

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

When Emotional Intensity Is High, Feelings Start to Look Like Facts

Permalink

<https://escholarship.org/uc/item/8k97300v>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Xiong, Tianyi

Tang, Xianglu

Wylie, Jordan

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

When Emotional Intensity Is High, Feelings Start to Look Like Facts

Tianyi Xiong (tx224@cornell.edu)^{1,3} & Xianglu Tang² & Jordan Wylie (jordan.wylie@cornell.edu)¹

¹Department of Psychology, Cornell University

²Department of Psychology, University of Michigan

³Applied Psychology, Beijing Normal-Hong Kong Baptist University

Abstract

Moral conflicts frequently involve ambiguous evidence, creating opportunities for emotion to shape what people find credible. Across three studies, we examine how emotional expressions influence the perceived credibility and usability of moral evidence. In Study 1, we analyze posts from the Reddit forum *r/AmITheAsshole*, finding that readers expressed more trust in posts that were emotionally intense. In Study 2, we use a social media-style post to manipulate emotional intensity to test perceived credibility of subjective and objective evidence. High emotional intensity of expressions narrowed the perceived difference in credibility between subjective and objective evidence. In Study 3, we find that emotional intensity also enhances evidence usability—higher perceived causality and counterfactual necessity are attributed to a person who delivers evidence with high emotional intensity (vs. low). This research highlights how the intensity of emotional expressions influences both the evaluation and use of moral evidence.

Keywords: morality; emotion; evidence

Introduction

Emotions play a large role in how people navigate their social worlds. People rely on them to communicate with others (e.g., Izard et al., 2001; Hess et al., 1995), infer what others are thinking (e.g., Ong et al., 2019; Van Kleef, 2010), and form expectations about how people will act in the future (Levine et al., 2018). How people display, recognize, decode, and interpret emotion is critical to adaptive social functioning (e.g., Izard et al., 2001; Jones et al. 2011), and how people evaluate information (see Clore & Huntsinger, 2007; Clore et al., 2012). Emotions shape what people feel is important (Forgas, 1995), bad or good (Frijda et al., 2000), and even morally significant (Wisneski & Skitka, 2017; Hofmann & Baumert, 2010).

Emotions also influence how people engage with information in the first place. People are more likely to avoid information they expect to be negative—to avoid encountering information which may affect their emotional state (see Golman et al., 2017 for review). Emotions, including the affective load elicited by others' emotional disclosures, can tax people's cognitive (e.g., Plass & Kalyuga, 2019; Pessoa, 2009), emotional resources (Mazzetti et al., 2020), and reduce the effort people devote to comparing the quality of competing claims (see Greifeneder et al., 2011). Emotion shapes what people engage with.

Emotional displays also shape whose claims are seen as credible or compelling (e.g., Kaufmann et al., 2003). For example, research finds that people tend to find an argument more persuasive when it is presented as an emotional appeal rather than as objective (i.e., facts; Kubin et al., 2021). In courtroom contexts, jurors tend to trust witnesses who

display emotion more than those who remain calm when testifying about heinous crimes (e.g., Salerno, 2021; Nitschke et al., 2019). That is, the intensity of emotional expressions both influences what information people engage with, and how they evaluate that information.

The moral domain is a particularly important context for understanding emotional intensity because moral decisions frequently involve emotional stimuli (e.g., Brady et al., 2017) and ambiguous evidence (Wong, 1992), which create opportunities for emotion signals to heavily influence interpretation (e.g., Greene et al., 2001; Blanchette & Richards, 2010; see also Haidt, 2001). When facts and clear standards are lacking, people rely on emotion as a cue to judge which arguments are factual and which evidence can be used to persuade others (see also Haidt, 2001). While prior work shows that emotions broadly shape moral reasoning—often more so for negative than positive emotions (e.g., Prinz, 2006; Ugazio et al., 2012; Cummins & Cummins, 2012; Yu et al., 2019)—far less is known about how the intensity of emotional expressions affects the evaluation of moral evidence in everyday moral conflicts.

Here, we examine two key aspects of moral evidence evaluation: perceived credibility (how much the evidence appears to be true) and perceived usability (how likely people are to consider using it to support a moral judgment). We predicted that both aspects are susceptible to the influence of intense emotional expressions. When emotional expressions are intense, people may judge evidence as more credible and thus may be more likely to see it as relevant, usable evidence when forming a moral judgment (Avramova & Inbar, 2013).

The Present Research

The present research integrates large-scale observational data and experimental methods to test whether the intensity of emotional expressions enhances the perceived credibility and usability of moral evidence. We examine this question across different types of negative emotion for care-related moral dilemmas. We chose care-related moral dilemmas because we expected that emotion language would be more natural for such dilemmas, and because research suggests negative emotions are more strongly associated with care-based moral reasoning than with other types of reasoning (e.g., fairness; e.g., Skoe et al., 2002). In Study 1, we analyze 1,794 posts and over 300,000 comments from the Reddit forum *Am I the Asshole* (AITA), a subreddit that provides a large, naturally occurring dataset of everyday moral dilemmas and judgments (see also Yudkin et al., 2025). We use natural language processing (NLP) to

examine whether the emotional intensity of a post influences commenters' trust in its arguments. We predicted that there would be a positive association between the frequency of expressing stronger negative emotion and linguistic markers of trust from commenters in everyday care-related moral dilemmas. In Study 2, we experimentally manipulate the intensity of emotional displays through post content and examine its influence on credibility judgments. We predicted that intense emotional expressions would lead to reduced differentiation between subjective and objective evidence, meaning people would rate subjective evidence as more credible than those where the intensity of an emotional expression is low. Finally, in Study 3, we test whether intensity of emotional expression also increases the perceived usability of moral evidence. That is, whether people not only see emotional signals as credible, but whether they think others see them as such as well. We predicted that an intense emotional expression (vs. control) would lead to increased evidence usability judgments. All studies were approved by our institutional IRB, and we report all sample sizes, exclusion criteria, manipulations, and measures in the manuscript.

Study 1

Study 1 aimed to establish that emotional intensity is generally related to linguistic markers of trust in care-related moral dilemmas using large-scale reddit data. Care-related posts here refer to reddit posts about the Care/Harm foundation from the Moral Foundations Theory (Graham et al., 2011), which involve concerns about others' welfare, suffering, or the perceived failure to provide care. We analyzed 1,794 labeled care-related posts and over 300,000 corresponding reader comments from the subreddit r/AITA. We predicted that there would be a positive relationship between the intensity of negative emotion in posts and the trust expressed in comments as indexed by a transformer-based trust score.

Methods

Dataset We used an r/AITA dataset (Liew, 2021), which contains posts and comments collected between 2018 and 2021. Using the *extended Moral Foundations Dictionary* (eMFD; Hopp et al., 2021), we extracted 1,794 care-related posts, along with over 300,000 associated comments. Posts with fewer than five comments were excluded. We then applied text preprocessing including personal information removal, lemmatization, stopword removal, and tokenization (Naseem et al., 2021).

Design and Procedure We first measured the emotional intensity of the post using the negative emotion (*negemo*) score in the *Linguistic Inquiry and Word Count Dictionary* (LIWC; Pennebaker et al., 2015; Tausczik & Pennebaker, 2010). The emotional intensity in each post was intended to reflect the strength of emotion that the post elicits in readers.

We then estimated commenter trust using a zero-shot transformer-based classifier (Hugging Face, 2024; Lewis et

al., 2020). The model first labeled each comment as either "believes the post" or "does not believe the post," with a confidence score. We then linearly mapped each output onto a 7-point scale. Comments labeled as "believes the post" were scored from 1 to 7 based on classifier confidence, while those labeled as "does not believe" were inversely scaled.

Next, we integrated the overall trust scores for each post by averaging the scores across all comments. To examine the relationship between emotional intensity and commenter trust, we conducted a partial Pearson correlation while controlling for total word count and pronoun usage (Tausczik & Pennebaker, 2010).

Results

We found that posts expressing higher emotional intensity were more likely to receive trust from commenters ($r = .12$, $p < .001$; see Figure 1). This suggests that in care-related moral dilemmas, higher emotional intensity of negative emotional expressions may increase the perceived trustworthiness of subjective moral arguments.

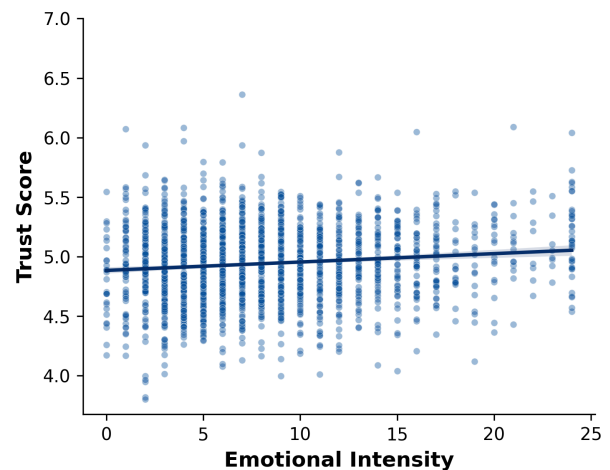


Figure 1: Emotional intensity predicts greater trust in posts with care-related moral dilemmas. Individual data points reflect the raw comments. Shading reflects the 95% confidence interval.

Interim Discussion

Overall, Study 1 provides evidence for our prediction: emotional intensity in posts increases trust from readers in everyday care-related moral dilemmas. However, since the findings are correlational, the causal role of emotion remains uncertain. In Study 2, we sought to directly test whether the emotional intensity of expressions increases the perceived credibility of moral evidence.

Study 2

In Study 2, we manipulated the intensity of emotional displays in a post while holding other content constant. We sought to test whether high emotional intensity leads people

to treat subjective evidence as if it were objective. One reason to expect this pattern is because emotionally intense content may affect people's mood and increase reliance on affective cues (Clore & Huntsinger, 2007; Clore et al., 2012). This may increase trust in otherwise subjective arguments. In contrast, objective evidence may be less likely to show these same biases due to the verifiability of it (Brashier & Marsh, 2020). As such, we predicted that high emotional intensity would lead to more credibility for subjective evidence compared to low emotional intensity, and that this effect would emerge for the emotion signaler (i.e., the "poster"), not the target of that signal (i.e., "the friend").

Methods

Participants We first conducted a power analysis using G*Power 3.1 (Faul et al., 2007) and found that 100 participants was sufficient to detect a mean difference in subjective vs. objective evidence differentiation between high and low emotional intensity conditions (with $d = 0.50$; alpha set to .05; power set to 80%). So we recruited a total of 102 participants who are Chinese and between 18 and 25 years old through *Credamo*, an online experiment platform in China. After excluding participants who failed the attention check ($N = 1$), we performed the analysis on a final sample of 101 participants ($M_{\text{age}} = 23.33$, $SD_{\text{age}} = 2.45$; Female = 72, Male = 29).

Design and Procedure We used a between-subjects design to examine whether emotional intensity of expressions influences perceived credibility of evidence in a care-related everyday moral dilemma. Consenting participants read about a care-related moral conflict between a "poster" and a close friend in the style of a Reddit post. The post describes a situation where someone's friend did not visit them during their surgery and hospital stay. Participants were randomly assigned to read either a high-emotion or low-emotion version of the post. We manipulated emotional intensity by adding crying emojis and emotional expressive wording in one version of the post (high-intensity) vs. not adding this information (low-intensity). After reading the post, participants first completed a manipulation check of the intensity of emotional expressions (how sad, angry, disappointed, and emotionally negative overall the poster is; rated on a 0 = *Not at all* to 100 = *Extremely* scale). We collapsed these items into a single index of negative emotion (Cronbach's alpha = .84), excluding anger because it was not sufficiently correlated with the other items.

After completing the manipulation checks, participants initiated a direct message (DM) conversation with the poster. They selected a question to ask, and the poster responded with a specific piece of evidence (see Figure 2).

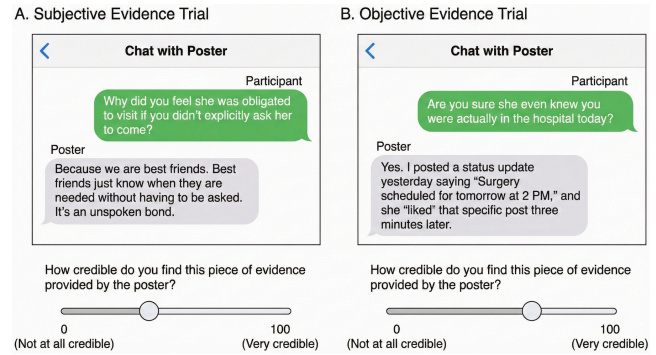


Figure 2: Examples of the credibility judgment task. Stimuli examples have been translated from Chinese.

After reading each piece of evidence, participants completed the evidence credibility judgment task. Participants rated a total of 12 pieces of evidence (6 for each side of the conflict) on a slider from 0 = *Not at all credible* to 100 = *Very credible*. For each side of the conflict (i.e., for the "poster" and "friend"), one objective item (i.e., a direct quote) served as a baseline and five were subjective (see Figure 2 for examples). All evidence arguments were provided by the poster through the DM. Evidence for the poster's side justified why the friend should have visited, while evidence for the friend's side offered alternative explanations for why she did not. Evidence was shown one at a time in randomized order, and brief dialogues were included to simulate online conversation. Figure 3 shows an overview of the experimental procedure. Participants were compensated upon completion of the study.

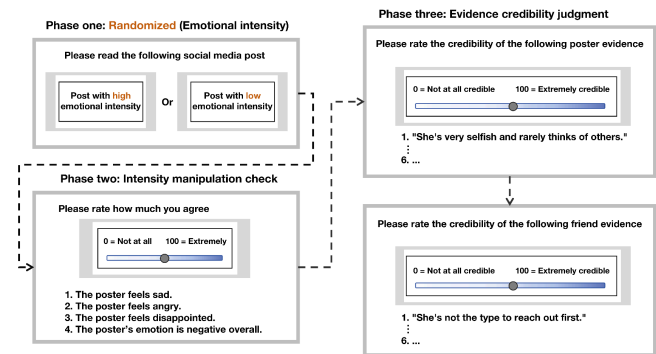


Figure 3: Overview of the experimental procedure. Stimuli examples have been translated from Chinese.

Results

Emotional Expression Intensity Manipulation Check

Results indicated that the manipulation of emotional intensity was successful. Participants in the high-intensity condition rated the poster as overall more emotionally negative ($b = 7.25$, $SE = 2.12$, $t(99) = 3.42$, $p < .001$, 95% $CI [3.04, 11.46]$, $d = 0.68$) than those in the low-intensity condition.

Subjective and Objective Evidence Differentiation We fit a linear regression model to compare people's differentiation between subjective and objective evidence as the outcome and emotional intensity of expressions as the predictor. We used the objective evidence credibility rating as a baseline. Difference scores were calculated as objective minus the average of subjective credibility ratings, such that lower scores indicate that subjective evidence was rated as more credible, approaching the credibility level of the objective ones. As predicted, participants under the high emotional intensity condition rated subjective evidence used by the poster as more credible, $b = -8.96$, $SE = 3.87$, $t(99) = -2.32$, $p = .023$, $d = -0.46$, 95% CI $[-16.64, -1.28]$ (see Figure 4). However, for the evidence that justifies the friend's side, there was no significant difference between conditions ($b = -1.47$, $SE = 3.90$, $p = .71$, $d = -0.08$). That is, high emotional intensity of expressions reduces people's differentiation between objective and subjective moral evidence when the evidence supports the emotionally intense side.

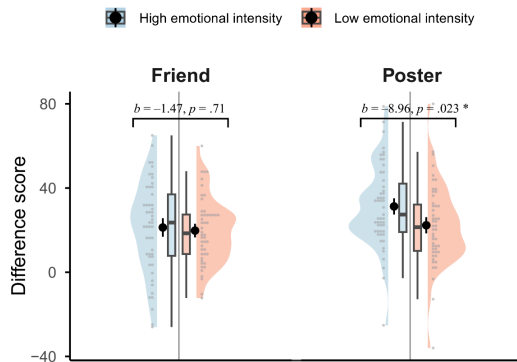


Figure 4: Emotional intensity of expressions increases the perceived credibility of subjective evidence. * $p < .05$.

Interim Discussion

Study 2 provides experimental support for the idea that emotional intensity of expressions can distort how people evaluate moral evidence. Much like in the observational data from Study 1, when participants read evidence delivered by an emotionally intense person, they are more likely to perceive it as credible. However, it is possible that the evidence being used by the friend (vs. the poster) differs in the quality of evidence vs. being uniquely affected by emotional intensity. As such, we matched the evidence quality in Study 3.

Study 3

Study 3 examined whether the perceived usability of evidence is affected by emotional intensity when evidence quality was matched. We test this using two different operationalizations of usability: the causal strength of moral evidence (Experiment 3a; i.e., whether the evidence can support the moral judgment being made; see Engelmann &

Waldmann, 2022), and counterfactual necessity (Experiment 3b; i.e., whether the moral judgment would have changed if the individual had taken this evidence into account; Phillips & Goodman, 2006). Then we integrated these two measures in a Bayesian model for a measure of overall usability. We predicted that participants under higher emotional intensity would perceive the evidence as (a) more causally supportive of the judgment, (b) more necessary to prevent opposing judgments, and (c) more usable overall.

Methods

Participants We recruited over 100 participants for each experiment to ensure a sufficient sample for the Bayesian analyses (Kruschke & Liddell, 2018), using the same inclusion criteria and effect size estimate as in Study 2. In Experiment 3a, after excluding people who failed attention check ($N = 9$), we conducted analyses on a final sample of 107 participants ($M_{\text{age}} = 22.66$, $SD_{\text{age}} = 2.09$; Female = 88, Male = 19). In Experiment 3b, after excluding people who failed attention check ($N = 11$), we conducted analyses on a final sample of 109 participants ($M_{\text{age}} = 22.42$, $SD_{\text{age}} = 1.85$; Female = 71, Male = 38).

Design and Procedure Both Experiment 3a and 3b used the same moral dilemma (the "post") as in Experiment 2. We again manipulated the intensity of emotional displays, but this time participants saw both high and low emotional intensity conditions in the same conversation (i.e., within-compared to between-subjects). This update is to control for between-subject variability in participants' use of the rating slider. For both Experiment 3a and Experiment 3b, the poster is always expressing high emotional intensity and the friend is expressing low emotional intensity. We also conducted pilot studies to collect paired evidence arguments to ensure that evidence used by the friend and poster was balanced (aside from emotional information) between the poster and the friend (i.e., the two sides of the moral conflict) and selected 6 balanced pairs (p 's $> .05$) to use in both studies.

After reading the post, participants completed the same manipulation check items as in Experiment 2. Then, participants either completed a causal judgment task (Experiment 3a) or a counterfactual necessity judgment task (Experiment 3b). In the causal judgment task, participants saw the following message from the poster: "A user online said my friend really should have come to visit." We then told participants that this meant they had shared evidence that made others side with them. Participants then saw the six pieces of evidence that the poster shared, one at a time. Each time, they judged how likely it was that the user supported the poster because of that piece of evidence. Responses were recorded on a slider from 0 = *Not at all likely* to 100 = *Very likely* (see Figure 5).

After rating all six pieces of poster-side evidence, participants were told to imagine a second person, A, who had chatted with the friend and sided with her, agreeing that it was reasonable for the friend not to visit the poster.

Participants then saw six pieces of evidence the friend might have shared with person A (e.g., "The hospital was far from my company, and I had to stay late that entire week") and rated how likely A's support for the friend was due to each piece of evidence.

The counterfactual necessity task followed the same procedure as the causal judgment task. However, the statements were framed counterfactually. That is, we told participants that a user opposed the poster after talking with her, and then asked them to rate: "The online user wouldn't have opposed her if the poster had shared the following evidence." We presented a similar scenario for the friend side and asked participants to rate: "A wouldn't have opposed her if the friend had shared the following reason." Participants received compensation upon completing the study.

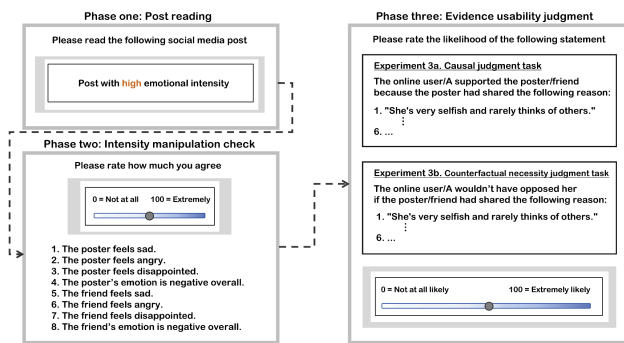


Figure 5: Overview of Experimental Procedure, including the examples of the causal (Experiment 3a) and counterfactual necessity (Experiment 3b) judgment task. Stimuli examples have been translated from Chinese.

Results

Emotional Expression Intensity Manipulation Check As in Experiment 2, we collapsed the emotion items into a single negative emotion index in both experiments (Experiment 3a: Poster Cronbach's alpha = .67; Friend Cronbach's alpha = .83; Experiment 3b: Poster Cronbach's alpha = .78; Friend Cronbach's alpha = .90). We again dropped anger for fit issues.

Results indicated that the emotional-intensity manipulation was successful in both Experiment 3a and 3b. In the causal judgment task (Experiment 3a), participants rated the poster as overall more emotionally negative ($b = 66.78$, $SE = 1.82$, $t(106) = 37.09$, $p < .001$, $95\% CI [63.21, 70.35]$, $d = 3.59$) than the friend. A similar pattern emerged in the counterfactual necessity judgment task (Experiment 3b): participants again perceived the poster as overall more emotionally negative ($b = 61.80$, $SE = 2.22$, $t(108) = 27.83$, $p < .001$, $95\% CI [57.40, 66.21]$, $d = 2.67$) than the friend.

Causal Judgment Task We first tested whether people perceived evidence from the high emotional intensity side as more causal in producing a supportive outcome. To test this, we fit a linear regression model with intensity side (poster =

high intensity vs. friend = low intensity) as a fixed effect. Results showed that participants gave higher causal ratings to evidence from the high emotional intensity side, $b = 13.54$, $SE = 1.02$, $t(106) = 13.26$, $p < .001$, $95\% CI [11.53, 15.54]$, $d = 1.03$.

Counterfactual Necessity Judgment Task We also examined counterfactual necessity judgments across different emotional intensity conditions. We again fit a linear regression model with the emotional intensity side as the predictor. As predicted, participants gave higher counterfactual necessity ratings to evidence from the high emotional intensity side, $b = 10.75$, $SE = 1.11$, $t(108) = 9.69$, $p < .001$, $95\% CI [8.57, 12.93]$, $d = 0.69$. This suggests that emotional intensity increases the perceived counterfactual necessity of moral evidence.

Bayesian Usability Estimation We then integrated the causal and counterfactual necessity ratings using a Bayesian model in PyMC to estimate the overall usability (θ_i) of each of the 12 evidence items. In this model, ratings were predicted by a latent usability parameter for each item, with a fixed effect for task type (causal vs. counterfactual necessity). We then compared posterior usability estimates between the high- and low-emotional-intensity conditions within each evidence pair. To test this, we examined the mean usability difference, the 95% Highest Density Interval (HDI), and the posterior probability that the evidence from the high-intensity side was more usable than the evidence from the low-intensity side.

Across four of six evidence pairs, mean usability differences ranged from approximately 2.99 to 5.77, with posterior probabilities $P(\theta_{\text{high}} > \theta_{\text{low}})$ exceeding 0.99 for each pair (see Figure 6). These results suggest that participants perceived evidence from the high emotional intensity side as more usable, as reflected by an integrated index of causal and counterfactual judgments.

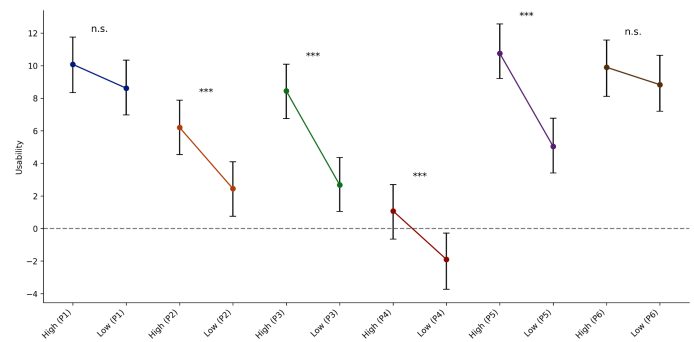


Figure 6: Posterior estimates of evidence usability by the emotional intensity of expressions. *** $P(\theta_{\text{high}} > \theta_{\text{low}}) > .99$, "n.s." indicates $P(\theta_{\text{high}} > \theta_{\text{low}})$ did not exceed .95. Points depict posterior means of usability (θ_i) for each evidence, with error bars indicating 95% HDIs.

Interim Discussion

Overall, results from Study 3 supported our hypotheses. For both causal and counterfactual necessity judgments, participants considered evidence from the high emotional intensity side as causing supportive moral judgments and as necessary to prevent opposing ones. We also found that most evidence pairs showed higher usability ratings for the emotional side using the Bayesian usability index. The only exceptions were found in Pairs 1 and 6, which described concrete reasons why the friend did not visit the poster (e.g., limited time). This suggests that the effect of emotional intensity may be limited to instances where the evidence is subjective or ambiguous.

General Discussion

Across three studies, we examined whether emotional intensity influences how people evaluate moral evidence in care-related moral dilemmas. Across a large-scale Reddit observational study and two experiments, we find emotional intensity increases both trust in and use of moral evidence. In Study 1, emotional intensity in Reddit posts predicted higher trust from readers. In Study 2, participants under the high emotional intensity condition judged subjective evidence as more credible. Finally, in Study 3, evidence from the high emotional intensity side was rated as more causally supportive and counterfactually necessary, suggesting greater usability in moral reasoning.

Emotion plays a central role in shaping how people approach and interpret information (e.g., Clore et al., 2012; Golman et al., 2017; Kubin et al., 2021). Our findings add to this literature by demonstrating that subjective evidence in moral contexts is similarly affected by emotional information. We found that the effect of emotional intensity occurred only for evidence from the emotionally intense side, and only in dilemmas in which the evidence is subjective. People see subjective evidence, in particular, as more credible when it is delivered with emotion.

These findings have a number of implications. Much of the language people use in everyday conversation is subjective. People rarely refer directly to facts in conversation, even when those conversations concern issues like politics, culture, and more (see Kubin et al., 2021). As such, these results help to clarify the ways that people use emotional cues to decide which claims seem credible or worth repeating. This fits well with prior research suggesting that emotional cues are the most effective way to go viral on the internet (Brady et al., 2017; Robertson et al., 2023). Further, these findings have important implications for legal contexts. As prior work has found (e.g., Salerno, 2021; Nitschke et al., 2019), jurors tend to be swayed by emotional displays. These findings fit well with that research as jurors are tasked with evaluating credibility under conditions of uncertainty. Future research should further investigate whether these same patterns emerge for people with legal training, like lawyers and judges.

One methodological limitation is that all studies relied on a single care-related scenario, focused on illness and

friendship. Future research should explore whether the effects presented here are generalizable to other moral scenarios. Next, our three studies used materials from different languages and cultural contexts. Study 1 analyzed Reddit posts from the United States, while Study 2 and Study 3 used Chinese vignettes. Emotional expression varies in intensity, connotation, and social appropriateness across languages, and its interpretation varies across cultures (e.g., Tsai, 2007; Mesquita & Walker, 2003), and East Asian norms tend to favor emotional restraint and contextual subtlety (Markus & Kitayama, 1991; Matsumoto et al., 2008). Similar linguistic expressions may carry different meanings across cultures. Future research should explore whether similar experimental patterns of results emerge in US contexts. Lastly, we included only six pairs of evidence arguments from both sides of the dilemma and collected participants' ratings of credibility and usability. Future work will expand the evidence set to include objective, subjective, and purely emotional information, and examine the effect of intentionality, capability, and foreseeability. We will also include a participant selection process beyond rating.

The results of these experiments extend work on how emotion shapes information processing (e.g., Salerno, 2021; Nitschke et al., 2019) to everyday moral judgment. Emotional intensity affects how credible and usable people consider subjective moral evidence, even for relatively minor infractions. That is, emotions shape both what kinds of moral judgments people make (see e.g., Pizarro, 2000), and also what counts as evidence in the first place. In moral conflicts, trust often favors the emotional.

References

- Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(2), 169-178. <https://doi.org/10.1002/wcs.1216>
- Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition and Emotion*, 24(4), 561-595. <https://doi.org/10.1080/02699930903132496>
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318. <https://doi.org/10.1073/pnas.1618923114>
- Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual review of psychology*, 71(1), 499-515. <https://doi.org/10.1146/annurev-psych-010419-050807>
- Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9), 393-399. <https://doi.org/10.1016/j.tics.2007.08.005>
- Clore, G. L., Gasper, K., & Garvin, E. (2012). Affect as information. In *Handbook of affect and social cognition* (pp. 122-145). Psychology Press.

- Cummins, D. D., & Cummins, R. C. (2012). Emotion and deliberative reasoning in moral judgment. *Frontiers in Psychology*, 3, 328. <https://doi.org/10.3389/fpsyg.2012.00328>
- Engelmann, N., & Waldmann, M. R. (2022). How causal structure, causal strength, and foreseeability affect moral judgments. *Cognition*, 226, 105167. <https://doi.org/10.1016/j.cognition.2022.105167>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, 117(1), 39–66. <https://doi.org/10.1037/0033-2909.117.1.39>
- Frijda, N. H., Manstead, A. S., & Bem, S. (2000). The influence of emotions on beliefs. *Emotions and beliefs: How feelings influence thoughts*, 1–9.
- Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, 55(1), 96–135. <https://doi.org/10.1257/jel.20151245>
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385. <https://doi.org/10.1037/a0021847>
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>
- Greifeneder, R., Bless, H., & Pham, M. T. (2011). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review*, 15(2), 107–141. <https://doi.org/10.1177/1088868310367640>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>
- Hess, U., Banse, R., & Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of personality and social psychology*, 69(2), 280–288.
- Hofmann, W., & Baumert, A. (2010). Immediate affect as a basis for intuitive moral judgement: An adaptation of the affect misattribution procedure. *Cognition and emotion*, 24(3), 522–535. <https://doi.org/10.1080/02699930902847193>
- Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2021). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior research methods*, 53(1), 232–246. <https://doi.org/10.3758/s13428-020-01433-0>
- Hugging Face. (2024). *facebook/bart-large-mnli* [Pretrained language model]. Hugging Face Model Hub. <https://huggingface.co/facebook/bart-large-mnli>
- Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological science*, 12(1), 18–23. <https://doi.org/10.1111/1467-9280.00304>
- Jones, C. R. G., Pickles, A., Falcato, M., Marsden, A. J. S., Happé, F., Scott, S. K., Sauter, D., Tregay, J., Phillips, R. J., Baird, G., & Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 52(3), 275–285. <https://doi.org/10.1111/j.1469-7610.2010.02328.x>
- Kaufmann, G., Drevland, G. C., Wessel, E., Overskeid, G., & Magnussen, S. (2003). The importance of being earnest: Displayed emotions and witness credibility. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 17(1), 21–34. <https://doi.org/10.1002/acp.842>
- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic bulletin & review*, 25(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- Kubin, E., Puryear, C., Schein, C., & Gray, K. (2021). Personal experiences bridge moral and political divides better than facts. *Proceedings of the National Academy of Sciences*, 118(6), e2008389118. <https://doi.org/10.1073/pnas.2008389118>
- Levine, E. E., Barasch, A., Rand, D., Berman, J. Z., & Small, D. A. (2018). Signaling emotion and reason in cooperation. *Journal of Experimental Psychology: General*, 147(5), 702–719. <https://doi.org/10.1037/xge0000399>
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7871–7880). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.703>
- Liew, J. L. (2021). *Reddit dataset for AmItheAsshole subreddit* [Data set]. Kaggle. <https://www.kaggle.com/datasets/jianloongliew/reddit>
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>
- Matsumoto, D., Yoo, S. H., Nakagawa, S., & Multinational Study of Cultural Display Rules. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology*, 94(6), 925–937. <https://doi.org/10.1037/0022-3514.94.6.925>

- Mazzetti, G., Guglielmi, D., & Topa, G. (2020). Hard enough to manage my emotions: How hardness moderates the relationship between emotional demands and exhaustion. *Frontiers in Psychology*, *11*, 1194. <https://doi.org/10.3389/fpsyg.2020.01194>
- Mesquita, B., & Walker, R. (2003). Cultural differences in emotions: A context for interpreting emotional experiences. *Behaviour Research and Therapy*, *41*(7), 777–793. [https://doi.org/10.1016/S0005-7967\(02\)00189-4](https://doi.org/10.1016/S0005-7967(02)00189-4)
- Naseem, U., Razzak, I., & Eklund, P. W. (2021). A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter. *Multimedia Tools and Applications*, *80*(28), 35239–35266. <https://doi.org/10.1007/s11042-020-10082-6>
- Nitschke, F. T., McKimmie, B. M., & Vanman, E. J. (2019). A meta-analysis of the emotional victim effect for female adult rape complainants: Does complainant distress influence credibility? *Psychological Bulletin*, *145*(10), 953–979. <https://doi.org/10.1037/bul0000206>
- Ong, D. C., Zaki, J., & Goodman, N. D. (2019). Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in cognitive science*, *11*(2), 338–357. <https://doi.org/10.1111/tops.12371>
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.
- Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, *13*(4), 160–166. <https://doi.org/10.1016/j.tics.2009.01.006>
- Phillips, C. V., & Goodman, K. J. (2006). Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *Emerging Themes in Epidemiology*, *3*, 5. <https://doi.org/10.1186/1742-7622-3-5>
- Pizarro, D. (2000). Nothing more than feelings? The role of emotions in moral judgment. *Journal for the theory of social behaviour*, *30*(3), 355–375. <https://doi.org/10.1111/1468-5914.00135>
- Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational psychology review*, *31*(2), 339–359. <https://doi.org/10.1007/s10648-019-09473-5>
- Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations*, *9*(1), 29–43. <https://doi.org/10.1080/13869790500492466>
- Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature human behaviour*, *7*(5), 812–822. <https://doi.org/10.1038/s41562-023-01538-4>
- Salerno, J. M. (2021). The impact of experienced and expressed emotion on legal factfinding. *Annual Review of Law and Social Science*, *17*(1), 181–203. <https://doi.org/10.1146/annurev-lawsocsci-021721-072326>
- Skoe, E. E., Eisenberg, N., & Cumberland, A. (2002). The role of reported emotion in real-life and hypothetical moral dilemmas. *Personality and Social Psychology Bulletin*, *28*(7), 962–973. <https://doi.org/10.1177/014616720202800709>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, *29*(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, *2*(3), 242–259. <https://doi.org/10.1111/j.1745-6916.2007.00043.x>
- Ugazio, G., Lamm, C., & Singer, T. (2012). The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion*, *12*(3), 579–590. <https://doi.org/10.1037/a0024611>
- Van Kleef, G. A. (2010). The emerging view of emotion as social information. *Social and Personality Psychology Compass*, *4*(5), 331–343. <https://doi.org/10.1111/j.1751-9004.2010.00262.x>
- Wisneski, D. C., & Skitka, L. J. (2017). Moralization through moral shock: Exploring emotional antecedents to moral conviction. *Personality and Social Psychology Bulletin*, *43*(2), 139–150. <https://doi.org/10.1177/0146167216676479>
- Wong, D. B. (1992). Coping with moral conflict and ambiguity. *Ethics*, *102*(4), 763–784.
- Yu, H., Siegel, J. Z., & Crockett, M. J. (2019). Modeling morality in 3-D: Decision-making, judgment, and inference. *Topics in cognitive science*, *11*(2), 409–432. <https://doi.org/10.1111/tops.12382>
- Yudkin, D. A., Goodwin, G. P., Reece, A., Gray, K., & Bhatia, S. (2025). A large-scale investigation of everyday moral dilemmas. *PNAS nexus*, *4*(5), pgaf119. <https://doi.org/10.1093/pnasnexus/pgaf119>