Disengaging from Polarized Debates: Promises and Pitfalls of Expressing Receptiveness on Social Media

Yvan Norotte[a]

Anne-Sophie Chaxel[b]

Author Note

[a]Yvan Norotte is a PhD student at HEC Paris, 1 rue de la libération, 78350 Jouy-en-Josas, France, Email: yvan.norotte@hec.edu

[b]Anne-Sophie Chaxel is an Associate Professor of Marketing at HEC Paris, 1 rue de la libération, 78350 Jouy-en-Josas, France, Email: chaxel@hec.fr.

Disengaging from Polarized Debates: Promises and Pitfalls of Expressing Receptiveness on Social Media

Abstract

Recent research has shown that acts of receptiveness can improve the outcomes of offline and dyadic discussions. In the present research, we more broadly investigate the effectiveness of acts of receptiveness in online environments, particularly on social media platform disengagement, in the presence of highly polarized debates. Three research questions are examined: Do acts of receptiveness help reduce platform disengagement in the presence of polarized debates? If yes, how? Does the presence of malevolent actors (i.e., trolls) significantly impact the effectiveness of acts of receptiveness, driving users instead toward greater platform disengagement? Study 1 shows that acts of receptiveness in a polarizing argument help curb platform disengagement because the viewer feels more receptive in return, akin to a reciprocity process. Study 2 shows that the presence of trolls moderates the relationship between acts of receptiveness and the viewer's experienced receptiveness, thereby negating their positive effect. Study 3 shows that the viewer's experienced receptiveness prevents platform disengagement in the presence of polarization. Finally, Study 4 provides direct evidence for the entire conceptual model. Overall, this research calls for the need to account for digital social dynamics when attempting to foster open dialogs on digital platforms.

Keywords: Receptiveness, Trolling, Platform disengagement, Inferences, Polarization

In a digital world where conflicting positions quickly escalate to the extremes, finding common ground on which to lay the base of a constructive discussion is becoming more difficult than ever. Accordingly, the increase in polarization was ranked as the third highest short-term risk to society by the World Economic Forum in its 2024 Global Risk Report, immediately after extreme weather events and misinformation (World Economic Forum, 2024). Polarization is associated with partisan social dynamics that bring like-minded individuals closer and inhibit the dialog between members of opposite camps (Clark & Winegard, 2020). The effects of polarization are particularly important on social media platforms, as we have seen with the recent backlash around Elon Musk's endorsement of an antisemitic post on X that led numerous advertisers to withdraw from the platform with heavy financial consequences (Mac & Conger, 2023). The presence of polarizing figures on social media can also hurt user experience, and polarizing user activity on social media are deterring people from certain platforms. Therefore, understanding the dynamics of user disengagement in the context of polarizing debates is essential for platforms. In this research, we take the perspective of user interactions within platforms and investigate whether the subjective experience of interactions with disagreeing others can increase users' willingness to disengage from the platform. Previous literature on conflict has shown that sending signals of openness in a debate (i.e., acts of receptiveness) can help restore dialog between two disagreeing individuals (Minson & Chen, 2021). However, we posit that acts of receptiveness may not be as effective in online environments like social media platforms because their characteristics differ drastically from those of common offline settings (Kozyreva et al., 2020). In particular, the presence of malevolent actors such as trolls poses specific challenges to online communication and may drive users away from the platform, i.e., stoke greater platform disengagement. In this research, we ask three research questions: Do acts of receptiveness help reduce platform disengagement in the presence of polarizing debates? If yes, how? Does the presence of malevolent actors (i.e., trolls) significantly impact the effectiveness of acts of receptiveness, driving users instead toward *greater* platform disengagement?

Theoretical Development

Acts of Receptiveness

Acts of receptiveness are defined as "behaviors or actions that signal a person's openness to ideas, arguments, and attitudes that are new or opposing to their own" (Hussein & Tormala, 2021, p. 229) and can take various forms, including asking elaboration questions (Chen et al., 2010), conveying uncertainty (Tormala, 2016), or actively listening (Itzchakov et al., 2023). For instance, to promote dialog with a disagreeing counterpart, one can show receptiveness by conceding "I cannot be entirely sure, but I believe that" before sharing one's opinion. Drawing insights from the social relations model (Back & Kenny, 2010), Minson & Chen (2021) theorized a process akin to a transfer of receptiveness where perceived acts of receptiveness from a disagreeing other induce individuals to experience more receptiveness themselves, thereby helping the discussion converge toward a certain level of open dialog. Through this virtuous circle, various types of acts of receptiveness were shown to be successful in boosting interpersonal evaluations and willingness to talk in the context of disagreements (Minson et al., 2023), helping to impair conflict escalation, and boosting future collaborative intentions between disagreeing parties (Yeomans et al., 2020). Based on the receptiveness literature, we thus propose the following hypothesis, replicating prior research in offline environments to online environments:

> Hypothesis 1: Individuals are more likely to experience receptiveness in the social media context when their counterpart uses acts of receptiveness (versus not).

However, existing empirical evidence supporting the transfer of receptiveness is limited to interactions between two individuals (Back & Kenny, 2010; J. A. Minson & Chen, 2021). In this research, we instead consider the broader context of public online discussions akin to forums or comment sections. This setting differs from prior research settings in two ways: a) information is automatically shared to the entire community, and b) anyone can jump into the discussion (unlike in a dyadic dialog). Due to the public nature of the interaction, the ways in which users are perceived by others can drive their behavior on the platform. Specifically, we take a motive-based approach by arguing that public online interactions allow for the proliferation of new actors whose presence in the environment changes the nature of user interactions by stirring suspicion of ulterior motives.

Trolling

An important challenge associated with communication in the digital world is the presence of individuals engaging in trolling: a "form of online harassment that involves posting provocative and inflammatory messages in order to disrupt the conversation and upset other people" (Kozyreva et al., 2023, p. 86). Trolling adds a new hidden actor whose actions are not aimed at defending or promoting a specific belief but at disrupting the conversation itself, and in the presence of trolls, receiving acts of receptiveness from another user may be perceived as manipulative. Drawing on insights from the persuasion knowledge model (Friestad & Wright, 1994), we propose that acts of receptiveness in the context of disagreement could be perceived as manipulation cues, thereby deterring other users from reciprocating (Campbell & Kirmani, 2000; Kirmani & Zhu, 2007). In other words, as the suspicion of an ulterior motive is likely to lead to negative attitudes (DeCarlo, 2005), we hypothesize that interacting in an online environment where trolls are known to be active should severely reduce the benefits of acts of receptiveness on experienced receptiveness.

> Hypothesis 2: The effect of acts of receptiveness on experienced receptiveness is lower in a trolling (versus helpful) forum.

Platform disengagement

Digital platforms are essential players in the digital world because they provide the means for online interactions. However, platforms are also independent actors that serve their own purposes. In this research, we investigate how platform objectives can be affected by the nature of their users' interactions in a polarized context. Specifically, we focus on a key driver of platform profitability: user engagement and, more specifically, user *disengagement* following exposure to conflicting views in a polarized context. Polarized interactions on platforms often turn into partisan clashes where users from each side are more preoccupied with defending their opinions and the people sharing them than opening up to dialog (Finkel et al., 2020; Wilson et al., 2020). Past research suggests that positive experiences of interactions with disagreeing others can counterbalance such polarization dynamics and receptive conversations between disagreeing users were shown to increase willingness to engage in future conflicting interactions (Minson & Chen, 2021; Yeomans et al., 2020). Therefore, users experiencing receptiveness during polarizing interactions should be more inclined to engage in future conflicting interactions and therefore less likely to disengage from the platform. In other words, we expect the level of receptiveness experienced in an online conversation with disagreeing others to determine disengagement from the platform.

> Hypothesis 3: Experienced receptiveness prevents platform disengagement.

Overview of studies

All studies were preregistered. Study 1 provides initial support for the transfer of receptiveness (H1) along with evidence for its benefits for platform disengagement intentions (H3). Study 2 provides additional support for the transfer of receptiveness using different stimuli (H1) and introduces a trolling condition to document its moderating effect on experienced receptiveness (H2). Study 3 tests the predictive power of experienced receptiveness on platform disengagement behaviors against alternative explanations (H3) . Together, Study 2 and Study 3 provide independent empirical evidence for the two parts of the conceptual model (Figure 1). Consistent with these results, Study 4 provides direct evidence for the entire conceptual model by testing all the hypotheses simultaneously using moderated mediation (H1, H2, H3).

Study 1: Acts of receptiveness in polarizing discussions prevent platform disengagement

Method

Procedure. A total of 299 US residents (M_{Age} = 42.15, SD = 14.67, 61% women) were recruited on Prolific in exchange for $.4. Participants first indicated whether they believed the government should prioritize mitigating over adapting to climate change (yes/no).)". Participants were then asked to imagine they were browsing a forum where they eventually posted a message to defend their opinion on the focal topic and received an answer coming from a very active user, user H, which consisted of a counter-attitudinal argument. Both arguments (in favor and against the focal topic) were pretested to be similarly persuasive. Participants then indicated agreement with nine statements on a 7-points Likert scale from 1 (strong disagreement) to 7 (strong agreement). The first four served as a measure of disengagement intentions (α = .94; adapted from Villanova & Matherly, 2023), the following three served as measure of experienced receptiveness (α = .85; extracted from Chen et al., 2010) and the last two as manipulation check.

Design. Participants were randomly assigned to one of two conditions (Arguments: AOR vs control) in a between subject design. We manipulated the answer from user H by either preceding the counter-attitudinal argument with an act of receptiveness: "I hadn't thought about it in that way, I may be open to reconsider my stance, because for me" (AOR condition) or not (no-AOR condition).

Results

Platform disengagement. The manipulation was successful. Consistent with H1, participants' mean disengagement intentions were significantly greater when receiving a counter-attitudinal argument alone (M_{no-AOR} = 3.68, SD = 1.64) than when the source of the argument also included an act of receptiveness (M_{AOR} = 3.06, SD = 1.37; t(296.88) = 3.66, p < .001). Consistent with H3, this effect was mediated by experienced receptiveness as both the effect of act of receptiveness on experienced receptiveness (a = .90, t(297) = 5.44, p < .001) and the effect of experienced receptiveness on platform disengagement when controlling for acts of receptiveness (b = -.63, t(296) = 13.57, p < .001) were significant. When controlling for experienced receptiveness, acts of receptiveness no longer predicted platform disengagement (c'= -.05, t(296) = 0.32, p = .75), indicating support for a full mediation. Using Monte Carlo bootstrapping (5000 simulations), the estimated indirect effect was -.571, with a 95% confidence interval between -.80 and -.35.

Study 2: The presence of trolling disrupts the transfer of receptiveness

Method

Procedure. We recruited 484 participants (M_{Age} = 39.92, SD = 12.82, 49% women) on Prolific in exchange for \$.4. The procedure was identical to Study 1 except from two changes: a) we used a different focal topic "Do you believe fossil fuel-based energy should be replaced with renewable energy? (yes/no)", and b) participants only answered the experienced receptiveness (α = .87) and manipulation checks questions.

Design. Participants were randomly assigned to one of four conditions in a 2 (condition: trolling vs. control) × 2 (arguments: AOR vs. no-AOR) between-subjects design. Before receiving the answer from user H, participants in the trolling condition were told that while browsing the forum, they had noticed that some active users were trying to waste others' time by engaging in sterile debates while maintaining a pretense of civility and commitment to reasonable debate (a common type of trolling referred to as sealioning; Kozyreva et al., 2023). In the control condition, those same users were described as genuinely engaging with others. Upon receiving the answer, the counter-attitudinal argument in the AOR condition started with "I'm fascinated by different opinions. This is very interesting, tell me, what led you to form such a unique perspective?". In the no-AOR condition, participants only saw the counter-attitudinal argument.

Results

Experienced receptiveness. Both receptiveness and trolling manipulations were successful. Consistent with H1 and H2, an ANOVA with experienced receptiveness as the dependent variable and both AOR types and condition as factors revealed a significant interaction ($F(1,480)$ = 5.90, p = .02). In the control condition, participants felt more receptive to opposing views when the respondent used an act of receptiveness ($M_{control-AOR}$ = 4.41, $SD_{control-AOR}$ = 1.48 vs. ($M_{control-noAOR}$ = 3.45, $SD_{control-noAOR}$ = 1.63; t(240.34) = 4.82, p <0.001). However, this effect disappeared in the trolling condition ($M_{trolling-AOR}$ = 3.89, $SD_{trolling-AOR}$ = 1.50 vs. $M_{control-AOR}$ 3.62, $SD_{control-AOR}$ = 1.69; t(228.87) = 1.28, p = .20). The results are displayed in Figure 2.

Study 3: Experienced receptiveness prevents platform disengagement

Method

Procedure. 399 US resident (M_{Age} = 40.81, SD = 12.71, 50% women) recruited on Connect participated in a 4-minutes long study for \$.6. After indicating their belief on a range of hot societal topics, including the focal topic of the study (drug rehab), participants were told the aim of the experiment was to gather their feedback on a new platform specialized in debate. After reading a short description of the platform, participants were informed they had been added to a debate occurring on the platform. Upon entering the debate, participants saw a pro-attitudinal post on the focal topic of the study, followed by two disagreeing comments from another user (user K). After reading the posts, participants decided whether they wanted to keep following or unfollow the debate (binary choice) and rated the platform on a Likert scale from 1 (Extremely bad) to 7 (Extremely good). Participants then completed a measure of mood (α = .85; extracted from Cutright et al., 2011), a measure of liking of user K using the feeling thermometer (Iyengar et al., 2012), and a measure of experienced receptiveness (α = .82; extracted from Chen et al., 2010).

Design. The experiment was correlational and only included one condition.

Results

Unfollowing. 175 participants (43.8%) chose to unfollow the debate on the platform. As preregistered, we ran a logistic regression with the unfollowing choice as dependent factor (coded as 0 = keep following and 1 = unfollowing), and experienced receptiveness, mood, liking, and their interactions as independent factors. Consistent with H3, experienced receptiveness was negatively related with unfollowing choice (β = -1.35, z = -7.57, p < .001).

Platform rating. The average rating of the platform was 4.84 (1.21). Consistent with H3, a linear regression with unfollowing choice as dependent factor and experienced receptiveness, mood, liking, and their interactions as independent factors found a positive correlation between experienced receptiveness and platform rating ($\beta = .51$, $t(391) = 7.417$, $p < .001$). In other words, the receptiveness experienced on a platform positively predicts the evaluation of the platform itself, above and beyond mood and liking.

Study 4: Testing the full conceptual model

Method

      Procedure. As preregistered, 600 US residents ($M_{Age} = 42.22$, SD = 14.50, 58% women) were recruited on Prolific to complete our study in exchange for $.4. The procedure was identical to Study 2 with only two changes: the focal topic (police violence) and we added a measure of platform disengagement intentions (the same as in Study 1).

      Design. As in Study 2, participants were randomly assigned to one of four conditions in a 2 (condition: trolling vs. helpful) × 2 (acts of receptiveness: AOR vs. no-AOR) between-subjects design.

Results

      Platform disengagement. Investigating the full conceptual framework as one, we ran a moderated mediation using PROCESS Model 7 (Hayes, 2013) with acts of receptiveness as the dependent variable, platform disengagement as the independent variable, experienced receptiveness as the mediator and condition as the moderator. As expected, the potential for acts of receptiveness to induce experienced receptiveness varied in each condition (see **Figure 3**). Specifically, when interacting in a troll forum, the presence of an act of receptiveness failed to increase experienced receptiveness (a $_{trolling}$ = .18, $t(596) = 0.98$, $p = .33$), while the effect remained significant in the helpful forum (a $_{helpful}$= .86, $t(596) = 4.63$, $p < .001$). The inability of acts of receptiveness to reduce platform disengagement in the trolling forum can thus be explained by a failure to induce experienced receptiveness. Conversely, acts of receptiveness successfully reduced platform disengagement in the helpful forum by increasing experienced receptiveness, with an estimated indirect effect of -.52 and a bootstrapping 95% confidence interval (5000 iterations) ranging between -.76 and -.30. Overall, the moderated mediation model was significant, with an index of moderated mediation of .42 and a bootstrapping 95% confidence interval (5000 iterations) ranging from .11 to .74.

Discussion

      This research shows that the presence of trolls inhibits the effectiveness of acts of receptiveness in facilitating polarized interactions. As such, we demonstrated that the determinants of experienced receptiveness online are different from those of private interactions and will therefore require different interventions to reduce polarization. In addition, we delineated the role of experienced receptiveness in platform disengagement, thereby revealing it as a key concept for both platform profitability and polarization reduction. While past research has shown how unacceptable language can drive people supporting the same brand to dissociate from the group and the brand itself (Villanova & Matherly, 2023), this research shows how positive language in the form of acts of receptiveness coming from people of the opposite belief camp can also drive disengagement due to the mere presence of bad actors in the environment. The same way brand disengagement can have negative repercussions for businesses in terms of consumer reach (Lee et al., 2018), disengagement from belief-exchange platforms reduces exposure to opposing views. These results contribute to extending our understanding of digital social dynamics.

References

Back, M. D., & Kenny, D. A. (2010). The Social Relations Model : How to Understand Dyadic Processes. *Social and Personality Psychology Compass*, *4*(10), 855-870. https://doi.org/10.1111/j.1751-9004.2010.00303.x

Campbell, M. C., & Kirmani, A. (2000). Consumers' Use of Persuasion Knowledge : The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent. *Journal of Consumer Research*, *27*(1), 69-83. https://doi.org/10.1086/314309

Chen, F. S., Minson, J. A., & Tormala, Z. L. (2010). Tell me more : The effects of expressed interest on receptiveness during dialog. *Journal of Experimental Social Psychology*, *46*(5), 850-853. https://doi.org/10.1016/j.jesp.2010.04.012

Clark, C. J., & Winegard, B. M. (2020). Tribalism in War and Peace : The Nature and Evolution of Ideological Epistemology and Its Significance for Modern Social Science. *Psychological Inquiry*, *31*(1), 1-22. https://doi.org/10.1080/1047840X.2020.1721233

Cutright, K. M., Wu, E. C., Banfield, J. C., Kay, A. C., & Fitzsimons, G. J. (2011). When Your World Must Be Defended : Choosing Products to Justify the System. *Journal of Consumer Research*, *38*(1), 62-77. https://doi.org/10.1086/658469

DeCarlo, T. E. (2005). The Effects of Sales Message and Suspicion of Ulterior Motives on Salesperson Evaluation. *Journal of Consumer Psychology*, *15*(3), 238-249. https://doi.org/10.1207/s15327663jcp1503_9

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533-536. https://doi.org/10.1126/science.abe1715

Friestad, M., & Wright, P. (1994). The Persuasion Knowledge Model : How People Cope with Persuasion Attempts. *Journal of Consumer Research*, *21*(1), 1-31. https://doi.org/10.1086/209380

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition : A Regression-Based Approach*. Guilford Publications.

Hussein, M. A., & Tormala, Z. L. (2021). Undermining Your Case to Enhance Your Impact : A Framework for Understanding the Effects of Acts of Receptiveness in Persuasion. *Personality and Social Psychology Review*, *25*(3), 229-250. https://doi.org/10.1177/10888683211001269

Itzchakov, G., Weinstein, N., Leary, M., Saluk, D., & Amar, M. (2023). Listening to understand : The role of high-quality listening on speakers' attitude depolarization during disagreements. *Journal of Personality and Social Psychology*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/pspa0000366

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, Not Ideology : A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, *76*(3), 405-431. https://doi.org/10.1093/poq/nfs038

Kirmani, A., & Zhu, R. (Juliet). (2007). Vigilant against Manipulation : The Effect of Regulatory Focus on the Use of Persuasion Knowledge. *Journal of Marketing Research*, *44*(4), 688-701. https://doi.org/10.1509/jmkr.44.4.688

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens Versus the Internet : Confronting Digital Challenges With Cognitive Tools. *Psychological Science in the Public Interest*, *21*(3), 103-156. https://doi.org/10.1177/1529100620946707

Kozyreva, A., Wineburg, S., Lewandowsky, S., & Hertwig, R. (2023). Critical Ignoring as a Core Competence for Digital Citizens. *Current Directions in Psychological Science*, *32*(1), 81-88. https://doi.org/10.1177/09637214221121570

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising Content and Consumer Engagement on Social Media : Evidence from Facebook. *Management Science*, *64*(11), 5105-5131. https://doi.org/10.1287/mnsc.2017.2902

Mac, R., & Conger, K. (2023, novembre 24). X May Lose Up to $75 Million in Revenue as More Advertisers Pull Out. *The New York Times*. https://www.nytimes.com/2023/11/24/business/x-elon-musk-advertisers.html

Minson, J. A., & Chen, F. S. (2021). Receptiveness to Opposing Views : Conceptualization and Integrative Review. *Personality and Social Psychology Review*, 10888683211061037. https://doi.org/10.1177/10888683211061037

Minson, J., Hagmann, D., & Luo, K. (2023). *Cooling heated discourse : Conversational receptiveness boosts interpersonal evaluations and willingness to talk*. OSF Preprints. https://doi.org/10.31219/osf.io/5w3dg

Tormala, Z. L. (2016). The role of certainty (and uncertainty) in attitudes and persuasion. *Current Opinion in Psychology*, *10*, 6-11.

Villanova, D., & Matherly, T. (2023). For Shame ! Socially Unacceptable Brand Mentions on Social Media Motivate Consumer Disengagement. *Journal of Marketing*, 00222429231179942. https://doi.org/10.1177/00222429231179942

Wilson, A. E., Parker, V. A., & Feinberg, M. (2020). Polarization in the contemporary political and media landscape. *Current Opinion in Behavioral Sciences*, *34*, 223-228. https://doi.org/10.1016/j.cobeha.2020.07.005

World Economic Forum. (2024). *Global Risks Report 2024*. https://www.weforum.org/publications/global-risks-report-2024/digest/

Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness : Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, *160*, 131-148. https://doi.org/10.1016/j.obhdp.2020.03.011

Figure Captions

Figure 1. Conceptual Framework of the Effect of Acts of Receptiveness on Platform Disengagement through Platform Disengagement.

Figure 2. The Influence of Acts of Receptiveness on Experienced Receptiveness in Study 2. Bars represent 95% confidence intervals of the means. **: $p < .01$; ***: $p < .001$.

Figure 3. Moderated Mediation Model of Study 4. Coefficients are unstandardized regression coefficients; *: $p < .05$; brackets: bootstrapping 95% confidence intervals (5000 iterations).
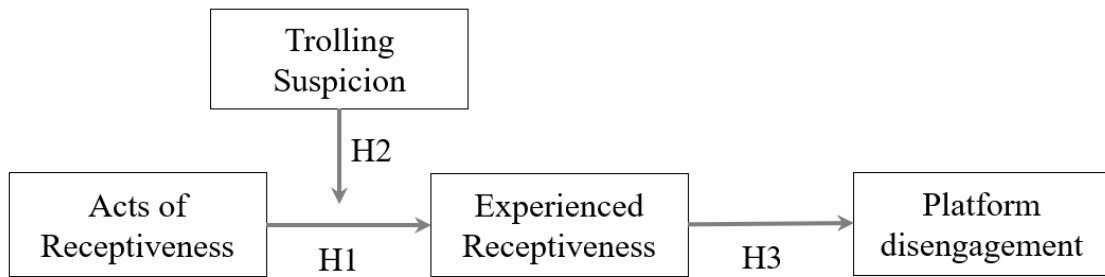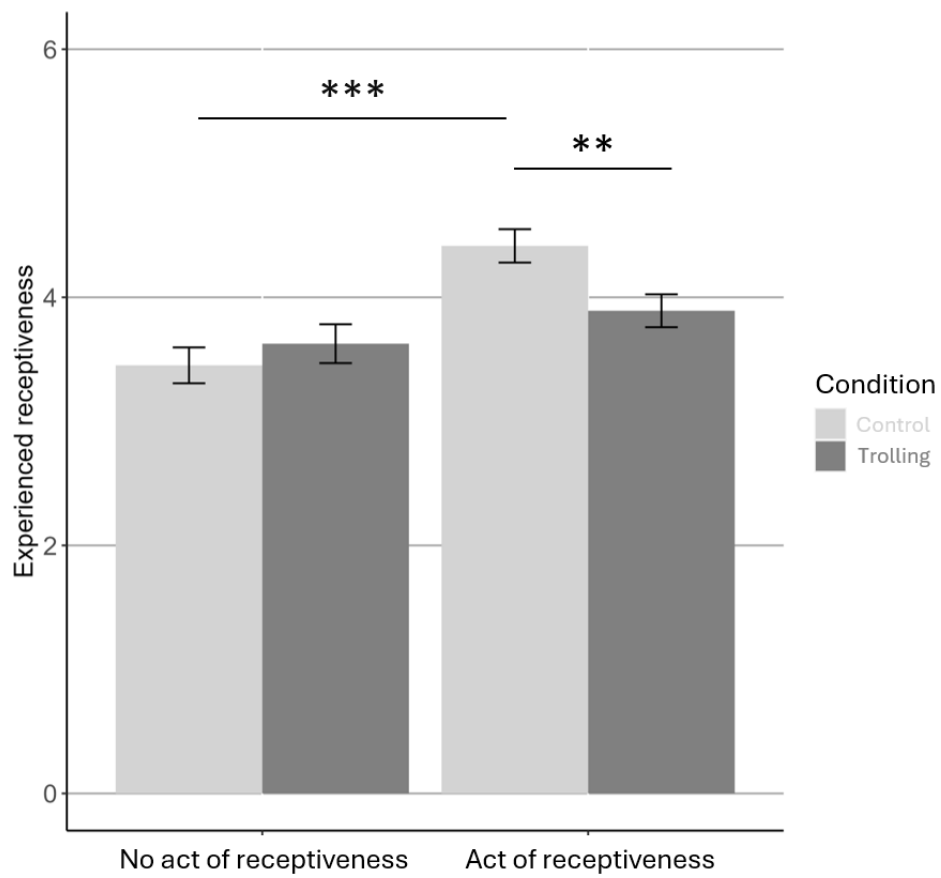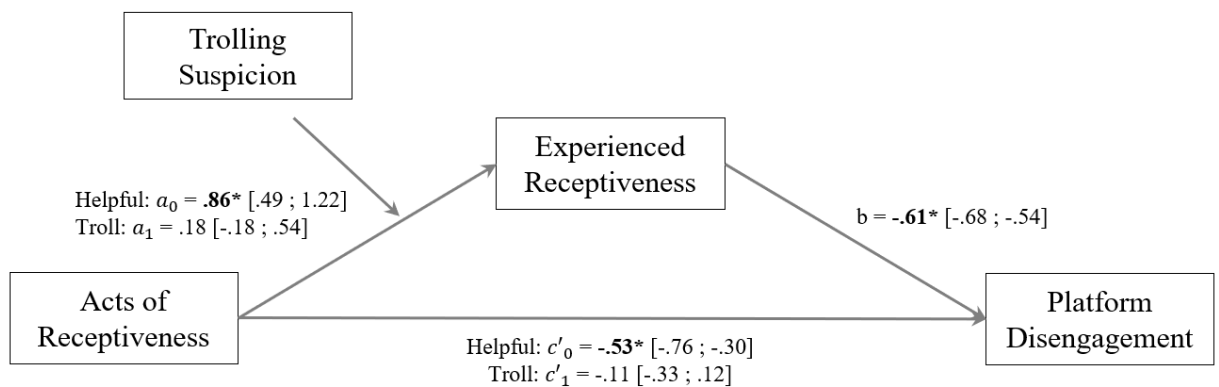
Figures

Figure 1

Figure 2

Figure 3