

**Performative Morality and Social Perception: Effects of Moral Grandstanding on
Perceived Likability**

Eline M. L. Jansen

S5199522

Department of Psychology, University of Groningen

PSB3E-BT15.2024-2025.3: Bachelor thesis

Group number: 27

Supervisor: dr. E. W. Meerholz

Second evaluator: dr. T. Kuppens

In collaboration with: Antonia Jager, Berk Muratoglu, Daan van der Veen, Lilly Schönleiter,
and Linde van Loon

June 26, 2025

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned

Abstract

This study investigates the social consequences of moral grandstanding, a recently defined concept, which refers to the use of public moral talk with the intention to gain status and social approval (Tosi & Warmke, 2016). It explores the effects of moral grandstanding on the perceived likability of a speaker and examines whether the observer's agreeableness moderates this relationship. In a between-subjects experiment, participants were randomly assigned to read a neutral or moral grandstanding comment within a simulated online discussion on gene modification. Agreeableness was measured using self-assessment, and the likability of the commenter was rated after exposure to the experimental manipulation. A manipulation check confirmed that our manipulation was effective. Results showed that commenters in the moral grandstanding condition were perceived as significantly less likable than those in the control condition. Against expectations, agreeableness did not moderate this relationship. These findings underscore the interpersonal costs of moral grandstanding. Although individuals use strategic moral discourse to boost their own image, such behavior may instead provoke resentment and reduce their likability in others' eyes.

Keywords: moral grandstanding, likability, agreeableness, interpersonal relationships, online discourse

Performative Morality and Social Perception: Effects of Moral Grandstanding on Perceived Likability

Susan B. Anthony was an American women's rights activist who was arrested and convicted in 1872 for voting illegally. In August 2020, President Donald Trump announced that he would pardon Anthony and stated that 'It should never have taken so long.' Many construed this action to be disingenuous, given Trump's history of misogynistic remarks. His pardon was viewed as a political move attempting to garner female voters by displaying a virtue that he actually did not possess (Rosa, 2023; Stollznaw, 2020). The term *virtue signaling* has often been used online to describe Trump's behavior (Bishara, 2020; Stollznaw, 2020), referring to public moral discourse aimed at boosting one's social status (McClay, 2018). While the term originated online, it has been subject to scientific investigation ever since. However, according to modern philosophers Tosi and Warmke (2016), it is too vague to use as a scientific concept, mainly because the word 'signaling' is inherently ambiguous. Instead, they propose *moral grandstanding*¹, which they define as the strategic use of moral talk to shape others' perceptions and judgements in order to gain social approval.

Moral grandstanding is a widespread phenomenon in modern society, occurring primarily online but also in everyday interactions and political settings (Tosi & Warmke, 2016). Given its rise in relevance, further investigation of this concept and its implications for our social interactions is essential. Grandstanding has been said to have far-reaching social consequences, such as contributing to polarization or undermining moral discourse (Tosi & Warmke, 2016). Yet, all such outcomes begin with the basic social process of how the grandstander is perceived by others. One crucial dimension of this perception is perceived

¹ The terms 'moral grandstanding' and 'grandstanding' are being used interchangeably in this report, they refer to the same concept.

likability, which can act as a heuristic in social judgement and can influence attitude formation of individuals (Zajonc, 1980). Understanding whether grandstanding increases or decreases likability is therefore a key first step in understanding its social impact. Further, individual differences in personality traits, mainly *agreeableness*, a personality trait reflecting an individual's cooperative nature, may moderate the way grandstanding is perceived (Graziano & Tobin, 2018). If likability is indeed an important factor in the context of moral grandstanding, then it is also important that we look at the factors that influence the relationship between grandstanding and likability. Therefore, the current study explores how moral grandstanding affects an observer's perception of the grandstander's likability, with the observer's agreeableness serving as a moderating factor.

Likability

Likability, which refers to the quality of being pleasant and easy to get along with, is a well-studied psychological construct with roots in evolution. Being well liked increased survival chances in prehistoric times, because it led to an individual's acceptance into social groups, which in turn provided access to resources and protection (Hare, 2016). Humans evolved to have a rather strong radar for picking up social cues to make a judgement about someone's likability. This too can be explained from an evolutionary perspective, since recognizing social cues allowed people to identify potential threats and filter out the untrustworthy among them (Hare, 2016). To this day, likability remains an influential and relevant variable in both the formation and maintenance of relationships. Likability can function as a mental shortcut in the formation of attitudes towards a person (Zajonc, 1980). Additionally, Pulles and Hartman (2017) found that individuals who are perceived as likable are more likely to be trusted and included in future interactions, which is crucial for building and maintaining relationships.

Former studies have shown that someone's perceived likability is influenced by a number of factors, three of which are particularly relevant to mention in this context. The first aspect that has been linked to a higher perceived likability is physical attractiveness. The so-called 'what is beautiful is good' stereotype states that people are more likely to attribute positive traits such as kindness, intelligence, and competence to physically attractive individuals. As a result, physically attractive people tend to be perceived as more likable (Dion et al., 1992). Secondly, vocal characteristics such as tone, pitch, and speech rate have been shown to influence how likable a speaker is perceived to be. Research has found that speakers who spoke more quickly and had a darker timbre, characterized by a smoother, deeper tone, were perceived as more likable (Weiss & Burkhardt, 2010). Thirdly, individuals who are viewed as authentic tend to be perceived as more likable, because authenticity fosters trust and relatability (Rivera et al., 2024; Rivera, 2020). In their experiments, Rivera et al. (2024) found evidence of a bidirectional relationship between authenticity and likability; highly likable people were perceived as more authentic and vice versa. Given the pivotal role perceived authenticity plays in determining perceptions of likability, it seems plausible that moral grandstanding, in which inauthenticity is a key component, is another factor that might strongly influence likability.

Moral Grandstanding

Moral grandstanding is the use of public moral talk, intended to shape others' perspective of oneself. Grandstanders engage in public moral discussion, not with the goal to express their views, but to elevate themselves above others. As Tosi and Warmke (2016) put it, "To grandstand is to turn one's contribution to public discourse into a vanity project" (p. 199). People engaging in moral grandstanding often seek to be seen as worthy of respect or even admiration for possessing a certain moral virtue, whether that be a heightened sense of empathy or keen moral awareness. Tosi and Warmke (2016) argue that moral grandstanding

is a diffuse social phenomenon comparable with other complex social phenomena such as love, blame, apology, and forgiveness. It is expressed through various behaviors and most commonly characterized by five key features: (1) *piling on*: repeating an argument without adding meaningful insight, showing little lack of actual engagement; (2) *ramping up*: making increasingly strong comments to outdo others and appear more morally committed; (3) *trumping up*: exaggerating an issue's severity to increase its moral significance; (4) *excessive emotional displays*: using strong emotion to signal deep moral involvement; and (5) *claims of self-evidence*: the tendency of insisting one's moral views as so obviously correct that disagreement reflects ignorance or immorality (Tosi & Warmke, 2016). Moral grandstanding manifests itself in different ways, whether that be in just one of the behaviors mentioned above or any combination of them.

Grandstanding has been said to have three key problematic effects. Firstly, Tosi and Warmke (2016) proposed that it is likely to lead to an unhealthy amount of cynicism about moral discourse. People may view moral discourse as performative rather than sincere, leading to distrust and increased cynicism. Secondly, the researchers suggested that moral grandstanding results in what they call 'outrage exhaustion.' This refers to people becoming desensitized to moral outrage. As mentioned above, grandstanding often manifests itself in the expression of strong emotions. An abundance of outraged expressions could result in individuals experiencing difficulties in recognizing or feeling genuine outrage when it is actually needed. Thirdly, Tosi and Warmke (2016) propose that it fosters group polarization, which refers to the tendency of a deliberating group to adopt more extreme positions over time. Behaviors like ramping up and trumping up encourage people to outdo each other, pushing like-minded people towards more extreme versions of their initial moral convictions. Group polarization in turn can lead to escalation of discussions between groups, which results in unproductive and uncompromising dialogue (Sunstein, 2002; Tosi & Warmke, 2016). In a

later study, Grubbs et al. (2020) found that *prestige-motivated grandstanding*, a subtype of moral grandstanding aimed specifically at gaining admiration, was consistently associated with more extreme ideological views and a greater affective polarization. The latter refers to stronger negative feelings towards opposing parties. They proposed that grandstanders often express exaggerated emotions that fuel polarization and agitation, which can deepen division between groups. *Dominance-motivated grandstanding*, the other subtype of moral grandstanding that has been established in the literature, where the aim is to dominate and intimidate others, was not related to polarization (Grubbs et al., 2020).

Moral grandstanding may negatively affect how likable a person is perceived to be. Moral grandstanding involves prioritizing self-promoting over the expression of genuine moral convictions when making a public moral statement (Tosi & Warmke, 2016). Due to their highly evolved nature for assessing someone's authenticity, people are likely to detect these self-serving motivations, leading to them viewing the other person as insincere and untrustworthy (Hare, 2016). Inauthentic behavior could undermine factors that underlie likability, like trust and feelings of connectedness. Trust, a core component of likability, can quickly fade when people sense that someone's actions are not driven by true moral conviction, but rather by the desire for external validation (Rivera et al., 2020). Furthermore, likability is closely tied to feelings of social connectedness and shared values (Rivera et al., 2024; Rivera, 2020). When people notice someone is partaking in moral grandstanding, it might create a distance between the observer and the grandstander, as it disrupts the honest and mutually respectful exchange of ideas. As a result, it is hypothesized that engaging in moral grandstanding will lead to lower likability ratings, as it undermines key relational factors such as trust, authenticity, and the sense of genuine connection with others (hypothesis 1). However, we believe this relationship may be moderated by another factor, namely agreeableness.

Agreeableness

Agreeableness is one of the five major personality domains in the Big Five personality traits (Goldberg, 1992; McCrae & Costa, 1999). Of these five domains, agreeableness is the one most concerned with individual differences in interpersonal behaviors. Overall, people who score high on agreeableness are altruistic, sympathetic, and cooperative (Graziano & Tobin, 2009). The vast quantity of literature on agreeableness shows the many consequences that agreeableness has in interpersonal interactions. A meta-meta analysis by Wilmot and Ones (2022) establishes eight general themes that describe the characteristics of agreeableness. These themes are (1) *self-transcendence*: aspirations to grow as a person; (2) *satisfaction*: the acceptance of and the adjustment to life; (3) *relational investment*: successes in the maintenance of positive relationships; (4) *teamwork*: smooth and effective coordination with others; (5) *work investment*: a great willingness to put forth effort; (6) *lower outcomes emphasis*: lower productivity; (7) *social norm orientation*: norm awareness and compliance; and (8) *social integration*: integrability into social roles and institutions.

People who score high on agreeableness are possibly more susceptible and more prone to manipulation; three key reasons for this proposed relation will be discussed. Firstly, Oyibo and Vassileva (2019) found that highly agreeable individuals are more influenced by social proof and comparison, as they tend to conform to social norms to maintain harmony and avoid conflict. Secondly, Evans and Revelle (2008) found that people who scored higher on agreeableness reported more trust in others when surveyed, although their actual behavior did not always match their self-reported trust. Thirdly, Ezrim et al. (2021) found that agreeableness predicted increased overall trust in others in a Nigerian student sample.

The effect of moral grandstanding on likability may depend on the observer's level of agreeableness. Overall, highly agreeable individuals tend to trust others and are less cynical or skeptical about others' intentions compared to those low in agreeableness (Ezrim et al.,

2021; Oyibo & Vassileva, 2019; Revelle, 2008). Although trust is generally a positive trait, it can become problematic when others engage in deceptive behaviors, such as moral grandstanding. We propose that agreeableness moderates the relationship between moral grandstanding and likability: highly agreeable individuals, being less skeptical, are more likely to take statements at face value and may overlook the covert intentions behind grandstanding. In contrast, individuals low in agreeableness are more likely to question the sincerity behind moral statements. Therefore, the impact of moral grandstanding on likability is expected to depend on the observer's level of agreeableness. Specifically, for people who score low on agreeableness, we expect that grandstanding results in lower likability ratings. For people who score high on agreeableness, we expect this negative effect to be much smaller, as they are less likely to interpret grandstanding negatively (hypothesis 2).

Present Research

We believe the current study contributes to the literature in three key ways. Firstly, it extends research on the phenomenon beyond the American cultural context. Up to the present, nearly all empirical research on moral grandstanding has been done in the USA (Grubbs et al., 2019; Grubbs et al., 2020; Grubbs et al., 2022). By examining the phenomenon in a different cultural context, the current research offers a broader understanding of it. Secondly, it will contribute to the general understanding of the phenomenon of moral grandstanding, a rather novel concept that is not explored enough up to the present. Most prior work on the subject has been philosophical in nature. There have only been a handful of key empirical studies on the topic, which makes this study a useful and important complement to the existing literature. Thirdly, it will broaden the understanding of moral grandstanding in a more specific context: the effect it has on interpersonal interactions. The only study that looked at the interpersonal consequences of moral grandstanding was done by Savejnaron et al. (2022). The other studies on moral grandstanding have looked at individual

predictors or motivations of partaking in moral grandstanding (Grubbs et al., 2022) and the effect it has on discourse in general (Grubbs et al., 2020). Since there have not been many studies regarding the effect that moral grandstanding has on interpersonal interactions, this study will make for a valuable addition to the literature.

Method

Participants & Design

Initially, the dataset consisted of 254 cases. In total, 106 cases were excluded from the study. This was done for several reasons. Firstly, 98 incomplete cases were removed from the data set. The seriousness check at the end of the survey was set as the criterion to decide whether a case was seen as incomplete: if a participant did not fill out the seriousness check, we could not verify whether they took part seriously or not. They were therefore excluded due to incompleteness. It is important to note that this group of considerable size included participants who viewed only the initial survey screen or exited early to complete the survey at a later point. Of the 98 removed, only 23 made it to the manipulation, and only 11 actually filled out any of the dependent variable items. This indicates that most never meaningfully engaged with the study or exited it to complete it later on, rather than dropping out midway. Secondly, of the remaining 156 cases, 2 were removed from the data set as they indicated that they had not participated seriously on a seriousness check, rendering their data unusable. Lastly, of the remaining 154 cases, 6 of them who no longer wanted their data to be used in our study after they were debriefed about its true purpose were also removed based on whether they indicated this wish. After exclusions, the final sample consisted of 148 participants (84 female, 61 male, and 3 other/prefer not to say), with ages ranging from 18 to 63 years ($M = 27.24$, $SD = 10.41$).

Participants were recruited using convenience sampling. Initially, the researchers reached out to their personal networks and shared the link to the online survey. The link to the

survey was shared via WhatsApp and posted online, for example, on Instagram and Reddit. When introducing the study, participants were told it was about the way online discussions were experienced; the true purpose of the study was therefore not disclosed but concealed with a cover story.

This study employed a between-subjects design, in which participants were randomly assigned to one of the two conditions of the independent variable, moral grandstanding: the moral grandstanding condition or the control condition. Likability served as the dependent variable and agreeableness as the moderator. This study was part of a bigger project, where additional variables were included; these are beyond the scope of this paper and will therefore not be discussed here. A full list of all the used variables can be found in Appendix A.

Materials & Procedure

The Ethics Committee of the Faculty of Behavioral and Social Sciences at the University of Groningen approved the study after reviewing it. The survey used for the online study was designed using Qualtrics. Participating in the study took about fifteen minutes, and participation was voluntary, meaning that the participants were able to stop the survey at any time.

Agreeableness

After giving their informed consent, participants first answered questions about our moderator; agreeableness, which refers to individual differences in being altruistic and harmonious in relations with others. To assess agreeableness, we used twelve items from the NEO Five-Factor Inventory and measured them on a 7-point Likert scale ranging from 1: strongly disagree to 7: strongly agree ($\alpha = .72$). This scale was chosen as it is commonly used and has been shown to be a valid and reliable measurement (Costa & McCrae, 2008; Graziano & Tobin, 2009). The order of the items was randomized to ensure that no order effects occurred. Examples of items from this scale are: “I tend to be cynical and skeptical of

others' intentions "(reverse coded) and "Most people I know like me" Appendix B can be consulted to view all items of the NEO-FFI agreeableness scale.

Moral grandstanding

After filling out questions about the moderating variable, participants were exposed to a simulated online discussion on gene modification, presented as a post on platform X, followed by a comment thread of four comments. The topic of gene modification was chosen intentionally, as the researchers anticipated a relatively balanced distribution of opinion for and against it, thereby minimizing potential bias related to the topic itself. Each comment responded directly to the previous one, forming a continuous dialogue. These commenters alternately expressed positions for or against gene modification, with the first commenter being against gene modification, the second one being in favor of it, and so on. Structuring the discussion to represent both pro- and anti-gene modification positions prevented potential bias due to the underrepresentation of either stance. Crucially, the fourth commenter's mode of expression varied by condition. Participants in all conditions were asked explicitly to focus on this comment and were also able to download it so that they could look back at it at later points in the study. In the control condition, the commenter expressed support for gene modification neutrally. In the moral grandstanding condition, the commenters' pro-gene modification stance incorporated explicit moral grandstanding behaviors. The behaviors, as defined by Tosi and Warmke (2016), were used to operationalize the phenomenon of moral grandstanding. The comment included the behaviors: "piling on," "trumping up," and "excessive emotional displays." An example of a sentence in the moral grandstanding condition is "I'm honestly deeply shocked (display of strong emotion) that some people still resist genetic modification, despite the overwhelming evidence that it could save millions of lives! (ramping up)." An example of a fragment in the control condition is "I understand that genetic modification is a complex and controversial issue... However, if we have the ability

to prevent genetic disorders and reduce suffering, shouldn't we at least consider it?" To view the complete manipulation, Appendix C can be consulted.

To avoid influencing participants' responses, all engagement metrics, such as likes, reshares, and comment counts, were deliberately removed from the discussion thread. The fake profile pictures and names were blurred, and participants were told this was for privacy reasons, in an attempt to make the online discussion look more realistic. The only thing altered between the two conditions was the fourth commenter's expression mode. This ensured that observed differences between conditions could be confidently attributed to moral grandstanding, thereby minimizing confounding variables.

Likability

After the manipulation, we presented participants with a measure of likability consisting of 11 items, which we measured on a 7-point Likert scale ranging from 1: strongly disagree to 7: strongly agree ($\alpha = .95$). The items were presented to participants in random sequences to avoid order effects. We chose to use the Reysen Likability Scale (Reysen, 2005), a commonly used measure for likability that has been shown to be valid and reliable. While this scale is primarily used to evaluate individuals in person, we adapted its principles to assess the likability of individuals through text by adjusting the phrasing of questions (e.g., "This person seems warm" becomes "This person seems warm through their writing," and "I would like to be friends with this person" becomes "I would be friends with this person, based on their comment"). See Appendix D to view all items of the original Reysen likability scale and its transformed items. The scale was carefully adapted to fit text-based content. While originally validated for face-to-face interactions, the core components of the scale remain applicable, though future research could further validate its use in text-based formats. As mentioned before, physical attractiveness (Reysen, 2005) and vocal characteristics like tone, frequency, and speech speed (Weiss & Burkhardt, 2010) are linked to higher perceived

likability. Since the current study relies on text-based interactions, factors like physical attractiveness and vocal characteristics are not relevant and will not influence likability ratings. Therefore, the item measuring physical attractiveness was removed from the scale and replaced with an item assessing how knowledgeable the person appears, which is more in line with the focus of the study. This adjustment helped to ensure that any observed effects were due to the content of the grandstanding rather than unrelated factors in the context of this study, like appearance or vocal characteristics.

Manipulation check

To verify the effectiveness of the manipulation used for the independent variable (moral grandstanding), participants rated their agreement with the statement: “The last person to comment (the one you were asked to focus on) was commenting primarily out of a desire to appear morally just, rather than out of genuine moral concern.” They were asked to rate their agreement with this statement on a 7-point Likert scale that ranged from (1): strongly disagree to (7): strongly agree. Additionally, demographic information was collected after the survey, including participants’ gender and age. They were also asked to fill out a seriousness check, indicating whether they took part in the study in a serious manner or just merely clicked through without any actual engagement. Following the data collection, participants received a comprehensive debriefing that explained the study’s true purpose. After the debrief, participants were asked to consent once again for their data to be used.

Results

The software programs used to analyze the data were IBM SPSS Statistics version 28 and PROCESS macro for SPSS (Hayes, 2013). All assumptions were met.

Manipulation Check

To test whether our manipulation worked as intended, we analyzed the results of the manipulation check. We ran an ANOVA with our grandstanding manipulation as the

independent variable and our manipulation check item, where we asked participants to what extent they felt the person was grandstanding, as the dependent variable. As expected, participants in the grandstanding condition ($M = 4.25$, $SD = 1.32$) scored higher on our manipulation check than participants in the control condition ($M = 3.20$, $SD = 1.40$). This positive effect was significant, $F(1, 146) = 24.12$, $p < .001$.

Hypotheses Tests

To test our hypotheses that grandstanding would lead to lower likability ratings and that agreeableness moderates this relationship, we ran an analysis with our grandstanding manipulation as the independent variable, likability as the dependent variable, and agreeableness as the moderator. In line with our first hypothesis, we found that participants in the grandstanding condition ($M = 3.03$, $SD = 1.03$) reported lower perceived likability of the commenter than those in the control condition ($M = 4.53$, $SD = 0.77$). This expected negative effect was significant, $t(144) = -9.88$, $p < .001$. Contrary to what was predicted in the second hypothesis, the agreeableness score of the participants did not moderate the relation between moral grandstanding and likability, $t(144) = -.23$, $p = .82$. The main effect of agreeableness was not significant, $t(144) = 1.67$, $p = .10$. While not significant, the observed value suggested that people high in agreeableness tended to like the commenter generally more.

Discussion

This study investigated how moral grandstanding impacts perceived likability and to what extent this relationship was moderated by the observer's agreeableness. The manipulation check confirmed our manipulation was effective. In the first hypothesis, it was expected that moral grandstanding would result in lower perceived likability ratings. The data supported this prediction; moral grandstanding significantly reduced perceived likability. In the second hypothesis, it was expected that the observer's level of agreeableness would

moderate the relationship between moral grandstanding and likability. However, the data did not support this prediction.

Our findings regarding the first prediction (i.e., the main effect of moral grandstanding) partially align with the existing literature in three key ways. Firstly, the findings align with Tosi and Warmke's (2016) theory that grandstanding may be perceived as performative or insincere, leading to negative social consequences, in this case for the grandstander themselves. Although the researchers did not provide data for this claim, the current study did, thereby reinforcing the credibility of their initial argument. Secondly, the findings align with the former work of Grubbs et al. (2020), in which they found grandstanding is associated with increased polarization and conflict. While their research focused on group-level division, such dynamics are likely to translate to interpersonal interactions as well. Grandstanding may contribute to social friction, which can reduce others' perception of the grandstanders' likability. The current study provides empirical support for this downstream interpersonal consequence, suggesting that consequences of grandstanding occur not only at the group level but also at the interpersonal level. Thirdly, our findings partially align with previous literature on moral grandstanding and its interpersonal consequences. Savejnaron et al. (2022) found that dominance-based moral grandstanding (grandstanding aimed at dominating and intimidating others) is associated with negative interpersonal consequences, such as lower social status and likability, which is in line with our findings. However, they also found that prestige-based moral grandstanding (grandstanding aimed at gaining admiration) was associated with more positive perceptions, such as higher social status, morality, and attractiveness, a pattern that contrasts this study's findings. One explanation for these inconsistent results is that the current study did not split up moral grandstanding into its two formerly established subtypes but rather looked at the effects of the general phenomenon. Overall, our findings seem to indicate that when no

distinction is made between subtypes, moral grandstanding leads to lower perceived likability.

When considering why we did not find the expected moderation effect of agreeableness, three possible explanations stand out. Firstly, it is possible that the moderation effect simply does not exist, or at least not within the specific context of this study, which involved an online discussion on X (formerly Twitter). Agreeableness is usually linked to being kinder and more empathetic in interpersonal relations (Graziano & Tobin, 2009); however, these traits may not carry over as strongly to online settings, where communication is often more extreme or reactive, resulting in less nuance in the way individual differences are expressed (Wu et al., 2023). The second explanation lies in a possible methodological shortcoming; the manipulation of moral grandstanding may have been too strong, possibly overshadowing any smaller moderating effects. As noted earlier, several additional moderating variables were also measured, yet none showed significant effects. This pattern of results suggests that the grandstanding manipulation may have been too overpowering. This raises an important issue inherent to experimental design: achieving a balance between subtlety and clarity. On one hand, the grandstanding must be clear for the manipulation to be effective; on the other hand, it should not be so blatant that it overrides potential smaller moderating effects. The choices made in designing the grandstanding condition were made deliberately, to ensure that the manipulation was noticeable and robust. Thirdly, the lack of moderating effects may also be understood from a theoretical perspective, in which moral grandstanding is viewed as an inherently overpowering phenomenon. Normally, agreeable people are more likely to perceive others positively (Graziano & Tobin, 2009); however, grandstanding might be universally disliked, to the point that even the most agreeable people find it off-putting and unlikable. Rather than indicating a methodological shortcoming, the lack of expected moderation effect could also reflect something true about human perception

of grandstanding. This interpretation is supported by the data; agreeableness had a near-significant positive association with liking overall, showing that the personality measure functioned as expected; however it seems like the trait cannot buffer against grandstanding. Further, the manipulation check showed that participants did not rate the grandstanding as overly blatant (scoring slightly above 4 on a 7-point scale). This makes it less likely that the manipulation was too overpowering; instead, it suggests that the phenomenon of grandstanding itself may be inherently overpowering.

Literature seems to indicate that the phenomenon of moral grandstanding is associated with negative outcomes, such as increased polarization (Tosi & Warmke, 2016; Grubbs et al., 2020). This study further expands the negative consequences of the phenomenon by looking at its interpersonal costs. On the basis of this study's findings, we can conclude that the general phenomenon of moral grandstanding leads to lower perceived likability, confirming the social cost of moral grandstanding in online discourse.

Limitations

This study provided important insights into the relationship between moral grandstanding, agreeableness, and likability; however, five limitations must be acknowledged. Firstly, the study consisted of an online experiment that used a self-selected sample. While this method allowed efficient data collection, it may also have introduced self-selection bias. Participants who voluntarily engage in studies on online discourse may already hold specific attitudes towards others' digital behavior. While not much is known about attitude differences regarding this topic, the established literature on self-selection bias shows that voluntary participants are oftentimes more opinionated and informed about the subject matter than the general population (Bethlehem, 2020; Lavrakas, 2008). The sample may therefore overrepresent individuals who are more active consumers of online moral discussions, who possibly have already developed evaluative frameworks for what is

considered acceptable or unacceptable in online behavior. As a result, the findings may not generalize to populations less familiar with or invested in online conversations. Future research could enhance generalizability by using stratified sampling methods or recruiting participants. Secondly, the sample used in this study may have been demographically too homogeneous, leading to a compromised generalizability of the current study's findings. The researchers recruited participants from their own networks, which led to a sample consisting mostly of people similar in age ($n = 148$, $M = 27.24$, $SD = 10.41$). Prior research by Roberts et al. (2006) has shown that the personality trait of agreeableness tends to increase with age. Our sample, consisting of predominantly young participants, was therefore not fully representative of the population and could have led to biased results. Future research could address this limitation by drawing from a bigger demographic pool, including people of entirely different ages. Thirdly, the use of a simulated discussion in the experimental manipulation might have undermined the ecological validity of this study. The discussion may not reflect real-world discourse accurately, as it lacks the use of natural language usually present online, such as slang and typing errors (Kunch et al., 2022). While the absence of informal language and typing errors in the comment thread was an intentional decision, made to avoid giving participants ideas about the commenter's credibility or knowledge, it could have led to biased results. If participants sensed the discussion to be fabricated, they may have responded differently (e.g., "This person is not real anyway, so I can be harsher in my judgement" or, oppositely, "This person is not real anyway, so therefore I could not bother to judge them harshly"). This limitation could be addressed by future research by using a real-life discussion that contains grandstanding behaviors. The grandstanding comment should then be neutralized in its wording. While this method would still not provide a perfect ecological validity, it would still make for a significant improvement. Fourthly, the artificial nature of the discussion could have introduced hypothesis confirmation bias. Since the

comment thread was fabricated by the researchers, the phrasing of the text could have unintentionally steered participants in the direction that confirmed the hypotheses the researchers had. This limitation too could be addressed by using a real discussion found on X or another social media platform. Lastly, it is also possible that demand characteristics influenced participants' responses. The moderators we measured prior to the manipulation may have given clues to the participants about the true purpose of the study. Despite the fact deception was used regarding the aim of the study (i.e., the study being about online discourse), they may have perceived the study to be about moral judgement and given their answers accordingly. This could have led to biased answers; participants might have tried to confirm the perceived hypothesis or, conversely, gone against it. Future research could address this by using a stronger, more convincing form of deception. This could, for instance, be done by adding more detail when introducing the study (e.g., this study is about your judgments of the writing clarity of online discourse). Another strategy to avoid participants guessing the true purpose of the study would be to add several distracting variables, such as memory and visual tasks, so the moral component seems just one of many. Due to time constraints, this study did not implement such variables; however, they could be a valuable addition for upcoming studies.

Future research

Whilst the current study provided important information about moral grandstanding and its social costs, many questions about this underexplored phenomenon still remain unanswered. Hence, we give three recommendations for future research. Firstly, it would be valuable to investigate the way moral grandstanding influences likability in a video or in a real-life situation, as grandstanding not only occurs online but also frequently occurs in real-life scenarios too (Tosi & Warmke, 2016). Real-world grandstanding includes certain vocal characteristics, facial expressions, and body language that significantly contribute to how

messages are interpreted (Duncan, 1969; Tosi & Warmke, 2016). Future research could incorporate videos or real-life interactions, where an actual discussion is being held between individuals, in order to see how social and nonverbal cues influence the perception of moral grandstanding. Real-life formats may also provide a more ecologically valid context in which traits like agreeableness, as well as multiple other factors, can entirely manifest and possibly moderate interpersonal evaluations (Wu et al., 2023). Secondly, future research could broaden the scope of existing literature by looking at the intrapersonal effects of moral grandstanding. Instead of focusing on how grandstanding is perceived by others, it might be valuable to investigate the way grandstanding influences the grandstander themselves. It is imaginable that grandstanding would lead to increased stress that results from the constant pressure of keeping up an image of oneself that is not or only partially congruent with one's actual values. Shedding light on the possible personal costs of performative moral talk may serve as another discouragement for individuals to partake in this behavior. Thirdly, using qualitative or mixed methods in future studies could add valuable insights to the existing body of literature on the topic. Qualitative approaches give participants a voice, allowing them to share their thoughts, feelings, and reasons behind their choices in more detail. This can make the findings richer and more meaningful by showing the context behind the numbers. Mixed methods, which combine both qualitative and quantitative elements, could additionally be of great value for future research because they allow for a comprehensive understanding of phenomena. Since moral grandstanding is a complex phenomenon, it requires methods that can address all of its aspects. Overall, these methods can make the research more complete and better reflect real human experience.

The current study's findings have an important practical implication; they suggest organizations, authorities, and individuals should be cautious about appearing overly moralistic in communications meant to persuade, as this will only turn people against them.

Being more nuanced when putting out online statements will cause people to perceive the messenger as more likable. This observation could help the field of social environmental psychology by giving them more clearly defined handles for how they ought to deliver a message to the public they want to persuade.

Conclusion

Moral grandstanding has been shown to have many negative effects. The current study further underscores its interpersonal costs, as it showed that grandstanders are perceived as less likable. While this relation was not moderated by agreeableness, the findings contribute to a growing understanding of what the effects of moral grandstanding are in digital contexts. When people grandstand, their aim is to achieve something positive, namely to earn the respect and admiration of others. Ironically, their behavior often has the opposite effect: it alienates people and turns them away. This creates a paradox: people seek status but end up causing others' resentment by doing so. All the observed effects and theorized consequences of moral grandstanding have been consistently negative, so therefore the takeaway message is the opposite of what Nike has been telling us for years: just don't do it!

References

- Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, 78(2), 161–188. <https://doi.org/10.1111/j.1751-5823.2010.00112.x>
- Bishara, H. (2020) Susan B. Anthony Museum Rejects Trump's Presidential Pardon <https://hyperallergic.com/583840/susan-b-anthony-museum-pardon-reject/>
- Bleckmann, E., Wieczorek, L. L., & Wagner, J. (2024). The Role of Agreeableness, Neuroticism, and Relationship-Specific Features in Self- and Other-Perceptions of Conflict Frequency in Adolescent Relationships with Parents and Peers. *Journal Of Youth And Adolescence*, 53(7), 1630–1645. <https://doi.org/10.1007/s10964-024-01951-6>
- Chaiken, S. (1980). *Heuristic versus systematic information processing and the use of source versus message cues in persuasion*. *Journal of Personality and Social Psychology*, 39(5), 752–766. <https://doi.org/10.1037/0022-3514.39.5.752>
- Costa, P. T., Jr., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment, Vol. 2. Personality measurement and testing* (pp. 179–198). Sage Publications, Inc. <https://doi.org/10.4135/9781849200479.n9>
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3), 285–290. <https://doi.org/10.1037/h0033731>
- Duncan, S., Jr. (1969). Nonverbal communication. *Psychological Bulletin*, 72(2), 118–137. <https://doi.org/10.1037/h0027795>
- Ezirim, G. E., Mbah, P. O., Nwagwu, E. J., Eze, I. C., Nche, G. C., & Chukwuorji, J. C. (2021). Trust and trustworthiness in a sub-Saharan African sample: Contributions of personality and religiosity. *Social Indicators Research*, 153(3), 1087–1107.

<https://doi.org/10.1007/s11205-020-02536-z>

Goldberg, L. R. (1992). *Goldberg's Big Five Questionnaire* [Database record]. APA

PsycTests. <https://doi.org/10.1037/t09696-000>

Graziano, W. G., & Tobin, R. M. (2018). Agreeableness: A three-level integration. In V.

Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE handbook of personality and individual differences: Applications of personality and individual differences*. (pp.

212–234). Sage Reference. <https://doi.org/10.4135/9781526451248.n9>

Grubbs, J. B., James, A. S., Warmke, B., & Tosi, J. (2022). Moral grandstanding, narcissism, and self-reported responses to the COVID-19 crisis. *Journal Of Research in*

Personality, 97, 104187. <https://doi.org/10.1016/j.jrp.2021.104187>

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral

grandstanding in public discourse: Status-seeking motives as a potential explanatory mechanism in predicting conflict. *PLoS ONE*, 14(10), e0223749.

<https://doi.org/10.1371/journal.pone.0223749>

Grubbs, J. B., Warmke, B., Tosi, J., & James, A. S. (2020). Moral grandstanding and political polarization: A multi-study consideration. *Journal of Research in Personality*, 88.

<https://doi.org/10.1016/j.jrp.2020.104009>

Hare, B. (2016). Survival of the Friendliest: Homo sapiens Evolved via Selection for

Prosociality. *Annual Review Of Psychology*, 68(1), 155–186.

<https://doi.org/10.1146/annurev-psych-010416-044201>

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis:

A regression-based approach. *New York: Guilford Press*.

<https://doi.org/10.1111/jedm.12050>

Kunch, Z., Serednytska, A., Vasylyshyn, I., Horodylovska, H., & Farion, I. (2022). The

Peculiarities of Spelling Rules in Formal, Informal Handwriting and Internet

- Communication. In *COLINS* (pp. 91-103). <https://ceur-ws.org/Vol-3171/paper11.pdf>
- Lavrakas, P. (2008). Encyclopedia of Survey Research Methods. In *SAGE Publications, Inc. eBooks*. <https://doi.org/10.4135/9781412963947>
- McClay, B. D. (2018). Virtue signaling. *The Hedgehog Review*, 20(2), 141-144.
- link.gale.com/apps/doc/A549154930/AONE?u=anon~ee4b495d&sid=googleScholar&xid=d424626f
- McCrae, R. R., & Costa, P. T., Jr. (1999). A Five-Factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). Guilford Press. <https://psycnet.apa.org/record/1999-04371-005>
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776–783. <https://doi.org/10.1037/h0043424>
- Oyibo, K., & Vassileva, J. (2019). The relationship between personality traits and susceptibility to social influence. *Computers in Human Behavior*, 98, 174–188. <https://doi.org/10.1016/j.chb.2019.01.032>
- Pulles, N. J., & Hartman, P. (2017). Likeability and its effect on outcomes of interpersonal interaction. *Industrial Marketing Management*, 66, 56–63. <https://doi.org/10.1016/j.indmarman.2017.06.008>
- Reysen, S. (2005). Construction of a New Scale: The Reysen Likability Scale. *Social Behavior and Personality: An International Journal*, 33(2), 201–208. <https://doi.org/10.2224/sbp.2005.33.2.201>
- Rivera, G. N., Kim, J., Kelley, N. J., Hicks, J., & Schlegel, R. J. (2024). Liking Predicts Judgments of Authenticity in Real-Time Interactions More Robustly Than Personality States or Affect. *Personality and Social Psychology Bulletin*, 0(0). <https://doi.org/10.1177/01461672231218758>

- Rivera, G. N. (2020). The Role of Honesty, Likability, and Assumed Morality on Judgments of Others' Authenticity (*Doctoral dissertation, Texas A&M University*).
<https://www.proquest.com/openview/72562b082feef9a7c44117c50e5c79ac/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y>
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1–25. <https://doi.org/10.1037/0033-2909.132.1.1>
- Rosa, S. C. (2023). Feminist, Non-Feminist, and Anti-Feminist Uses of Feminist Memory. *Histoire Sociale*, 56(116), 275–300. <https://doi.org/10.1353/his.2023.a914564>
- Savejnarong, T., Pornsukjantra, P., & Manley, H. (2022). The interpersonal consequences of prestige and dominance-based moral grandstanding. *Personality and Individual Differences*, 194, 111656. <https://doi.org/10.1016/j.paid.2022.111656>
- Sedlár, M. (2023). Trust in strangers and friends: The roles of agreeableness, open-mindedness, perspective taking, and trustworthiness. *Nordic Psychology*, 75(3), 295–310. <https://doi.org/10.1080/19012276.2022.2094823>
- Stollznaw, K. (n.d.). 'Virtue signalling', a slur meant to imply moral grandstanding that might not be all bad. *The Conversation*.
<https://theconversation.com/virtue-signalling-a-slur-meant-to-imply-moral-grandstanding-tha-might-not-be-all-bad-145546>
- Sunstein, C. R. (2002). The Law of Group Polarization. *Journal of Political Philosophy*, 10(2), 175–195. <https://doi.org/10.1111/1467-9760.00148>
- Tosi, J., & Warmke, B. (2016c). Moral grandstanding. *Philosophy & Public Affairs*, 44(3), 197–217. <https://doi.org/10.1111/papa.12075>
- Weiss, B., & Burkhardt, F. (2010). Voice attributes affecting likability perception. *Interspeech*

2022, 2014–2017. <https://doi.org/10.21437/interspeech.2010-570>

Wilmot, M. P., & Ones, D. S. (2022). Agreeableness and its consequences: A quantitative review of meta-analytic findings. *Personality and social psychology review*, 26(3), 242-280. <https://doi.org/10.1177/10888683211073007>

Wu, W., Mitchell, P., & Lv, Y. (2023). Consistency in personality trait judgments across online chatting and offline conversation. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1077458>

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151–175. <https://doi.org/10.1037/0003-066X.35.2.151>

Appendix A

List of all Measured Variables

<u>IV</u>	<u>M</u>	<u>DV</u>
Moral Grandstanding	Agreeableness	Likability
Moral Grandstanding	Agreement with the topic	Perceived consensus
Moral Grandstanding	Optimism/pessimism	Perception of Moral grandstanding
Moral Grandstanding	Political Identity Strength	Ideological extremism
Moral Grandstanding	Agreement or disagreement with the topic	Perceived credibility of the speaker
Moral Grandstanding	Moral sensitivity	(Dis)engagement from topic

Appendix B

Items of the NEO-FFI Agreeableness Scale

Item no.	Item	Flipped (f)
1	I try to be kind to everyone I meet.	
2	I often get into arguments with my family and co-workers.	(f)
3	Some people think I'm selfish and egotistical.	(f)
4	I would rather cooperate with others than compete with them.	
5	I tend to be cynical and skeptical of others' intentions.	(f)
6	I believe that most people will take advantage of you if you let them.	(f)
7	Most people I know like me.	
8	Some people think of me as cold and calculating.	(f)
9	I am hard-headed and tough-minded in my attitudes.	(f)
10	I generally try to be thoughtful and considerate.	
11	If I don't like people, I let them know it.	(f)
12	If necessary, I am willing to manipulate people to get what I want.	(f)

Appendix C

Manipulation

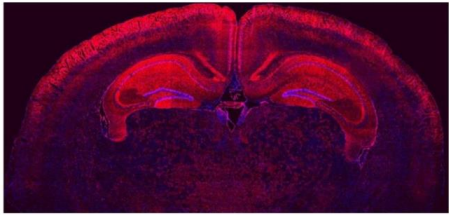
Figure C1

Control condition

Marco Johnson
@marcoj_official

More precise new gentle gene-edit approach developed by scientists. Promises great potential for future use in human embryos for disease and disability prevention. While there are ethical discussions to be had, this could be a game-changer for medicine.

<https://health.ucdavis.edu/news/headlines/could-a-new-medical-approach-fix-faulty-genes-before-birth/2024/10>



RESEARCH | October 24, 2024

Could a new medical approach fix faulty genes before birth?

By Nadine A Yehya

A new study in mice shows a unique mRNA delivery method for in-utero gene editing for neurodevelopmental conditions

2:02 PM · Nov 3, 2024

37 Retweets 5 Quote Tweets 120 Likes

Nov 3

I just don't think we should mess with nature like this. Once we start editing genes where do we draw the line? It feels like playing god.

Nov 4

I understand the concern but genetic modification has the potential to eliminate serious diseases. Wouldn't we want to prevent unnecessary suffering if we can?

Nov 4

But there are so many ethical issues! What if only rich people can afford it and it deepens inequality? This could create a huge social divide

Nov 5

I understand that genetic modification is a complex and controversial issue. There are ethical concerns and careful regulation necessary. However, if we have the ability to prevent genetic disorders and reduce suffering shouldn't we at least consider it? Scientific process has always faced scepticism but if used responsibly this technology could save lives and improve public health.

Figure C2

Experimental Condition

Marco Johnson
@marcoj_official

More precise new gentle gene-edit approach developed by scientists. Promises great potential for future use in human embryos for disease and disability prevention. While there are ethical discussions to be had, this could be a game-changer for medicine.

<https://health.ucdavis.edu/news/headlines/could-a-new-medical-approach-fix-faulty-genes-before-birth/2024/10>



RESEARCH | October 24, 2024

Could a new medical approach fix faulty genes before birth?

By Nadine A Yehya

A new study in mice shows a unique mRNA delivery method for in-utero gene editing for neurodevelopmental conditions

2:02 PM · Nov 3, 2024

Nov 3

I just don't think we should mess with nature like this. Once we start editing genes where do we draw the line? It feels like playing god

Nov 4

I understand the concern but genetic modification has the potential to eliminate serious diseases. Wouldn't we want to prevent unnecessary suffering if we can?

Nov 4

But there are so many ethical issues! What if only rich people can afford it and it deepens inequality? This could create a huge social divide

Nov 5

I'm honestly deeply shocked that some people still resist genetic modification despite the overwhelming evidence that it could save millions of lives! How can anyone in good conscience stand in the way of such obvious progress? That is clearly not ethical in any way. I stand by those who care about our future generations and who embrace this technology instead of clinging to outdated fears. Refusing to support it is irresponsible and egoistic! It is that simple... 😞

Appendix D

Items of the Reysen likability Scale and its Transformed Items

Table D1

List of original items

Item no.	Item
1	This person is friendly.
2	This person is likable.
3	This person is warm.
4	This person is approachable.
5	I would ask this person for advice.
6	I would like this person as a coworker.
7	I would like this person as a roommate.
8	I would like to be friends with this person.
9	This person is physically attractive.
10	This person is similar to me.
11	This person is knowledgeable.

Table D2

List of Transformed items

Item no.	Item
1	This person is friendly based on their comment.
2	This person is likable based on their comment.
3	This person seems warm through their writing.
4	This person comes across as approachable through their comment.
5	I would ask this person for advice based on their comment.
6	I would like this person as a coworker based on their online reasoning.
7	I would enjoy living with this person as a roommate based on their writing.
8	I would like to be friends with this person based on their comment.

9	This person seems knowledgeable based on their comment.
10	This person seems similar to me based on their comment.
11	This person seems trustworthy based on their comment.

Appendix E

Usage of AI

Hereby I acknowledge the use of AI tools in the process of writing this thesis. I used Open AI's ChatGPT (<https://chatgpt.com/>) for improvement of sentence structure and grammatical flow. Additionally it assisted me with my literature search. I used Google's notebookLM (<https://notebooklm.google.com/>) to summarize sources and help me answer questions according to the articles. I declare that AI did not generate any of the content I presented in my work, nor did it develop any of the arguments or analysis presented in this thesis.