



When You Are Wrong on Facebook, Just Admit It

Wrongness Admission Leads to Better Interpersonal Impressions on Social Media

Adam K. Fetterman^{1,2} , Nicole L. Muscanell^{2,3}, Dongjie Wu¹, and Kai Sassenberg^{2,4}

¹Department of Psychology, University of Houston, Houston, TX, USA

²Social Processes Lab, Leibniz-Institut für Wissensmedien, Tübingen, Germany

³Defense Equal Opportunity Management Institute, JHT, Incorporated, USA

⁴Department of Psychology, University of Tübingen, Germany

Abstract: Intellectually humble behavior, like admitting when you are wrong, leads to better impression formation. However, online social networks (OSNs) have changed the impression formation process. We investigated the impact of wrongness admission on impression formation during an OSN argument. In four experiments ($N = 679$), participants witnessed a user engage in wrongness admission, refuse to admit, or not respond, in an argument on a Facebook wall. Participants reported their impressions of whether they would be willing to interact with the (non) admitting user. User reputation ratings and interaction intentions were higher in the admission (vs. nonadmission) condition. The latter effect was mediated by user reputation ratings. Wrongness admission appears to have a positive impact on impression formation on OSNs.

Keywords: intellectual humility, impression formation, social media, reputation, wrongness admission

Arguments on online social networks (OSNs), such as Facebook and Twitter, have become a part of everyday life (Anderson et al., 2018; Neubaum et al., 2021). This is not surprising because OSNs create a space for people to easily express their opinions, which can encourage open dialog. However, OSNs have also changed the impression formation process (Walther & Whitty, 2021). Instead of forming impressions based on face-to-face interactions, OSNs allow people to look at people's online behavior as cues to their personalities (Utz, 2010). Furthermore, people using OSNs often form impressions through passive observation (Qin et al., 2021). Therefore, an OSN user's behavior during an online argument might have an impact on the impressions passive viewers form of the user (Orben & Dunbar, 2017). One such behavior that can occur during an online argument is a user's response when they discover that their viewpoint is factually wrong. With potentially hundreds (or more, depending on their privacy settings) of passive witnesses, the user can admit that they are wrong or avoid doing so. Based on the literature of impression formation on OSNs and intellectual humility, the OSN user's best course of action, here, is to publicly admit that they are wrong. The current set of experiments investigated whether this is indeed the case by testing

whether such *wrongness admission* during a Facebook argument impacts passive viewers' impression formation.

Intellectual Humility and Wrongness Admission

Intellectual humility is defined by Leary et al. (2017) as "recognizing that a particular personal belief may be fallible, accompanied by an attentiveness to limitations in the evidentiary basis of that belief and to one's own limitations in obtaining and evaluating relevant information" (p. 793). Those who score high on broad measures of intellectual humility, based on this definition, are more agreeable and open-minded (Leary et al., 2017). More recently, Porter et al. (2021) refined intellectual humility by developing a classification framework introducing different forms of intellectual humility. This classification is based on whether the intellectual humility is focused on the self or others – that is, about one's own or another person's beliefs – and whether it is internal or expressed – that is, about self-reflective awareness and cognitions compared to behavioral manifestations. In relation to the self/expressed quadrant of their framework, these observable

behaviors can be public or private. It is within public component of the self/expressed quadrant of intellectual humility that we locate a concept we call “wrongness admission.”

Wrongness admission is a form of public attitude change that involves expressing that one has changed a previously held attitude or belief (Fetterman et al., 2019). In this sense, wrongness admission is about recognizing and publicly acknowledging that a belief is *inaccurate*. Wrongness admission is not about being morally wrong or about behavioral errors. Willingness to engage in wrongness admission, as we define here, is positively correlated with agreeableness, openness to experience, honesty/humility, and emotional intelligence (Fetterman et al., 2019). Furthermore, those who score high in intellectual humility are more likely to engage in wrongness admission during an argument (Rodríguez et al., 2019). Wrongness admission, then, is an observable behavior that cues not only intellectual humility but also positive interpersonal traits (e.g., trustworthiness and friendliness).

If intellectual humility and wrongness admission, more specifically, are cues to positive interpersonal traits, then they should be reflected in observer reports of intellectually humble people. Indeed, Meagher et al. (2021) found that those who scored higher on intellectual humility, in general, tended to be rated by interaction partners as more agreeable and open-minded, less arrogant, and dominant, and that interaction partners felt more satisfaction when conversing with intellectually humble others (Meagher et al., 2021). Based on this work, intellectual humility appears to have a positive impact on impression formation, notably due to its cue of positive interpersonal traits.

There has been relatively little work looking more specifically at the impact of wrongness admission on impression formation. There are two notable exceptions. First, a study by Fetterman and Sassenberg (2015) had published scientists read one of two scenarios. In one scenario, after learning about a convincing failed replication of one of their research findings, a hypothetical scientist admits that they were wrong about the finding. In the other condition, the scientist refuses to admit that they were wrong about the finding. The results indicated that participants in the admission condition rated the target scientist as more competent and trustworthy than participants in the refuse condition. This was the first study to suggest that wrongness admission has a positive impact on impression formation.

In the second notable exception, John et al. (2019) investigated wrongness admission in business settings. They had participants give an entrepreneurial pitch, during which a panel contradicted some of the participants' statements. The results showed that those who backed down (i.e., wrongness admitters) were more likely to be

advanced to a final competition round and were rated as more competent by observers. Furthermore, these competence ratings impacted, in a positive way, whether the observers wanted to hire the admitter or invest in their ideas. Wrongness admission, again, appears to have a positive impact on impression formation.

Intellectual humility and wrongness admission appear to cue positive interpersonal traits for observers forming impressions. Yet, OSNs are unique environments in which people manage and form impressions (Walther & Whitty, 2021). Passive observers on OSNs form spontaneous impressions, without much context, of strangers as they scroll through their posts (Levodashka & Utz, 2017). Therefore, it is important to understand how wrongness admission on OSNs impacts impression formation.

Impression Formation on OSNs

Humans have an inherent need to form and maintain relationships due to their incredible social nature (Baumeister & Leary, 1995). To ensure a mutually beneficial relationship, people rely on reputation to predict the likely behavior of potential interaction partners, even on OSNs (Tennie et al., 2010). To get a sense of someone's reputation, people make inferences about that person's personality (i.e., impression formation). People form impressions of others along two fundamental dimensions: communion and agency (Abele et al., 2016; Abele & Wojciszke, 2014; Asch, 1946; Cuddy et al., 2008; Fiske et al., 2007). The agency dimension includes traits related to achievement and competence, while the communion dimension includes social variables, such as trust and friendliness. People often form impressions about these traits quickly, sometimes within seconds, based on non-verbal cues (Ambady & Rosenthal, 1992), which are remarkably accurate (Funder, 2012). However, given the rise of OSNs, impression formation has changed and to some extent moved online.

Since OSNs have become an everyday feature of most people's social lives (Seidman, 2013), much research has focused on impression formation and management on OSNs (Bacev-Giles & Haji, 2017). OSNs, however, have changed impression formation. For example, much OSN-based impression formation is passive in that the people are not interacting with those for whom they are forming impressions (Orben & Dunbar, 2017). Due to this lack of interaction, according to the hyperpersonal model (Walther & Whitty, 2021), people often form impressions based on textual and pictorial information posted by users. Of course, this material can be highly curated by the users (Krämer & Winter, 2008). Therefore, according to the warranting principle (Walther et al., 2009), viewers prefer

information that is less subject to user manipulation. Indeed, prior work shows that viewers form fewer positive impressions from user-generated information, likely due to the perceived untrustworthiness of this information (Utz, 2010).

Prior work investigating impression formation on OSNs has revealed specific traits and online behaviors that lead to more positive impressions, or liking. Much of this work has shown that communal behavior tends to lead to the most positive impressions. For example, Buffardi and Campbell (2008) found that self-promotion led to narcissistic impressions, which was related to lower ratings of communion. Furthermore, Stopfer et al. (2014) found that communal and open-minded people were more liked on OSNs. More recently, Qin et al. (2021) found that positive self-disclosure on OSNs led to higher ratings of trustworthiness and likeability. As noted, viewers might not form as many positive impressions from user-generated information (Utz, 2010). However, the negative impact of user-generated information can be overridden if the user engages in self-deprecation, a cue of humility (Austin et al., 2021; Bareket-Bojmel et al., 2016).

Overall, it seems that impression formation on OSNs is common and spontaneous (Levordashka & Utz, 2017). Although impression formation is different on OSNs than offline situations (Walther & Whitty, 2021), these impressions tend to be accurate (Stopfer et al., 2013). Furthermore, people tend to form the most positive impressions for those on OSNs who display communal, open, and humble online behaviors. We noted above that intellectual humility and wrongness admission tend to cue such positive traits and lead to positive impressions offline. Intellectual humility and wrongness admission could lead to positive impression formation online as well.

Wrongness Admission on OSNs

Due to (a) the commonality in arguments on OSNs (Anderson et al., 2018; Neubaum et al., 2021), (b) people's tendency to form spontaneous impressions of strangers while passively scrolling through their posts (Levordashka & Utz, 2017), and (c) the impact of passive judgments in impression formation in general (Quadflieg & Penton-Voak, 2017) and on OSNs (Waggoner et al., 2009), it seems important to test how wrongness admission during OSN arguments impacts the impression formation process. In fact, Utz (2010), drawing from the Brunswik lens model (Brunswik, 1956; Gigerenzer & Kurz, 2001), suggests that the impressions formed from cues on OSNs create a lens through which viewers make inferences about individuals. Intellectually humble behaviors, such as

wrongness admission, appear to cue positive interpersonal traits, which could shape a viewer's lens of an online wrongness admitter. This leads to our first hypothesis.

Hypothesis 1 (H1): Scores on user reputation will be more positive for participants who witness a stranger in an OSN argument engage in wrongness admission than who witness a stranger refusing to admit.

Among the primary uses of OSNs is the forming of social connections (Alhabash & Ma, 2017). Impression formation serves the function of informing the viewer of whether they should engage in future interactions with the target (Gable & Reis, 2001). In the context of OSNs, people form impressions with less contextual information than in face-to-face interactions (Walther & Whitty, 2021). Since (a) people are primarily looking for social connections and thus forming impressions based on interpersonal traits on OSNs (Tennie et al., 2010), (b) friendships often form based on judgments of agreeableness (Harris & Vazire, 2016), and (c) wrongness admission appears related to such traits, wrongness admission in an OSN argument might be a cue to passive viewers that the user is someone to interact with in the future. In fact, Hagá and Olson (2017) found that people were more likely to want to interact with people who displayed outward intellectual humility, such as wrongness admission, due to their increased perceptions of niceness. This leads to our second and third hypotheses.

Hypothesis 2a (H2a): Participants who witness a stranger in an OSN argument will indicate that they are more willing to interact with the user in the future if that user engages in wrongness admission compared to a user who refuses to do so.

Hypothesis 2b (H2b): Participants who witness a stranger in an OSN argument will indicate that they are more willing to interact with the user in the future if that user engages in wrongness admission compared to a user who refuses to do so and this effect is mediated through increases in positive impression formation.

Current Experiments

Across four experiments, we tested our broad hypothesis that wrongness admission on OSNs would have a positive impact on impression formation. We focused all of our experiments on Facebook. According to statista.com, Facebook is still the most used OSN as of October 2021, with 2.9 billion active users worldwide. Even so, the studies

were run between 2014 and 2016, and during that time, Facebook use was ubiquitous (Duggan & Smith, 2016).

We designed our OSN materials in a similar manner to that of previous investigations investigating the impact of minimal information on impression formation on OSNs (e.g., Austin et al., 2021; Kaye et al., 2020) and based on the idea that viewers tend to discredit user-curated OSN profile information (Utz, 2010) and form strong impressions when passively observing interactions (Quadflieg & Penton-Voak, 2017). Therefore, we provided only the textual information and removed all identifying characteristics and profile information. For the impression formation outcome measures, we created measure of general reputation, which includes items that reflect communal and competence traits, similar to that of previous investigations of impression formation on OSNs (e.g., Austin et al., 2021; Bacev-Giles & Haji, 2017; Batenburg & Bartels, 2017). In all four experiments, participants read a staged argument on a Facebook wall – a section of Facebook that allows users to post information and engage in discussion – between two users. The argument was focused on a made-up food additive and reflected common themes of health food discourse on OSNs at the time of the study (Munro et al., 2015) to increase realism. The differences between the experiments are provided in Table 1.

Open Science Disclosures

The current experiments contain all but two studies that we have conducted on the topic of wrongness admission on Facebook. The two studies not included consisted of undergraduate thesis projects that do not meet the standards of publishable research. The data from Experiment 2B included questionnaires relevant, but not related to the

current hypotheses. This questionnaire data, but not the data we report here, were published in Fetterman et al. (2019).

We initially intended to measure general user reputation. However, as the project and our theorizing progressed, we thought that it might be more informative to split the general reputation items into separate communion and competence scores. When we did so, the two scores were highly positively correlated (all $r_s > .70$). Furthermore, a parallel analysis on the data from all four experiments suggested that a single reputation factor was most appropriate. Therefore, we returned to our original general user reputation measure.

Beyond these disclosures, we report how we determined our sample sizes, all data exclusions, all manipulations, and all measures in the studies. We did not pre-register our hypotheses. Data, code, and full materials are available on the Open Science Framework (<https://osf.io/a4tpg/>).

Experiment 1A and Experiment 1B

Method

Participants

In Experiment 1A, we recruited 100 participants on Amazon Mechanical Turk based on the assumption that 50 per cell would be sufficient. We removed participants who did not accurately respond to an attention check or did not complete the experiment. Of the participants who completed the task, 88 (41 female; $M_{age} = 31.00$, $SD_{age} = 7.84$) answered the attention check correctly and 91% indicated that they use Facebook. For Experiment 1B, we performed an incorrect a priori power analysis, based on which we recruited 250 participants on Amazon Mechanical Turk. Of those participants, 238 participants (100 female) accurately responded to the attention check and 92% indicated that they use Facebook. A sensitivity analysis suggested that we had the power to detect a minimum effect size of $\eta^2_{1A} = .08$ and $\eta^2_{1B} = .04$ when $\alpha = .05$ (two-tailed) and power = .80.

Materials and Procedures

Participants were randomly assigned to an admission or a refuse (Experiment 1A and Experiment 1B) or a no-response (Experiment 1B only) condition. After providing informed consent, participants read instructions and said that we were interested in people's responses to Facebook arguments, that we had access to a corpus of Facebook arguments, and that we would randomly select one of the arguments for them to read. In reality, all participants read the same argument. Only the final post from the target user differed and served as the

Table 1. Breakdown of the differences across experiments and final sample sizes

Experiment	Conditions present	Wording of condition	Final <i>N</i>
Experiment 1a	1. Admission 2. No admission	Wording 1	88
Experiment 1b	1. Admission 2. No admission 3. No response	Wording 1	238
Experiment 2a	1. Admission 2. No admission 3. No response	Wording 2	171
Experiment 2b	1. Admission 2. No admission 3. No response	Wording 2	183

Note. Wording of condition = differences in the wording of the final post of the target Facebook user based on feedback from participants.

manipulation. Participants also read that they were to pay close attention to Participant A, but that they should pay attention to all users. They could not advance until 2 minutes had past.

The Argument

Participants read a screenshot of a fabricated argument between two Facebook users. Two of us staged the argument on our own Facebook pages to increase realism. To further increase realism, we examined real arguments on OSNs and modeled the same argument tactics. For example, the *users* called-out biases, utilized and called-out logical fallacies, cited blogs as sources, and resorted to ad hominem attacks. However, we preplanned the argument

such that, while there was room for debate, Participant B clearly had the facts on their side and had the more compelling argument. Participant B also cited scientific papers, instead of blogs. We did this to avoid too many participants thinking that the admitter was inappropriately engaging in wrongness admission. We removed the identities of the actors and labeled them Participant A and Participant B. As Participant A was the focus of the experiment, we highlighted their posts with a red border. See Figure 1 for the full argument stimulus.

The final post by Participant A contained the manipulation. In one (admission) condition, Participant A ended the conversation by posting, “Thanks for the info and conversation. I was wrong and you were right.” In the



Figure 1. Argument stimulus viewed by all participants in all experiments.

other (refuse) condition, Participant A ended the conversation by posting, “Thanks for the info and conversation. I guess we agree to disagree.” Therefore, the only thing that differed in the conditions was the second sentence in the final post. In Experiment 1B, participants in the no-response condition read, instead of a final post, “There were no further responses by Participant A.”

We conducted a pilot study ($N = 50$) to confirm that most people thought that Participant B was more correct. Participants read a neutral (i.e., no final post) version of the conversation and judged whether they thought Participant A or B was correct. Indeed, 70.59% of participants agreed that Participant B was correct. This was significantly different than 50%, $t(48) = 3.20$, $p = .002$, $d = 0.90$, indicating that perceptions of correctness were not random.

User Reputation Ratings

Participants, in both studies, rated their agreement (1 = *strongly disagree*; 5 = *strongly agree*) with 18 statements about Participant A with items related to communion (e.g., “Participant A is a nice person”) and to competence (e.g., “Participant A has little education”). We reverse scored the negatively worded items and averaged across all items to create a user reputation score (1A: $M = 3.06$, $SD = 0.77$, $\alpha = .94$; 1B: $M = 2.83$, $SD = 0.87$, $\alpha = .94$).

Interaction Intentions

To measure whether participants would be willing to interact, online or offline, with the target user in the future (i.e., testing *Hypotheses 2a* and *2b*), participants indicated their level of agreement (1 = *strongly disagree* to 5 = *strongly agree*) with four statements (e.g., “I would like to meet this person”). We reverse scored the negatively worded items and averaged across these items to create an interaction intentions score (1A: $M = 2.88$, $SD = 1.03$, $\alpha = .84$; 1B: $M = 2.81$, $SD = 0.87$, $\alpha = .80$).

Attention Check

To assess whether participants were aware of the *admission* or *refusal* (Study 1A and Study 1B) or *no response* (Experiment 1B), we administered a memory test at the end of the experiments. Participants indicated whether

Participant A engaged in wrongness admission. As noted, we also used this memory test as an attention check to exclude participants.

Results

We tested our hypotheses by conducting ANOVA with Admission Condition as the independent variable and user reputation and interaction intentions as the dependent variables (for inferential statistics, see Table 2). In Experiment 1A, there were significant Admission Condition effects on user reputation ratings and interaction intentions. Supporting Hypothesis 1, user reputation ratings were significantly higher in the admission condition than the refuse condition. Supporting Hypothesis 2a, interaction intentions were higher in the admission condition than the refuse condition. To test Hypothesis 2b, we performed a mediation analysis, with 10,000 bootstrapping samples, and found a significant indirect effect of Admission Condition on interaction intentions through user reputation ratings (see Table 4).

For Experiment 1B, failing to replicate Experiment 1A, none of our hypotheses were supported.

Discussion and Experiment 2A and Experiment 2B

Our hypotheses were supported in Experiment 1A. It appears that wrongness admission on OSNs serves as a cue of communal and competence traits and leads observers to be willing to interact with the wrongness admitting user. However, we were unable to replicate these findings in Experiment 1B. Importantly, however, wrongness admission did not lead to negative impression formation in this experiment, as user reputation scores were descriptively higher in admission condition than the nonadmission conditions.

We were surprised that we did not find a significant difference in user reputation ratings between the admission and nonadmission conditions in Experiment 1B.

Table 2. Effect of condition on impression formation and means and SDs by condition, Experiment 1A and Experiment 1B

Study (DF)	Experiment 1A (1, 87)							Experiment 1B (2, 235)					
	<i>F</i>	η^2	90% CI	<i>M</i> (<i>SD</i>)		<i>M</i> (<i>SD</i>) _{NR}	<i>F</i>	η^2	90% CI	<i>M</i> (<i>SD</i>)		<i>M</i> (<i>SD</i>) _{NR}	
			<i>M</i> (<i>SD</i>) _{Admit}	Refuse	<i>M</i> (<i>SD</i>) _{Admit}		Refuse			<i>M</i> (<i>SD</i>) _{Admit}	Refuse		
User reputation ratings	6.10*	.07	0.01, 0.17	3.28 (0.75)	2.88 (0.75)	NA	2.12	0.02	0.00, 0.05	2.93 (0.71)	2.70 (0.71)	2.85 (0.73)	
Interaction intentions	8.69**	.09	0.02, 0.20	3.22 (1.01)	2.59 (0.88)	NA	0.63	0.01	0.00, 0.03	2.79 (0.86)	2.89 (0.85)	2.74 (0.92)	

Note. *** $p < .001$; ** $p < .01$; * $p < .05$. DF = Degrees of Freedom, NA = Not applicable, NR = no-response condition.

Table 3. Effect of condition on impression formation and means and SDs by condition, Experiment 2A and Experiment 2B

Study (DF)	Experiment 2A (2, 168)						Experiment 2B (2, 180)					
	<i>F</i>	η^2	90% CI	<i>M</i> (SD) _{Admit}	<i>M</i> (SD) _{Refuse}	<i>M</i> (SD) _{NR}	<i>F</i>	η^2	90% CI	<i>M</i> (SD) _{Admit}	<i>M</i> (SD) _{Refuse}	<i>M</i> (SD) _{NR}
User reputation rating	8.21***	.09	0.03, 0.16	3.24 (0.71)	2.75 (.65)	2.92 (0.64)	5.32**	0.06	0.01, 0.11	3.17 (0.72)	3.05 (0.80)	2.95 (0.17)
Interaction intentions	3.45*	.04	0.00, 0.09	3.14 (0.88)	2.86 (.83)	2.75 (0.82)	1.40	0.02	0.000, 0.051	3.04 (0.72)	2.95 (0.69)	3.05 (0.80)

Note. *** $p < .001$; ** $p < .01$; * $p < .05$. NR = no-response condition.

Therefore, we reviewed the responses to an optional feedback question at the end of both experiments to determine what, if anything, led to differences in the results. Notably, some participants in Experiment 1B indicated that the wrongness admission response of the user seemed sarcastic and insincere. Therefore, in Experiment 2A and Experiment 2B, we edited the final post in the admission and refuse conditions. Participants rated the politeness and sincerity of the final post in Experiment 2A and Experiment 2B. These ratings were well above the midpoint (>6.00 for politeness and >5.50 for sincerity, both out of 7) in the admission condition. We do not report further results with these ratings as we cannot compare them to the participants' reactions to the final posts in the previous experiments and because the ratings were in regard to the post, not the OSN user.

All of our experiments, including Experiment 2A, relied on Amazon Mechanical Turk's participant pool. While this participant pool is more diverse and attentive (Buhrmester et al., 2011), there are problems, too (Arechar et al., 2017). Therefore, the purpose of Experiment 2B was to replicate Experiment 2A in a nononline sample.

Method

Participants

We recruited 191 participants in Experiment 2A from Amazon Mechanical Turk. Of the 191 participants, 171 participants (100 female) completed the experiment and passed the attention check and 98% reported using Facebook. For Experiment 2B, we created 226 slots via a psychology department's online SONA system. Participants were undergraduates at a large university in the Southwest United States. Of the 226 undergraduate students who participated, 182 (116 female) completed the experiment and passed the attention check and 85% reported using Facebook. These sample sizes were determined by our ability to recruit as close to 200 participants as possible. According to a sensitivity analysis, we had the power to detect a minimum effect size of $\eta^2 = .05$ in both studies.

Materials and Procedures

The materials and procedures were nearly identical to those of Experiment 1B. However, we modified the final posts in the admission and refuse conditions. The final post

in the admission condition was as follows: "So, I read through the sources you posted . . . The evidence is pretty strong. I guess I am wrong and you are right on this. Thanks for posting those links and thanks for the conversation!" The final post in the refuse condition was identical except that the third sentence read, "I still think I am right and you are wrong."

We created scores for user reputation ratings (2A: $M = 2.99$, $SD = 0.70$, $\alpha = .94$; 2B: $M = 3.03$, $SD = 0.61$, $\alpha = .92$) and interaction intentions (2A: $M = 2.93$, $SD = 0.86$, $\alpha = .77$; 2B: $M = 3.06$, $SD = 0.74$, $\alpha = .69$), as in the previous experiments. Participants also responded to the same attention check question, which we used to exclude participants who were not paying attention.

Results

We tested our hypotheses by conducting ANOVAs with Admission Condition as the independent variable and user reputation ratings and interaction intentions as the dependent variables (for inferential statistics, see Table 3). In Experiment 2A, there were significant Admission Condition effects on user reputation ratings and interaction intentions. To test whether this condition effect supported Hypothesis 1 and Hypothesis 2a, we ran contrast analyses comparing the admission condition to the two nonadmission conditions. Supporting Hypothesis 1, user reputation ratings were significantly higher in the admission condition than the two nonadmission conditions, $b = 0.13$, $t(168) = 3.81$, $p < .001$. Supporting Hypothesis 2a, interaction intentions were significantly higher in the admission condition than the two nonadmission conditions, $b = 0.11$, $t(168) = 2.52$, $p = .013$. To test Hypothesis 2b, we performed a mediation analysis, with 10,000 bootstrapping samples, with the admit condition contrast coded as 1 and the nonadmit conditions as 0, and found a significant indirect effect of Admission Condition on interaction intentions through user reputation ratings (see Table 4).

In Experiment 2B, there was a significant Admission Condition effect on user reputation ratings, but not interaction intentions. To test whether this condition effect supported Hypothesis 1, we ran contrast analyses comparing the admission condition to the two nonadmission

Table 4. Mediation analyses of condition effects on interaction intentions through user reputation ratings

Experiment	A path	B path	C path	C' path	95% CI for the indirect effects
1A	0.26**	0.85***	0.30**	0.09	0.09, 0.79
1B	—	—	—	—	—
2A	0.28***	0.79***	0.19*	−0.03	0.19, 0.62
2B	0.23**	0.55***	0.11	−0.01	0.07, 0.34

Note. Standardized regression coefficients are depicted for each path (** $p < .01$; *** $p < .001$). A path = condition to mediator; B path = mediator to outcome; C path = condition to outcome; C' path = condition to outcome controlling for the mediator.

conditions. Supporting Hypothesis 1, user reputation ratings were significantly higher in the admission condition than the two nonadmission conditions, $b = 0.09$, $t(180) = 3.12$, $p = .002$. To test Hypothesis 2b, we performed a mediation analysis and found a significant indirect effect of Admission Condition on interaction intentions through user reputation ratings (see Table 4).

Mini Meta-Analysis

Because there were variations in the materials, sample sizes, and effect sizes across the experiments, we conducted a meta-analysis of the effect of admission condition compared to the nonadmission conditions on user reputation scores and interaction intentions to get a truer estimate of the actual effect sizes. We used Goh et al.'s (2016) mini meta-analytic strategy, in which we converted the effect size estimates across studies to r s and then calculated a weighted (by N) mean effect size. The meta-analytic effect size was moderate for reputation, $r = .20$ (95% CI [.14, .26]), and small for interaction intentions, $r = .11$ (90% CI [.05, .17]).

Discussion

Our hypotheses were partially supported in Experiment 2A and Experiment 2B. Wrongness admission in an OSN argument served as a cue of communal and competence traits and led observers to indicate a willingness to interact with the admitting user. However, in Experiment 2B, there was no direct impact of wrongness admission on interaction intentions. Even so, wrongness admission on OSNs appears to lead to positive impression formation outcomes.

General Discussion

With the ubiquitous use of OSNs and the growing presence of arguments on these sites, it is important to understand how online behavior in these online situations impacts impression formation. Here, we investigated the impact of

a form of intellectual humility: wrongness admission. In four studies, we found general support for our hypotheses. Those who witnessed an OSN user engage in wrongness admission rated that user as higher in communion and competence traits compared witnessing a user not engaging in wrongness admission, supporting Hypothesis 1. Furthermore, we found that those in the wrongness admission condition were more likely to indicate interest in interacting with the admitting user compared to those in the nonadmission conditions, supporting Hypothesis 2a. However, this latter effect was weaker, and all effects on interaction intentions across experiments were likely due to participants' increased communion and competence ratings of the user, supporting Hypothesis 2b.

Implications and Theoretical Considerations

It seems obvious that wrongness admission, as opposed to refusing to admit, is a better strategy for those having arguments and managing their impressions on OSNs. Intellectual humility is considered a virtue (Chancellor & Lyubomirsky, 2013; Roberts & Wood, 2007), and people seem to like intellectually humble people (Meagher et al., 2021). The results of Fetterman and Sassenberg (2015) and John et al. (2019) suggest that people know that wrongness admission is the best option. Yet, this behavior is not as common as it should be. In fact, Fetterman et al. (2019) found in a daily diary study that, when given a chance to admit, 43% of participants refused to engage in wrongness admission. When asked why they refuse to admit, most people indicate impression management concerns (John et al., 2019).

If people are hesitant to engage in wrongness admission in face-to-face situations, they may be even more hesitant on OSNs, since many passive viewers can witness the admission and OSNs are a place where people spend considerable time managing their impressions (Krämer & Winter, 2008; Lee & Jang, 2019; Ranzini & Hoek, 2017; Seidman, 2013). Yet, our results suggest that, because wrongness admission serves as a cue of communion and competence, those who might refuse to engage in wrongness admission during an OSN argument might be making an impression management mistake. Wrongness

admission, in these situations, appears to *improve* the impressions formed of users by passive viewers.

It is understandable that these findings might seem counterintuitive to some, especially since the act of wrongness admission is itself an admission of incompetence. We suggest a similar mechanism for wrongness admission that Brooks et al. (2015) suggested for advice seeking. Wrongness admission serves as a cue of intellectual humility, communion, and competence. Although the admitter (like the advice seeker) is telling onlookers that they have been incompetent in this instance, it suggests that they are willing to work together and that they are competent enough to recognize faulty knowledge and change it. Because OSN users are seeking social connections (Alhabash & Ma, 2017) and, perhaps, future friends (Bacev-Giles & Haji, 2017), they are forming impressions online (Tennie et al., 2010) with relatively little contextual information (Walther & Whitty, 2021). Any behavior that serves as a cue of these communal and competence traits should lead to positive impression formation. Wrongness admission, as we have noted, is one such cue.

Wrongness admission on OSNs might be important for reasons beyond impression formation and management, however. As more people engage in debate on the internet, the more OSNs have an impact on behaviors and the spread of misinformation (Allcott & Gentzkow, 2017). OSN-based debates do not shy away from controversial and consequential topics such as politics, religion, or other important societal issues (Anderson et al., 2018; Neubaum et al., 2021), and these debates can be equally frustrating and engaging (Wang & Silva, 2018). If people avoid wrongness admission in these debates, there will likely be no, or worse, resolution to these issues. It could further affect the widening division in the world of politics (Van Bavel & Pereira, 2018) and inspire violent reactions to false information – particularly for those who become more extreme in their faulty beliefs instead of engaging in wrongness admission (i.e., the Backfire Effect; Nyhan & Reifler, 2010). Further research is underway to investigate ways to encourage wrongness admission on OSNs.

Limitations and Future Directions

In the current experiments, we focused on internal, rather than on external, validity. We wanted to limit the impact of other information readily available on OSNs (e.g., profiles, pictures, and demographics). This decision is in line with prior research investigating the effects of minimal information on OSNs (e.g., Austin et al., 2021; Kaye et al., 2020). However, it limits the current investigation because it is unclear how this extra information might impact

impression formation in relation to wrongness admission on OSNs. For example, it could be that the identity of the admitting user moderates the effects we found here in some way. Future research is needed, which subtly adds in information to the online stimuli to test this and other hypotheses. Indeed, we consider the current investigation as a launching point for much more research on the outcomes of wrongness admission and intellectual humility on OSNs.

In a similar vein, our studies were limited to supposedly real, but fake, topics and an argument between strangers. This means that the participants did not have prior attitudes toward the topic of conversation, nor the users. It is likely that prior attitudes toward both affect the impressions formed by passive viewers of OSN users who engage in wrongness admission. For example, it may be that Donald Trump engaging in wrongness admission on OSNs is seen positively by his supporters and negatively by his detractors – or vice versa, given the reactions to his late 2021 support of COVID-19 vaccinations. It may be that a user engaging in wrongness admission about the benefits of intermittent fasting is seen negatively by intermittent fasters, but positively by nonfasters. Future research should investigate these possibilities.

We presented mediation models suggesting that wrongness admission (vs. nonadmission) led to better user reputation ratings, which then related to interaction intentions. From a theoretical perspective, this makes sense. The purpose of impression formation is to determine who one wants to interact with, even on OSNs (Tennie et al., 2010). However, we did not test the full causal model, and many have cast doubts on the type of cross-sectional mediation analyses we conducted (e.g., Maxwell & Cole, 2007; Rohrer, 2018). Therefore, we suggest due caution in interpreting these results, consider them preliminary, and encourage further research investigating the full causal chain.

Conclusion

People are wrong a lot and for a variety of reasons (Schulz, 2010). However, the ratio of instances of wrongness to admissions is far from equal (Fetterman et al., 2019). It is likely that people do not want to broadcast their incompetence on OSNs. However, wrongness admission on OSNs not only allows people to change their attitudes to become more factual (being competent) but also leads to the formation of more positive impressions (appearing communal and competent). Therefore, wrongness admission on OSNs appears to lead to better impression formation outcomes than not admitting. At least, that is

what we can conclude until someone provides evidence that we are wrong. If such a time comes, we will never admit it.

References

- Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness-communion with warmth and morality. *Frontiers in Psychology*, 7, 1810. <https://doi.org/10.3389/fpsyg.2016.01810>
- Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in experimental social psychology* (Vol. 50, pp. 195–255). Academic Press. <https://doi.org/10.1016/b978-0-12-800284-1.00004-7>
- Alhabash, S., & Ma, M. (2017). A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? *Social Media + Society*, 3(1), 2056305117691544. <https://doi.org/10.1177/2056305117691544>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, 111(2), 256–274. <https://doi.org/10.1037/0033-2909.111.2.256>
- Anderson, M., Toor, S., Ranie, L., & Smith, A. (2018). *Activism in the social media age*. Pew Research Center. <http://www.pewinternet.org/2018/07/11/activism-in-the-social-media-age/>
- Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. (2017). Turking overtime: How participant characteristics and behavior vary over time and day on Amazon mechanical Turk. *Journal of the Economic Science Association*, 3(1), 1–11. <https://doi.org/10.1007/s40881-017-0035-0>
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41, 258–290. <https://doi.org/10.1037/h0055756>
- Austin, A. B., Costabile, K. A., & Smith, L. (2021). Social judgments, social media, and self-deprecation: Role of information source and valence on trait and favorability judgments. *Journal of Media Psychology: Theories, Methods, and Applications*. <https://doi.org/10.1027/1864-1105/a000299>
- Bacev-Giles, C., & Haji, R. (2017). Online first impressions: Person perception in social media profiles. *Computers in Human Behavior*, 75, 50–57. <https://doi.org/10.1016/j.chb.2017.04.056>
- Bareket-Bojmel, L., Moran, S., & Shahar, G. (2016). Strategic self-presentation on Facebook: Personal motives and audience response to online behavior. *Computers in Human Behavior*, 55, 788–795. <https://doi.org/10.1016/j.chb.2015.10.033>
- Batenburg, A., & Bartels, J. (2017). Keeping up online appearances: How self-disclosure on Facebook affects perceived respect and likability in the professional context. *Computers in Human Behavior*, 74, 265–276. <https://doi.org/10.1016/j.chb.2017.04.033>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, 61(6), 1421–1435. <https://doi.org/10.1287/mnsc.2014.2054>
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.
- Buffardi, L. E., & Campbell, W. K. (2008). Narcissism and social networking web sites. *Personality and Social Psychology Bulletin*, 34(10), 1303–1314. <https://doi.org/10.1177/0146167208320061>
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk. *Perspectives on Psychological Science*, 6(1), 3–5. <https://doi.org/10.1177/1745691610393980>
- Chancellor, J., & Lyubomirsky, S. (2013). Humble beginnings: Current trends, state perspectives, and hallmarks of humility. *Social and Personality Psychology Compass*, 7(11), 819–833. <https://doi.org/10.1111/spc3.12069>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, 40, 61–149. [https://doi.org/10.1016/S0065-2601\(07\)00002-0](https://doi.org/10.1016/S0065-2601(07)00002-0)
- Duggan, M., & Smith, A. (2016). *The political environment on social media*. Pew Internet & American Life Project. <http://www.pewinternet.org/2016/10/25/the-political-environment-on-social-media>
- Fetterman, A. K., Curtis, S., Carre, J., & Sassenberg, K. (2019). On the willingness to admit wrongness: Validation of a new measure and an exploration of its correlates. *Personality and Individual Differences*, 138, 193–202. <https://doi.org/10.1016/j.paid.2018.10.002>
- Fetterman, A.K. (2021). *Data for "When you are wrong on Facebook, just admit it wrongness admission leads to better interpersonal impressions on social media"*. <https://osf.io/a4tgp/>
- Fetterman, A. K., & Sassenberg, K. (2015). The reputational consequences of failed replications and wrongness admission among scientists. *PLoS One*, 10(12), e0143723. <https://doi.org/10.1371/journal.pone.0143723>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177–182. <https://doi.org/10.1177/0963721412445309>
- Gable, S. L., & Reis, H. T. (2001). Appetitive and aversive social interaction. In J. Harvey & A. Wenzel (Eds.), *Close romantic relationships: Maintenance and enhancement* (pp. 169–194). Erlbaum.
- Gigerenzer, G., & Kurz, E. M. (2001). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 342–348). Oxford University Press.
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10(10), 535–549. <https://doi.org/10.1111/spc3.12267>
- Hagá, S., & Olson, K. R. (2017). 'If I only had a little humility, I would be perfect': Children's and adults' perceptions of intellectually arrogant, humble, and diffident people. *The Journal of Positive Psychology*, 12(1), 87–98. <https://doi.org/10.1080/17439760.2016.1167943>
- Harris, K., & Vazire, S. (2016). On friendship development and the Big Five personality traits. *Social and Personality Psychology Compass*, 10(11), 647–667. <https://doi.org/10.1111/spc3.12287>
- John, L. K., Jeong, M., Gino, F., & Huang, L. (2019). The self-presentational consequences of upholding one's stance in spite of the evidence. *Organizational Behavior and Human Decision Processes*, 154, 1–14. <https://doi.org/10.1016/j.obhdp.2019.07.001>
- Kaye, L. K., Wall, H. J., & Hird, A. T. (2020). Less is more when rating extraversion: Behavioral cues and interpersonal perceptions on the platform of facebook. *Psychology of Popular Media*, 9(4), 465–474. <https://doi.org/10.1037/ppm0000263>
- Krämer, N. C., & Winter, S. (2008). Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media*

- Psychology*, 20(3), 106–116. <https://doi.org/10.1027/1864-1105.20.3.106>
- Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., Deffler, S. A., & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin*, 43(6), 793–813. <https://doi.org/10.1177/0146167217697695>
- Lee, S. Y., & Jang, K. (2019). Antecedents of impression management motivations on social network sites and their link to social anxiety. *Media Psychology*, 22(6), 890–904. <https://doi.org/10.1080/15213269.2019.1580588>
- Levodashka, A., & Utz, S. (2017). Spontaneous trait inferences on social media. *Social Psychological and Personality Science*, 8(1), 93–101. <https://doi.org/10.1177/1948550616663803>
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12(1), 23–44. <https://doi.org/10.1037/1082-989X.12.1.23>
- Meagher, B. R., Leman, J. C., Heidenga, C. A., Ringquist, M. R., & Rowatt, W. C. (2021). Intellectual humility in conversation: Distinct behavioral indicators of self and peer ratings. *The Journal of Positive Psychology*, 16(3), 417–429. <https://doi.org/10.1080/17439760.2020.1738536>
- Munro, K., Hartt, C. M., & Pohlkamp, G. (2015). Social media discourse and genetically modified organisms. *The Journal of Social Media in Society*, 4(1), 38–65.
- Neubauer, G., Cargnino, M., Winter, S., & Dvir-Gvirsman, S. (2021). “You’re still worth it”: The moral and relational context of politically motivated unfriending decisions in online networks. *PLoS One*, 16(1), e0243049. <https://doi.org/10.1371/journal.pone.0243049>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Orben, A. C., & Dunbar, R. I. M. (2017). Social media and relationship development: The effect of valence and intimacy of posts. *Computers in Human Behavior*, 73, 489–498. <https://doi.org/10.1016/j.chb.2017.04.006>
- Porter, T., Baldwin, C. R., Warren, M. T., Murray, E. D., Cotton Bronk, K., Forgeard, M. J. C., Snow, N. E., & Jayawickreme, E. (2021). Clarifying the content of intellectual humility: A systematic review and integrative framework. *Journal of Personality Assessment*, 1–13. <https://doi.org/10.1080/00223891.2021.1975725>
- Qin, Y., Cho, H., Li, P., & Zhang, L. (2021). First impression formation based on valenced self-disclosure in social media profiles. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.656365>
- Quadflieg, S., & Penton-Voak, I. S. (2017). The emerging science of people-watching: Forming impressions from third-party encounters. *Current Directions in Psychological Science*, 26(4), 383–389. <https://doi.org/10.1177/0963721417694353>
- Ranzini, G., & Hoek, E. (2017). To you who (I think) are listening: Imaginary audience and impression management on Facebook. *Computers in Human Behavior*, 75, 228–235. <https://doi.org/10.1016/j.chb.2017.04.047>
- Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford University Press.
- Rodriguez, D., Hook, J. N., Farrell, J. E., Mosher, D. K., Zhang, H., Van Tongeren, D. R., Davis, D. E., Aten, J. D., & Hill, P. C. (2019). Religious intellectual humility, attitude change, and closeness following religious disagreement. *The Journal of Positive Psychology*, 14(2), 133–140. <https://doi.org/10.1080/17439760.2017.1388429>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Schulz, K. (2010). *Being wrong: Adventures in the meaning of error*. ECCO.
- Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, 54(3), 402–407. <https://doi.org/10.1016/j.paid.2012.10.009>
- Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2013). Being popular in online social networks: How agentic, communal, and creativity traits relate to judgments of status and liking. *Journal of Research in Personality*, 47(5), 592–598. <https://doi.org/10.1016/j.jrp.2013.05.005>
- Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2014). Personality expression and impression formation in online social networks: An integrative approach to understanding the processes of accuracy, impression management and meta-accuracy. *European Journal of Personality*, 28(1), 73–94. <https://doi.org/10.1002/per.1935>
- Tennie, C., Frith, U., & Frith, C. D. (2010). Reputation management in the age of the world-wide web. *Trends in Cognitive Sciences*, 14(11), 482–488. <https://doi.org/10.1016/j.tics.2010.07.003>
- Utz, S. (2010). Show me your friends and I will tell you what type of person you are: How one’s profile, number of friends, and type of friends influence impression formation on social network sites. *Journal of Computer-Mediated Communication*, 15(2), 314–335. <https://doi.org/10.1111/j.1083-6101.2010.01522.x>
- Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, 22(3), 213–224. <https://doi.org/10.1016/j.tics.2018.01.004>
- Waggoner, A. S., Smith, E. R., & Collins, E. C. (2009). Person perception by active versus passive perceivers. *Journal of Experimental Social Psychology*, 45(4), 1028–1031. <https://doi.org/10.1016/j.jesp.2009.04.026>
- Walther, J. B., Van Der Heide, B., Hamel, L. M., & Shulman, H. C. (2009). Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using facebook. *Communication Research*, 36(2), 229–253. <https://doi.org/10.1177/0093650208330251>
- Walther, J. B., & Whitty, M. T. (2021). Language, psychology, and new media: The hyperpersonal model of mediated communication at twenty-five years. *Journal of Language and Social Psychology*, 40(1), 120–135. <https://doi.org/10.1177/0261927X20967703>
- Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, 81, 73–83. <https://doi.org/10.1016/j.chb.2017.11.041>

History

Received June 11, 2021

Revision received January 3, 2022

Accepted February 3, 2022

Published online March 18, 2022

Authorship

Adam K. Fetterman and Nicole L. Muscanell conceived of and designed the studies. A.K. Fetterman managed the data collection. A.K. Fetterman and Dongjie Wu conducted the analyses. Adam K. Fetterman led the writing of the manuscript. All authors edited the manuscript.

Open Data

We report how we determined our sample sizes, all data exclusions, all manipulations, and all measures in the studies. Data, code, and materials can be accessed through the Open Science Framework (OSF): <https://osf.io/a4tpg/>. We did not preregister our hypotheses.

ORCID

Adam K. Fetterman

 <https://orcid.org/0000-0001-5957-7670>

Adam K. Fetterman

Department of Psychology

University of Houston

3695 Cullen Boulevard

Houston, TX 77204-5022

USA

akfetterman@uh.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.