.4); no respondent completed the survey in under 30 seconds.

Measures

Three hypothetical scenarios represented the deviance categories: technical deviance (exploiting a streaming configuration error), social deviance (spreading anonymous political misinformation), and identity deviance (creating a fake professional networking profile). For each scenario, participants provided (a) a moral acceptability rating on a 7-point scale; (b) a projective norm estimate (0–100% slider estimating the percentage of average internet users who would engage in the behavior); and (c) a personal behavioral intention rating on a 7-point scale under conditions emphasizing anonymity. Three additional items measured anonymity attitudes on 7-point agreement scales: free expression without self-disclosure, separation of online and offline identities, and the general belief that anonymity changes behavior. Table 1 maps each survey item to the study constructs; the full instrument, including vignette texts, exact item wording, scale anchors, item order, and scoring, appears in Appendix A.

A Total Deviance Propensity Score was computed as the mean of the three behavioral intention items. Because the three scenarios deliberately sample heterogeneous categories of deviance, this score is treated as a broad composite index of deviance propensity rather than a validated unidimensional scale; its internal consistency ($\alpha = .668$) is reported for transparency and is expectedly modest given the heterogeneity of item content. The composite is used only as the outcome in the regression analyses; all other analyses are conducted and reported at the scenario level. The three anonymity attitude items showed low composite reliability ($\alpha = .546$) and were therefore treated as separate indicators in all inferential analyses; composite-level anonymity results are reported only descriptively.

Table 1. Mapping of survey items to study constructs

Construct	Vignette	Item(s)	Response format
Technical deviance: acceptability (injunctive norm), projective norm (descriptive norm), behavioral intention	A: streaming configuration error	Q4, Q5, Q6	7-pt acceptability; 0–100% slider; 7-pt likelihood
Social deviance: acceptability, projective norm, behavioral intention	B: anonymous political misinformation	Q7, Q8, Q9	7-pt acceptability; 0–100% slider; 7-pt likelihood
Identity deviance: acceptability, projective norm, behavioral intention	C: fake professional networking profile	Q10, Q11, Q12	7-pt acceptability; 0–100% slider; 7-pt likelihood
Anonymity attitudes: free expression; identity separation; anonymity changes behavior (separate indicators)	—	Q13a, Q13b, Q13c	7-pt agreement
Total Deviance Propensity Score (composite index)	All scenarios	Mean of Q6, Q9, Q12	1–7

Procedure

The study received IRB approval. Participants reviewed a consent statement describing the study as investigating “digital engagement and norms” (less than full disclosure intended to prevent priming). Forced-response validation ensured complete data on all scenario items. Twelve participants (5.9%)

skipped at least one of the three anonymity attitude items (eight skipped two items; four skipped one); analyses involving these items used listwise deletion, yielding $n = 190$ complete cases for the regression model. All participants received a full debriefing upon completion, disclosing the study's true purpose.

Data Analysis

Because the same participants evaluated all three scenarios, repeated-measures analyses were used for cross-scenario comparisons. Friedman tests were conducted to compare scenarios on acceptability ratings and behavioral intention ratings, followed by Wilcoxon signed-rank tests for pairwise comparisons, with effect sizes computed as $r = |Z|/\sqrt{N}$. Paired t -tests assessed within-scenario acceptability–intention gaps; Pearson and Spearman correlations tested projective norm–intention associations; and multiple linear regression tested anonymity attitudes as predictors of deviance propensity. All tests were two-tailed, with $\alpha = .05$.

No a priori power analysis was conducted; the sample size was determined by available panel resources. A post hoc sensitivity analysis indicated that with $n = 190$, $\alpha = .05$, and power = .80, the three-predictor regression could detect effects as small as $f^2 \approx .06$ (equivalent to $R^2 \approx .05$). The observed model effect ($f^2 = .78$) substantially exceeded this threshold; the primary regression findings were therefore not power-limited, although smaller effects such as pairwise scenario contrasts warrant more cautious interpretation.

Because the median completion time was short, all key analyses were re-estimated excluding the 32 respondents (15.8%) who completed the survey in under 60 seconds. The projective norm–intention correlations ($r = .34, .41$, and $.41$) and the regression results (identity separation $\beta = .48$) were substantively unchanged, indicating that the findings are not attributable to insufficient-effort responding.

Results/Findings

Descriptive Statistics

Acceptability ratings differed significantly across scenarios, Friedman $\chi^2(2) = 30.72$, $p < .001$. Technical deviance was rated near-neutral ($M = 3.99$, $SD = 2.11$) with a notably bimodal distribution and nearly equal peaks at the extremes. Social deviance drew the strongest condemnation ($M = 3.03$, $SD = 2.13$; 61.9% unacceptable), and identity deviance fell between the two ($M = 3.22$, $SD = 2.00$; 54.9% unacceptable). Pairwise Wilcoxon signed-rank tests showed that technical deviance was rated as significantly more acceptable than both social deviance ($W = 2062.5$, $Z = -5.69$, $p < .001$, $r = .40$) and identity deviance ($W = 2308.0$, $Z = -5.27$, $p < .001$, $r = .37$), whereas social and identity deviance did not differ significantly ($W = 2308.0$, $Z = -1.85$, $p = .064$, $r = .13$). (The identical W values for the latter two comparisons arise from different numbers of non-zero paired differences and are coincidental.)

All projective norms exceeded 50%: technical deviance averaged 71.4% ($SD = 20.43$), social deviance averaged 57.1% ($SD = 24.37$), and identity deviance averaged 53.0% ($SD = 26.94$). This rank ordering mirrored both acceptability and behavioral intention.

Behavioral intention followed the same hierarchy, Friedman $\chi^2(2) = 38.00$, $p < .001$. Over half of respondents (50.5%) endorsed technical deviance, whereas approximately 28–30% endorsed social and identity deviance. Pairwise Wilcoxon signed-rank tests indicated that technical deviance received significantly higher behavioral-intention ratings than both social deviance ($W = 2381.5$, $Z = -5.60$, $p < .001$, $r = .39$) and identity deviance ($W = 2349.5$, $Z = -6.19$, $p < .001$, $r = .44$), while social and identity deviance did not differ significantly ($W = 2468.5$, $Z = -0.02$, $p = .982$, $r < .01$). The Total Deviance Propensity Score was $M = 3.54$ ($SD = 1.67$). A paired t -test confirmed that behavioral intention significantly exceeded moral acceptability for technical deviance, $t(201) = 2.83$, $p = .005$, $d = 0.20$, but not for social deviance, $t(201) = 1.00$, $p = .320$, $d = 0.07$, or identity deviance, $t(201) = -0.43$, $p = .667$, $d = -0.03$.

For anonymity attitudes, agreement that anonymity changes behavior was near-universal ($M = 5.73$; 82.3% agreement). Free expression was moderately above neutral ($M = 4.36$, $p = .003$). The identity separation belief was statistically indistinguishable from neutral ($M = 3.92$, $p = .573$).

Taken together, these comparisons answer RQ3: respondents did not treat all forms of online deviance alike. Technical deviance was judged significantly more acceptable, and was endorsed at significantly higher rates, than either social or identity deviance, which did not differ from one another.

Correlational Analyses

All scenario-level correlations between projective norms and behavioral intentions were positive, moderate, and highly significant, confirming that the projective approach captures individual differences in deviance propensity.

Table 2. Pearson correlations between projective norms and behavioral intentions

Scenario pair	<i>r</i>	<i>p</i>	Spearman ρ
Projective norm A → Intention A	.306	< .001	.334
Projective norm B → Intention B	.420	< .001	.405
Projective norm C → Intention C	.423	< .001	.391
Average projective norm → Total deviance	.382	< .001	—

Note. $N = 202$. All correlations significant at $p < .001$.

Acceptability and intention were strongly correlated within each scenario ($r = .704-.798$, all $p < .001$). The identity separation belief showed the strongest bivariate association with Total Deviance Propensity ($r = .636$, $p < .001$), followed by free expression ($r = .534$, $p < .001$). The widely endorsed belief that anonymity changes behavior was not significant ($r = .119$, $p = .100$).

These results directly answer RQ1: participants who believed that most other people would engage in deviant online behavior were themselves significantly more willing to do the same, and this held across all three scenarios ($r = .306-.423$, all $p < .001$).

Regression Analyses

Total Deviance Propensity (TDP) was regressed on the three anonymity attitude items using ordinary least squares:

$$\text{TDP} = b_0 + b_1(\text{Free Expression}) + b_2(\text{Identity Separation}) + b_3(\text{Anonymity Changes Behavior}) + e \quad (1)$$

where TDP is the mean of the three behavioral intention items (1–7; see Measures and Table 1) and the predictors are the three 7-point agreement items. Table 3 presents the results.

Table 3. Multiple regression: Anonymity attitudes predicting total deviance propensity

Predictor	<i>B</i>	<i>SE</i>	95% CI	<i>t</i>	<i>p</i>	β
Intercept	0.356	0.446	[−0.523, 1.235]	0.800	.425	—
Free expression	0.194	0.073	[0.050, 0.338]	2.658	.009	0.190
Identity separation	0.449	0.062	[0.326, 0.572]	7.219	< .001	0.514
Anonymity changes behavior	0.100	0.067	[−0.033, 0.232]	1.481	.140	0.082

Note. $n = 190$. $R^2 = .438$, Adj. $R^2 = .429$, $F(3, 186) = 48.23$, $p < .001$. CI = confidence interval.

The model explained 43.8% of the variance. Identity separation was the dominant predictor ($\beta = 0.514$, $p < .001$). In substantive terms, each additional point of agreement that one's online and offline identities are separate was associated with a 0.45-point increase in deviance propensity on the 7-point scale ($B = 0.449$, 95% CI [0.326, 0.572]), holding the other attitudes constant. Free expression contributed a smaller unique effect ($B = 0.194$, 95% CI [0.050, 0.338]), and the general belief that anonymity changes behavior contributed no significant unique effect ($B = 0.100$, 95% CI [-0.033, 0.232]). Regression diagnostics indicated no multicollinearity (all VIFs ≤ 1.70) and no significant heteroskedasticity (Breusch-Pagan $p = .308$); residuals departed modestly from normality (Shapiro-Wilk $p = .003$), which is common with Likert-based composites and unlikely to bias inference at this sample size.

Demographic controls were then added, with age coded at bracket midpoints, gender as a male indicator, and education coded ordinally. The expanded model explained 49.1% of the variance ($n = 189$); age was the only significant demographic predictor ($B = -0.026$, $p < .001$), and identity separation remained the dominant predictor ($\beta = .47$, $p < .001$). Respondents aged 18–34 scored substantially higher on deviance propensity ($M = 3.74$) than those 55 and older ($M = 2.01$), $t(103) = 7.37$, $p < .001$, $d = 1.44$. Gender ($p = .698$) and education ($p = .260$) were not significant.

These results answer RQ2 affirmatively: the belief that one's online identity is separate from one's offline identity was the strongest predictor of willingness to engage in online deviance ($\beta = 0.514$, $p < .001$), whereas the general belief that anonymity changes behavior was not a significant predictor.

Discussion of Findings

The central question was whether projective peer estimates could serve as valid indirect indicators of deviance propensity. The individual-level correlations ($r = .306$ – $.423$, all $p < .001$) consistently supported this relationship across all three scenarios, confirming that the false consensus projection mechanism operated as theorized (Ross et al., 1977). The weaker correlation for technical deviance may reflect a ceiling effect, since the projective norm of 71.4% reduced individual variation. Researchers using this approach should note that projective estimates are most useful when the behavior is not already seen as nearly universal.

A clear hierarchy emerged. Technical deviance was rated as the most acceptable and received the highest level of endorsement, whereas social and identity deviance drew stronger disapproval and did not differ from one another. This may suggest that the public draws a qualitative moral distinction between exploiting a system and deceiving people. The finding that 50.5% would exploit the streaming error, while only 40.6% judged it acceptable, may point to a moral override effect: the perceived absence of consequences reduced moral restraint. This interpretation is consistent with Suler's (2004) distinction between benign and toxic disinhibition.

The regression analysis showed that attitudes toward anonymity explained 43.8% of the variance, which is a substantial effect. However, the three attitude items did not work in the same way. The strongest predictor was the belief that a person's online identity is separate from their offline identity ($\beta = 0.514$), which closely reflects Suler's (2004) concept of dissociative anonymity.

The shared belief that anonymity changes behavior (82.3% agreement) did not predict personal deviance. This null finding is theoretically important. Most people recognize that anonymity can affect behavior; those who are willing to engage in deviance seem to be set apart by a more personal belief that their online and offline identities are psychologically separate. For measurement purposes, items assessing personal identity separation therefore appear to be more informative than items assessing general awareness of anonymity's effects. The inter-item correlations reinforce this distinction: free expression and identity separation were substantially correlated ($r = .63$), whereas the belief that anonymity changes behavior was

essentially uncorrelated with both ($r = .12$ and $.04$). The low composite reliability of the anonymity items reflects multidimensionality rather than item noise, general awareness of anonymity's effects is empirically distinct from personally internalized identity-separation beliefs.

Younger respondents reported substantially higher deviance propensity ($d = 1.44$), a pattern consistent with developmental accounts of still-maturing self-regulation and risk evaluation (Steinberg, 2008) and with cohort differences in digital socialization: respondents who came of age online may perceive sub-criminal digital deviance as more normative than older cohorts do. Because age was measured in categorical brackets, a fine-grained moderation test was not feasible here; whether age moderates the relationship between identity separation and deviance propensity, for example, whether identity separation is a stronger risk marker among younger users, is an important question for future research using continuous age measurement.

Respondents consistently estimated that most average users would engage in deviance, even in scenarios they personally condemned. This cynical norm, or the shared belief that misconduct is the statistical default, may itself act as a permissive signal through the diffusion of responsibility mechanism described in Bandura's (1991) framework of moral disengagement. Across the two most strongly condemned scenarios, approximately 28–30% of respondents endorsed the behavior. This minority appeared across different types of scenario content and warrants additional research.

Implications for Research/Practice

For researchers, the present findings support the use of the projective vignette as a valid indirect measure of deviance propensity. This approach may be especially useful in contexts where direct self-report is vulnerable to social desirability bias or self-presentational concerns. The results also provide clear guidance for instrument development. In particular, anonymity-related items did not perform equally. Measures of dissociative anonymity, defined here as the belief that one's online identity is meaningfully separate from one's offline identity, were more informative than items assessing only general awareness that anonymity can influence behavior. Future research should incorporate this when developing screening instruments and other measures intended to assess risk for online deviance.

The data reveals that anonymity should not be treated as a single factor. The fact that identity separation predicted deviance, whereas general awareness of anonymity's effects did not, indicates that broad measures may obscure the psychological mechanisms most closely associated with behavioral risk. This may have implications not only for measurement, but also for theory development. Studies relying on undifferentiated measures of anonymity may underestimate the role of dissociative processes in shaping deviant behavior. Future work should continue to distinguish between general beliefs about anonymity and more personally internalized beliefs about the separation of online and offline identity.

Directions for future research. Because the projective vignette performed consistently across scenarios, additional studies should examine whether this method retains its validity across different populations, platform environments, and forms of misconduct. Such work would help clarify the conditions under which projective estimates are most useful. In addition, the repeated tendency of respondents to assume that most average users would engage in deviance suggests that perceived norms may warrant greater attention as an independent factor. Future research should examine whether these cynical expectations reflect projection or whether they serve as permissive signals that increase willingness to offend, weaken moral restraint, or reinforce processes of moral disengagement.

For law enforcement, cybersecurity, and platform governance, the hierarchy of deviance identified in this study suggests the need for adjusted intervention strategies. Technical deviance was perceived as both highly common and morally ambiguous, indicating that it may be less responsive to normal messaging and

more responsive to structural solutions. System hardening, friction-based controls, and stronger abuse-prevention mechanisms may be more effective than the appeal to shared norms.

The stable endorsement of social and identity deviance by approximately 30% of respondents under anonymous conditions carries important implications for anonymous forums and other open and loosely monitored environments. This pattern may help explain the persistence of behaviors such as impersonation, deceptive profiling, and other forms of identity-based misuse, including those associated with disinformation efforts and fake professional profiles.

The data also reveals that intervention strategies should be aligned with the specific type of deviance involved. When misconduct is perceived as common and low in moral seriousness, efforts aimed solely at changing attitudes may have limited impact, making opportunity reduction a more promising approach.

Social and identity deviance may require interventions that emphasize accountability, identity verification, moderation, and reductions in anonymity-enabled deception. Threat assessment and early intervention efforts may benefit from focusing less on broad attitudes toward anonymity and more on beliefs that reflect psychological separation between the online and offline self. To the extent that identity separation functions as a stronger marker of deviance propensity, it may offer a more precise basis for identifying elevated risk.

Anonymous or loosely monitored environments may require different design responses depending on the form of harm most likely to occur. Where technical deviance is normalized, platforms may need stronger access controls, greater procedural friction, and more robust system-level safeguards. Where social and identity deviance persist among a stable minority of users, platforms may also require design features that improve traceability, reduce impersonation, and make deceptive identities more difficult to establish and maintain. Taken together, these findings suggest that both research and practice would benefit from moving beyond generalized assumptions about anonymity and toward a more precise understanding of the specific beliefs and conditions that make deviant behavior more likely.

Limitations

Several limitations should be considered when interpreting the present findings. The sample was recruited through a paid panel, was skewed female ($\chi^2 = 4.18, p = .041$), and was highly educated, with 34.7% of respondents reporting a master's degree, which limits generalizability. The modest sample size ($N = 202$) also limits statistical power and the precision of parameter estimates. In addition, the 98.5% mobile completion rate may have affected slider-based responses. Item-level missingness on the anonymity attitude matrix reduced the number of complete cases available for analyses involving those items ($n = 190$). The anonymity attitude items showed low internal consistency as a set ($\alpha = .546$); accordingly, findings involving these items should be interpreted at the item level, and any composite-level anonymity result should be regarded as less stable than the item-level findings reported here.

The median completion time of 2.4 minutes was short for an instrument containing three vignettes. The vignettes were brief (two sentences each) and nearly all respondents completed the survey on mobile devices, where single-tap and slider formats are answered quickly; nonetheless, because no attention-check items were included, insufficient-effort responding cannot be fully ruled out. The robustness analyses reported above indicate that excluding the fastest respondents left the key findings substantively unchanged, and any residual inattentive responding would most plausibly introduce random noise and attenuate the observed associations rather than inflate them.

Nearly all data were collected on a single day (February 20, 2026). Single-day collection could introduce temporal bias if responses were shaped by contemporaneous news coverage related to cybersecurity, misinformation, or online fraud. The authors are not aware of a salient event of this kind coinciding with

the collection window, but the possibility cannot be excluded, and replication across collection periods would strengthen confidence in the estimates.

Measurement limitations should also be acknowledged. Acceptability scale anchors and behavioral intention wording varied slightly across scenarios (see Appendix A), which may have introduced measurement artifacts. More broadly, the use of three specific scenarios may not capture the full range of online deviance, and differences across scenarios may reflect variation in perceived harm, realism, or severity as well as deviance type. The study also relied on vignette responses and self-reported intentions rather than observed behavior, leaving the findings vulnerable to social desirability bias. Because the study used a cross-sectional, correlational design, the findings identify associations among variables but do not permit causal conclusions. Finally, incomplete disclosure was addressed through debriefing, but it remains possible that partial concealment of the study purpose influenced responses.

Conclusions

This study supports the use of a projective vignette methodology grounded in the false consensus effect as a valid indirect approach to assessing online deviance propensity that direct self-report measures may not fully capture. The strongest predictor of deviance willingness was not the general belief that anonymity changes behavior, which was endorsed by most respondents, but the more personal belief that one's online self is meaningfully separate from one's offline identity. This finding places the study's theoretical contribution within Suler's (2004) concept of dissociative anonymity and identifies a potentially useful indicator for future measurement and screening efforts. The hierarchy of deviance observed in the data, in which technical exploitation was viewed as relatively normal while interpersonal deception retained greater moral seriousness, suggests that interventions should be adjusted to the type of deviance involved rather than applied uniformly across contexts.

Data availability:

Survey materials, statistical summaries, and supporting visualizations referenced in this study are publicly available on the project GitHub Pages site here: <https://rmucj.github.io/deviance.propensity.data/>

Acknowledgements

The authors acknowledge the use of generative AI tools (Anthropic Claude) during the preparation of this manuscript. AI assistance was used in three capacities: identifying and locating relevant research literature, reviewing and correcting grammar, and performing statistical analyses on the survey data (including Pearson correlations, linear regressions, and descriptive statistics). All research design, data collection, interpretation of findings, and scholarly conclusions are the sole work of the authors. The authors reviewed and verified all AI-generated output for accuracy prior to inclusion.

References

- Alexander, C. S., & Becker, H. J. (1978). The use of vignettes in survey research. *Public Opinion Quarterly*, 42(1), 93–104. <https://doi.org/10.1086/268432>
- Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45–103). Lawrence Erlbaum.
- Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31(2), 101–119. <https://doi.org/10.1080/0305724022014322>

- Borsari, B., & Carey, K. B. (2003). Descriptive and injunctive norms in college drinking: A meta-analytic integration. *Journal of Studies on Alcohol*, 64(3), 331–341. <https://doi.org/10.15288/jsa.2003.64.331>
- Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions. *Computers in Human Behavior*, 23(6), 3038–3056. <https://doi.org/10.1016/j.chb.2006.09.001>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Higgins, G. E., Fell, B. D., & Wilson, A. L. (2006). Digital piracy: Assessing the contributions of an integrated self-control theory and social learning theory using structural equation modeling. *Criminal Justice Studies*, 19(1), 3–22. <https://doi.org/10.1080/14786010600615934>
- Hinduja, S., & Patchin, J. W. (2008). Cyberbullying: An exploratory analysis of factors related to offending and victimization. *Deviant Behavior*, 29(2), 129–156. <https://doi.org/10.1080/01639620701457816>
- Holt, T. J., & Bossler, A. M. (2014). An assessment of the current state of cybercrime scholarship. *Deviant Behavior*, 35(1), 20–40. <https://doi.org/10.1080/01639625.2013.822209>
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), 434–443. <https://doi.org/10.1016/j.chb.2011.10.014>
- Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, 102(1), 72–90. <https://doi.org/10.1037/0033-2909.102.1.72>
- Perkins, H. W., & Berkowitz, A. D. (1986). Perceiving the community norms of alcohol use among students. *International Journal of the Addictions*, 21(9–10), 961–976. <https://doi.org/10.3109/10826088609077249>
- Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, 123(3), 238–259. <https://doi.org/10.1037/0033-2909.123.3.238>
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279–301. [https://doi.org/10.1016/0022-1031\(77\)90049-X](https://doi.org/10.1016/0022-1031(77)90049-X)
- Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, 28(1), 78–106. <https://doi.org/10.1016/j.dr.2007.08.002>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet*. Simon & Schuster.

- Wall, D. S. (2001). Cybercrimes and the Internet. In D. S. Wall (Ed.), *Crime and the Internet* (pp. 1–17). Routledge.
- Wang, X., Gao, Y., & Chen, B. (2025). Gender differences in cyberbullying tolerance, traditional bullying victimization and their associations with cyberbullying perpetration among Beijing adolescents. *Child & Family Social Work*. Advance online publication. <https://doi.org/10.1111/cfs.13079>
- Wason, K. D., Polonsky, M. J., & Hyman, M. R. (2002). Designing vignette studies in marketing. *Australasian Marketing Journal*, 10(3), 41–58. [https://doi.org/10.1016/S1441-3582\(02\)70157-2](https://doi.org/10.1016/S1441-3582(02)70157-2)
- Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. *Nebraska Symposium on Motivation*, 17, 237–307.
- Zimmerman, A. G., & Ybarra, G. J. (2016). Online aggression: The influences of anonymity and social modeling. *Psychology of Popular Media Culture*, 5(2), 181–193. <https://doi.org/10.1037/ppm0000038>

Appendix A. Survey Instrument

The instrument was administered via SurveyMonkey Audience. Items were presented in a fixed order (no randomization); no attention-check items were included. Forced-response validation was applied to all scenario items. Items Q1–Q3 collected demographics (age bracket, gender, and highest education completed). Region, device type, and household income were captured by the panel provider.

Scenario A — Technical deviance. “A user discovers a configuration error on a media streaming website that allows them to access premium content without paying. They decide to use this error to watch movies for free.”

Q4 (moral acceptability): “How acceptable do you find this behavior?” 7-point scale: Completely Unacceptable, Mostly Unacceptable, Somewhat Unacceptable, Neutral, Somewhat Acceptable, Mostly Acceptable, Completely Acceptable.

Q5 (projective norm): “In your estimation, what percentage of AVERAGE internet users would take advantage of this error?” 0–100 slider.

Q6 (behavioral intention): “If you were in this situation and certain you would not be caught, how likely is it that YOU would do the same?” 7-point scale: Extremely Unlikely to Extremely Likely.

Scenario B — Social deviance. “A user enters an online political forum using an anonymous account. They share a sensationalized, unverified rumor about a candidate to see how people react and to stir up debate.”

Q7 (moral acceptability): “How acceptable do you find this behavior?” 7-point scale: Completely Unacceptable, Unacceptable, Somewhat Unacceptable, Neutral, Somewhat Acceptable, Acceptable, Extremely Acceptable.

Q8 (projective norm): “In your estimation, what percentage of AVERAGE internet users engage in this type of behavior?” 0–100 slider.

Q9 (behavioral intention): “If you were anonymous, how likely is it that YOU would do the same?” 7-point scale presented from Extremely Likely to Extremely Unlikely (response options reverse-ordered relative to Q6 and Q12); responses were coded 1–7 such that higher values indicate greater likelihood, consistent with Q6 and Q12.

Scenario C — Identity deviance. “A user creates a profile on a professional networking site using a fake name and credentials to access industry reports and view competitor profiles without being tracked.”

Q10 (moral acceptability): “How acceptable do you find this behavior?” 7-point scale: Extremely Unacceptable, Unacceptable, Somewhat Unacceptable, Neutral, Somewhat Acceptable, Acceptable, Extremely Acceptable.

Q11 (projective norm): “In your estimation, what percentage of AVERAGE internet users do this?” 0–100 slider.

Q12 (behavioral intention): “If you needed information, how likely is it that YOU would do the same?” 7-point scale: Extremely Unlikely to Extremely Likely.

Anonymity attitudes (Q13). “Please indicate your level of agreement with the following statements” (7-point: Strongly Disagree to Strongly Agree): (a) “I feel that the online space allows me to express myself freely without revealing who I am.” (b) “I believe that my online interactions are largely separate from my real-life identity.” (c) “Anonymity allows people to behave in ways they normally wouldn’t.”

Scoring. The Total Deviance Propensity Score is the mean of the three behavioral intention items (Q6, Q9, Q12), each scored 1–7 with higher values indicating greater willingness to engage in the behavior. Acceptability items were scored 1 (most unacceptable anchor) to 7 (most acceptable anchor). Minor anchor-label differences across scenarios are noted in the Limitations section.